Inter-platform Reproducibility of Liver Stiffness Measured with Two Different Point Shear Wave Elastography Techniques and Two-Dimensional Shear Wave Elastography using the Comb-Push Technique

Hwaseong Ryu³, Su Joa Ahn¹, Jeong Hee Yoon¹, Jeong Min Lee¹,²*

¹Department of Radiology, Seoul National University Hospital
²Institute of Radiation Medicine, Seoul National University College of Medicine
³Department of Radiology, Pusan National University Yangsan Hospital

*Corresponding author: Jeong Min Lee, M.D.

E-mail: jmsh@snu.ac.kr
Inter-platform Reproducibility of Liver Stiffness Measured with Two Different Point Shear Wave Elastography Techniques, and Two-Dimensional Shear Wave Elastography using the Comb-Push Technique

Type of Manuscript: Original Article
ABSTRACT

OBJECTIVE: To compare the technical success and reliable measurement rates of three shear wave elastography (SWE) techniques, and to assess the inter-platform reproducibility of the resultant liver stiffness measurements.

MATERIALS AND METHODS: This prospective study comprised 54 patients with liver diseases. Liver stiffness (LS) measurements were obtained using two-point SWE (pSWE) techniques (VTQ and S-Shearwave), two-dimensional (2D) SWE, and transient elastography (TE) serving as the reference standard. Technical success rates and reliable measurements of the three techniques were compared. LS values measured using the three techniques and TE were correlated using Spearman correlation coefficients and 95% Bland-Altman limits of agreement. Intra-class correlation coefficients (ICC) were used to analyze the inter-platform reproducibility of LS measurements.

RESULTS: Three SWE techniques and TE showed similar technical success rates ($p = 0.682$) but demonstrated significant differences in the reliability of LS measurements ($p =0.006$) and mean LS measurements ($p <0.0001$). Despite strong correlations ($r =0.73-0.94$) between SWE systems, various degrees of inter-platform reproducibility (ICC = 0.58-0.92) were observed for the three SWE techniques. The best agreement was observed between S-Shearwave and TE (ICC = 0.92), and the worst agreement was observed between 2D-SWE and TE (ICC = 0.58). In Bland-Altman analysis, tendency toward lower LS values with the three SWE techniques than TE in patients with F3 and F4 was observed.

CONCLUSION: Significant inter-system variability was observed in LS measurements of the three SWE techniques. Therefore, LS values measured using different SWE techniques should not be used interchangeably for longitudinal follow-up.
Keywords: Liver fibrosis; Liver stiffness; Ultrasound elastography; Shear wave elastography
INTRODUCTION

Chronic liver disease (CLD) stemming from hepatitis B or C viral infections, alcohol abuse, and non-alcoholic fatty liver disease (NAFLD) is a serious health concern worldwide [1] as liver fibrosis, its most common outcome, often results in cirrhosis, liver failure, and portal hypertension [2]. The progression of fibrosis to cirrhosis is also accompanied by a number of sequelae including distortion of the hepatic architecture and vasculature, deterioration of hepatic function, and increased risk of hepatocellular carcinoma (HCC) [3]. For the prognosis and management of patients with this disease, a previous investigation had shown that the amount and progression of liver fibrosis can be determining factors [4]. In addition, recent research on the molecular pathogenesis of liver fibrosis has shown that hepatic cellular recovery may be possible with the removal of fibrogenic stimuli [5]. Therefore, although it may be challenging, monitoring liver fibrosis remains an important clinical endeavor [6-8].

Liver biopsy has been considered the reference standard for the assessment of liver fibrosis to date. However, this assumption has recently been challenged owing to increasing awareness of its drawbacks [9,10] including invasiveness leading to severe complications [11], sampling error [12], and considerable inter- and intra-observer variability [13-15]. Thus, in recent years, noninvasive assessment of liver fibrosis has experienced explosive growth, and a wide spectrum of noninvasive methods ranging from serum assays to imaging techniques have been developed [10]. In particular, noninvasive imaging techniques such as transient elastography (TE), shear wave elastography (SWE), and MR elastography have had increasingly important roles in
liver fibrosis assessment [16-19]. Indeed, several studies have already demonstrated that liver stiffness (LS) measurements from TE (FibroScan, Echosens, Paris, France) correlate well with the advanced fibrosis stage of the liver [20,21] and that the diagnostic performances of point SWE (pSWE), and two-dimensional SWE (2D-SWE) using an acoustic radiation force impulse (ARFI) to generate shear waves were similar to that of TE according to a meta-analysis study [17]. The major benefit of ultrasound (US)-based SWE techniques over TE would be its add-on function during B-mode imaging which can allow the assessment of the underlying liver morphology, and screening for HCC in addition to stiffness measurements [20]. Yet, although any systemic bias would be critically important to rule out during the diagnosis and follow-up of patients with CLD, there have been few studies to date that have defined the reproducibility of the numerous types of SWE systems [22,23]. According to a recent experimental phantom study performed by the Ultrasound Shear Wave Speed technical committee of the Radiological Society of North America Quantitative Imaging Biomarker Alliance (QIBA), a statistically significant difference in shear wave speed estimates among commercial SWE systems were reported to the order of 12% although these findings have yet to be validated in clinical studies [1,24]. Furthermore, the percentage of unreliable LS measurements using SWE techniques was estimated to range between 6.7% ~ 10.4% for pSWE techniques and 10.2 ~23% for 2D SWE techniques [22,25-28]. From this results, we hypothesized that considerable variation between SWE systems would also be observed in patients.

Therefore, we prospectively evaluated the technical success rates, reliable LS measurements rates, and inter-platform reproducibility of LS measurements of two
kinds of p-SWE techniques and 2D-SWE using the comb-push technique in patients with chronic liver diseases.

MATERIALS AND METHODS

This prospective study was performed with approval from our Institutional Review Board. This prospective study was planned to be performed for four months, and the expected number of patient was ninety. Written informed consents were obtained from all patients prior to enrollment in this study.

Study population

Among the patients who were referred to the Department of Radiology at our institution for image-guided tumor ablation between May and September of 2017, the patients with suspected chronic liver disease or liver cirrhosis who were agreed to participate in this study were enrolled. The exclusion criteria were as follows: 1) age younger than 18 years; 2) patients who cannot hold their breath longer than 5 seconds; 3) patients who had undergone right hepatectomy; 3) patients who had multiple treated tumors in the right lobe of the liver. In total, 54 patients were included in our study. US examinations including SWE were performed for estimation of liver fibrosis and portal hypertension prior to tumor ablation.

Technical success rates and reliability of measurements of the three SWE techniques were assessed in all 54 patients. However, comparisons of LS measurements between the techniques were done only in patients in whom reliable LS values were obtained in all
three SWE systems and TE. Therefore, only 31 patients (24 men, 7 women; mean age, 66.6 years ± 9.49; age range, 38–80 years) were included for comparison of the inter-platform reproducibility of SWE techniques after excluding seven cases of technical failure and 16 cases of unreliable results at one or more techniques (Fig. 1). BMIs of all patients were recorded (mean BMI, 23.5 kg/m² ± 3.30; range, 18.2–30.7 kg/m²).

Etiologies of the chronic liver diseases in our study patients were chronic hepatitis B (n=26; 83.9%), chronic hepatitis C (n=2; 6.5%), and chronic non-viral hepatitis such as non-alcoholic steatohepatitis, alcoholic liver disease, and primary biliary cirrhosis (n=3; 9.6%).

Assessment of the degree of liver fibrosis in patients was made based on TE, as TE is the most validated method for liver fibrosis evaluation [29-31]. LS cut off values of TE were selected according to the latest meta-analysis data [30]: 7.9 kPa for moderate fibrosis (F ≥ 2), 8.8 kPa for severe fibrosis (F ≥ 3), and 11.7 kPa for liver cirrhosis (F = 4).

In the 31 patients who had reliable LS measurements in all examinations, the most common fibrosis stage was liver cirrhosis (F4) (17/31, 54.8%), followed by severe fibrosis (F3), (6/31, 19.4%), mild (F1) or no liver fibrosis (F0), (5/31, 16.1%), and moderate (F2) liver fibrosis (3/31, 9.7%) (Table 1).

**SWE examinations**

All patients underwent US examination after fasting for more than 6 hours. All US examinations were performed by one radiologist (O.O.O) who had six years of experience in US-based elastography including pSWE, 2D-SWE, and TE (> 200 examinations) and had 20 years of experience with abdominal US examinations.
At first, conventional B-mode sonography using a 4 MHz convex probe was used to assess the focal liver lesion during planning US examination to determine the feasibility of ablation therapy. After that, LS measurements were performed using the intercostal approach while patients were placed in the supine position with their right arm in maximum abduction during the SWE examination. LS measurements of each patient was done with S-Shearwave using the Samsung RS 80A US system (Samsung Medison, Seoul, Korea), Virtual Touch Quantification (VTQ) using the Siemens Acuson S2000 Virtual Touch US system (Siemens AG, Erlangen, Germany), 2D-SWE with the comb-push (CP) technique using the LOGIQ S8 US system (GE Healthcare, Wauwatosa, WI, USA), and TE using FibroScan (Echosens, Paris, France) added to LOGIQ S8 within a 24 hour interval for each patient. For VTQ, and S-Shearwave, a region of interest (ROI) was placed in the right anterior segment of the liver at a depth of 2.0 cm from the liver capsule so as not to include any focal liver lesions or vessels. Similarly, for 2D-SWE, a 1 × 1 cm² region-of-interest (ROI) was placed in the right anterior segment of the liver, taking care to avoid large vessels and areas with artifacts, 2.0 cm away from the Glisson capsule, and less than 6 cm deep from the transducer (Fig. 2).

The operator who conducted SWE examination performed all FibroScan examinations. The operator had performed more than 100 TE examinations and carried out all TE examinations according to the manufacturer’s recommendations: the tip of the transducer probe (M+ probe or XL+ probe when prompted by the automatic probe selection tool) was placed on the skin between the ribs over the right lobe of the liver and valid LS measurements were obtained under the guidance of M-mode monographic images. During LS measurement, the patients were instructed to hold their breath while avoiding
deep inspiration or expiration. At least 10 valid measurements were performed in every patient for every methods of SWE.

**Definition of Technical Failure and Reliable (or Unreliable) Measurement**

Technical failure of SWE methods and TE was defined as failure to acquire 10 valid measurements after at least 15 trials [22]. If the interquartile range (IQR)/median LS ratio was higher than 30%, the result was regarded to be an unreliable measurement [32]. To avoid any potential bias, the summary of the serial measurements of each technique was not made available to the operator until the three SWE techniques and TE examinations were completed [22].

**Statistical analysis**

LS values were expressed in kPa for the S-Shearwave and 2D-SWE techniques, while the LS values in m/s from VTQ were converted to the Young's modulus [33]. Continuous data were summarized as means and data range, and categorical data were summarized as counts and percentages. The Friedman test with Bonferroni correction was used to compare the technical failure rates, and unreliable measurement rates between the three different SWE imaging systems. For comparison of BMI and fibrosis stage between patients with reliable LS and unreliable LS measurements, Student’s t-test and the Mann-Whitney test were used. The Wilcoxon signed-rank test was used to compare LS measurement in pairwise analysis. Spearman correlation coefficients and two-way mixed model intra-class correlation coefficients (ICCs) with 95% confidence intervals (CIs) were obtained to evaluate the agreement between the different SWE techniques. Correlation coefficients were classified using the following
definitions: 0.0-0.19, very weak; 0.2-0.39, weak; 0.40-0.59, moderate; 0.60-0.79, strong; and 0.80-1.0, very strong [34]. Agreement based on ICCs was classified using the following definitions: 0-0.39, poor; 0.40-0.59, fair; 0.60-0.74, good; and 0.75-1.0, excellent [35]. In addition, Bland-Altman analysis was used to evaluate method-related variations using the mean value between the different SWE systems. 95% limits of agreement, as well as the coefficient of reproducibility (CR = 1.96 × SD of bias) were determined for inter-platform variability for the liver LS measurements. The coefficient of variation (CV) of the LS values between the SWE techniques was also calculated. Areas under the receiver operating characteristic curve (AUROC) were built for VTQ, S-Shearwave, and 2D-SWE to detect significant liver fibrosis (F ≥ 2) using the LS values of TE as the reference standard. Optimal cutoff values were determined using the highest Youden Index and the DeLong test was used to compare AUROC curves. All statistical analyses were performed using commercially available software programs (SPSS version 23, SPSS, IBM, Armonk, NY, USA; or MedCalc version 16, MedCalc Software, Mariakerke, Belgium) with a p-value of less than 0.05 considered to indicate a statistically significant difference.

RESULTS

Technical failure and unreliable measurement rates

Ten LS measurements of the three SWE techniques and TE were made successfully in 47 of the 54 patients (87%) (Fig. 1). LS measurements were not able to be obtained with
VTQ in one patient (1/54, 1.9%), with S-Shearwave in three patients (3/54, 3.7%), with 2D-SWE in two patients (3/54, 3.7%), and with TE in three patients (3/54, 3.7%). In addition, LS values could not be measured with both VTQ and TE in one patient, with both S-Shearwave and TE in one patient, and with both 2D-SWE and TE in one patient. There were no significant differences in technical success rates between the SWE techniques as well as TE ($p = 0.682$).

Among the 47 patients with technically successful LS measurements, reliable LS measurements were obtained in 97.9% (46/47) of patients with VTQ, 100% (47/47) with S-Shearwave, 83.0% (39/47) with 2D-SWE, and 85.1% (40/47) with TE. There was a significant difference in reliable LS measurement rates between the SWE techniques ($p = 0.006$). According to pairwise analysis, significant differences (P<0.017 after Bonferroni correction) were observed in the reliable measurement rates of S-Shearwave and 2D-SWE ($p = 0.005$). Conversely, no significant differences were observed in reliable measurement rates between VTQ and S-Shearwave ($p = 0.317$), and VTQ and 2D-SWE ($p = 0.020$).

**Correlation and interplatform reproducibility of LS values across different SWE techniques**

Mean LS values for the two different pSWE techniques and 2D-SWE were significantly different from those of TE ($p<0.0001$) (Table 2). According to pairwise analysis, significant differences were observed in mean LS values between VTQ and TE ($p<0.0001$), S-Shearwave and TE ($p<0.0001$), and 2D-SWE and TE ($p=0.001$). CVs for the SWE techniques ranged between 20.8 and 40.6 (Table 3).
The ICC of LS measurements for all SWE techniques was 0.87, indicating excellent agreement (95% CI: 0.16, 0.94). When assessing the agreement between each of the three SWE techniques and TE (n=3), pairwise ICCs ranged from 0.58 to 0.92. The best agreement was observed between S-Shearwave and TE (ICC = 0.92, r = 0.94). The worst agreement was observed between 2D-SWE and TE (ICC = 0.58, r = 0.88). In patients with F4, S-Shearwave showed best correlation with TE, and 2D-SWE showed worst correlation with TE (Fig. 3). In addition, Bland-Altman plots for reproducibility between TE and other SWE techniques showed a tendency toward lower LS values with the three SWE techniques than with TE in patients with F3 and F4 (Fig. 4).

**Performance of the three SWE techniques in detecting significant fibrosis (F ≥ 2)**

Using the LS values of TE as the reference standard, VTQ and S-Shearwave showed an AUROC of 0.90 (95% CI 0.74-0.98) and an AUROC of 0.99 (95% CI 0.860-1.000) in detecting significant fibrosis, respectively. 2D-SWE had an AUROC of 0.97 (95% CI 0.835-0.999). In the pairwise ROC curve comparison, the AUROCs of the three SWE techniques for the prediction of significant fibrosis were not significantly different (p = 0.163-0.612) (Fig. 5).

**DISCUSSION**

In our study comprised of patients with chronic liver diseases, we compared the LS measurements obtained from three commercially available SWE systems, each from
different manufacturers. From this comparison, we found that although there were no significant differences in technical success rates, the pSWE methods (VTQ and S-Shearwave) showed significantly higher rates of reliable LS measurements than 2D-SWE. Our study results are in good agreement with that of a previous study by Sporea et al. who also reported a significantly higher percentage of reliable LS measurements with VTQ compared with TE and 2D-SWE (supersonic shear imaging, SSI) [36]. Our study also demonstrated that the mean LS values of the two pSWE techniques and 2D-SWE were significantly different ($p = 0.006$) although the overall ICC for LS measurements for all SWE techniques ($n$=4) was 0.87, indicating excellent agreement (95% CI: 0.16, 0.94). In addition, the mean LS values obtained with the three SWE techniques were significantly lower than that obtained with TE. Our study results are quite similar to those of a previous study by Bende et al. which demonstrated substantially lower LS values with 2D-SWE than with TE [37]. In addition, our study demonstrated that S-Shearwave and VTQ correlated with TE more than 2D-SWE in patients with liver cirrhosis ($F=4$). This inter-system variability could be attributed to a number of system factors, in particular, shear wave vibration frequency and bandwidth, as well as the software’s method of calculating shear wave speed [38]. Therefore, we believe that as inter-system variability was consistently observed across the different SWE techniques, different cut-off values for fibrosis staging should be used for the two pSWE systems, 2D-SWE, and TE.

We also found in our study that although the three SWE techniques showed different optimal cut-off values (6.8-8.73 kPa) for diagnosing significant fibrosis ($\geq F2$), the
AUROCs of the three SWE techniques were not significantly different for the detection of significant fibrosis ($p = 0.163-0.612$) when using cut-off values of TE as the reference standard. Our results are in good agreement with the results of other previous studies [22,36,39,40] including that of Gerber et al. [40] who reported no significant differences in AUROCs between 2D-SWE, pSWE, and TE in the diagnosis of significant fibrosis and Sporea et al. [36] who also published the similar diagnostic accuracies of VTQ and 2D-SWE (SSI) in the diagnosis of significant fibrosis. Thus, although comparing the results obtained by different elastography techniques may be challenging due to non-standardized reported parameters, differing shear-wave frequencies, and other technical parameters [24], pSWE systems and 2D-SWE seem to show similar accuracy in fibrosis staging.

There are some limitations in our study that need to be mentioned. Because this study was intended for patients who were hospitalized for image-guided hepatic tumor ablation, the study population was relatively small and showed disproportionate distribution of liver fibrosis grades. This may have limited our assessment of the diagnostic performance of each SWE system in fibrosis staging. Including outpatients might be helpful to overcome this limitation, but performing repeated SWE examination in addition to TE examination is difficult in outpatient based environment. In addition, one radiologist performed all SWE examination, so we did not analyze intra or inter-observer variability. However, we believe that our study may serve as the first step toward a future study to evaluate inter-platform reproducibility of SWE systems. Secondly, a histological diagnosis of fibrosis staging was not performed in our study.
However, the primary goal of our study was to evaluate the inter-system variability of LS measurements rather than comparing the diagnostic performance of each SWE system.

In conclusion, although the three commercially available SWE techniques showed similar technical success rates, significant inter-system variability was observed in LS measurements. Therefore, LS values measured using different SWE techniques should not be used interchangeably for longitudinal follow-up, and cut-off values established for one SWE technique should not be applied to other SWE techniques.
Reference

cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454-462.


Fig. 1. Study design

Each one patient showed technical failure by both VTQ and TE, both S-Shearwave and TE, and both 2D-SWE and TE

Fig. 2. LS measurements of the SWE techniques

A, B, C. LS measurements were performed using VTQ (A), S-Shearwave (B), 2D-SWE (C), and TE (D). The ROIs were placed in the right anterior segment of the liver, avoiding vascular structures.

D. Slope of the line at the right panel of TE measurement indicates shear wave speed

Fig. 3. Correlation between SWE techniques

Scatter diagrams show strong correlations of LS measurements between VTQ and TE (A), S-Shearwave and TE (B), and 2D-SWE and TE (C) in this study

Fig. 4. Bland-Altman plots of SWE techniques

Bland-Altman plots demonstrate difference in LS values between VTQ and TE (A), S-Shearwave and TE (B), and 2D-SWE and TE (C). The solid blue line in the middle represents the mean of LS values obtained from each pair of three systems and TE, and the dotted brown lines define ± 1.96 standard deviations (SD), with associated 95% confidence intervals indicated by thin blue lines.
Fig.5. Comparison of ROC curves

The AUROCs of the VTQ, S-Shearwave, and 2D-SWE for the prediction of significant fibrosis (F ≥ 2) were not significantly different (p = 0.163-0.612)
Table 1. Patient characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>66.6 ± 9.49 yrs (38-80)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24 (77.4%)</td>
</tr>
<tr>
<td>Female</td>
<td>7 (22.6%)</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td>23.5 ± 3.30 kg/m²</td>
</tr>
<tr>
<td>Underweight (BMI &lt;18.5 kg/m²)</td>
<td>1 (3.2%)</td>
</tr>
<tr>
<td>Normal weight (BMI =18.5-24.9 kg/m²)</td>
<td>9 (29.0%)</td>
</tr>
<tr>
<td>Overweight (BMI =25-29.9 kg/m²)</td>
<td>20 (64.6%)</td>
</tr>
<tr>
<td>Obese (BMI &gt;30 kg/m²)</td>
<td>1 (3.2%)</td>
</tr>
<tr>
<td><strong>Etiology of liver disease</strong></td>
<td></td>
</tr>
<tr>
<td>Chronic hepatitis B</td>
<td>26 (83.9%)</td>
</tr>
<tr>
<td>Chronic hepatitis C</td>
<td>2 (6.5%)</td>
</tr>
<tr>
<td>Chronic non-viral hepatitis (NAFLD, alcoholic, PBC)</td>
<td>3 (9.6%)</td>
</tr>
<tr>
<td><strong>Fibrosis grade</strong>*</td>
<td></td>
</tr>
<tr>
<td>&lt;F2</td>
<td>5 (16.1%)</td>
</tr>
<tr>
<td>F2</td>
<td>3 (9.7%)</td>
</tr>
<tr>
<td>F3</td>
<td>6 (19.4%)</td>
</tr>
<tr>
<td>F4</td>
<td>17 (54.8%)</td>
</tr>
</tbody>
</table>

* Distribution of liver fibrosis stages using the transient elastography cut-off values proposed by a previous meta-analysis study (reference #31).

Note. – BMI = body mass index, NAFLD = nonalcoholic fatty liver disease, PBC = primary biliary cirrhosis
Table 2. Mean LS values obtained using two point SWE techniques, 2D-SWE, and transient elastography

<table>
<thead>
<tr>
<th></th>
<th>Mean LS value (kPa)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VTQ</td>
<td>10.5 ± 5.05 (3.12-21.1)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>S-Shearwave</td>
<td>12.2 ± 6.70 (3.70-31.5)</td>
<td></td>
</tr>
<tr>
<td>2D-SWE</td>
<td>10.6 ± 2.83 (5.64-17.6)</td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>15.1 ± 8.67 (3.80-39.6)</td>
<td></td>
</tr>
</tbody>
</table>

* Mean LS values obtained by each of three SWE techniques were compared with TE in pairwise manner using the Wilcoxon signed-rank test.

Note. – VTQ = Virtual Touch™ Quantification, 2D-SWE = two-dimensional shear wave elastography, TE = transient elastography
<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>ICC</th>
<th>CV(%)</th>
<th>CR(%)</th>
<th>Mean bias (%)</th>
<th>BALA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VTQ vs S-Shearwave</strong></td>
<td>0.77 (0.57, 0.88)</td>
<td>0.84 (0.65, 0.92)</td>
<td>28.1 (20.1, 36.6)</td>
<td>8.73 (7.00, 11.6)</td>
<td>-13.8 (-25.3, -2.13)</td>
<td>[-75.8, 48.3]</td>
</tr>
<tr>
<td><strong>S-Shearwave vs 2D-SWE</strong></td>
<td>0.84 (0.69, 0.92)</td>
<td>0.73 (0.44, 0.87)</td>
<td>22.5 (16.2, 29.1)</td>
<td>9.31 (7.46, 12.4)</td>
<td>4.98 (-5.30, 15.3)</td>
<td>[-49.9, 59.9]</td>
</tr>
<tr>
<td><strong>2D-SWE vs VTQ</strong></td>
<td>0.73 (0.50, 0.86)</td>
<td>0.76 (0.51, 0.89)</td>
<td>30.9 (22.1, 40.3)</td>
<td>6.93 (5.55, 9.21)</td>
<td>8.50 (-4.76, 21.8)</td>
<td>[-62.3, 79.3]</td>
</tr>
<tr>
<td><strong>VTQ vs TE</strong></td>
<td>0.78 (0.58, 0.89)</td>
<td>0.72 (0.19, 0.89)</td>
<td>40.6 (28.7, 53.6)</td>
<td>14.2 (11.4, 18.9)</td>
<td>-31.7 (-43.9, -19.5)</td>
<td>[-96.7, 33.4]</td>
</tr>
<tr>
<td><strong>S-Shearwave vs TE</strong></td>
<td>0.94 (0.88, 0.97)</td>
<td>0.92 (0.58, 0.97)</td>
<td>20.8 (15.0, 26.8)</td>
<td>8.70 (6.97, 11.6)</td>
<td>-18.7 (-25.5, -11.8)</td>
<td>[-55.2, 17.9]</td>
</tr>
<tr>
<td><strong>2D-SWE vs TE</strong></td>
<td>0.88 (0.76, 0.94)</td>
<td>0.58 (0.06, 0.80)</td>
<td>34.6 (24.6, 45.3)</td>
<td>15.3 (12.2, 20.3)</td>
<td>-23.3 (-35.7, -10.9)</td>
<td>[-89.5, 42.9]</td>
</tr>
</tbody>
</table>

**Note.**  - r = Spearman r correlation coefficient, ICC = Intra-class correlation, CV = coefficient of variation (%), CR = coefficient of reproducibility (%), BALA = Bland-Altman limits of agreement (%)
Total patients considered eligible (n = 54)

Excluded (n = 23)

- VTQ (n=2)
  - Technical failure 1
  - Unreliable data 1

- S-Shearwave (n=3)
  - Technical failure 3
  - Unreliable data 0

- 2D-SWE (n=11)
  - Technical failure 3
  - Unreliable data 8

- TE (n=10)
  - Technical failure 3
  - Unreliable data 7

Total patients included in statistical analysis (n = 31)
Liver stiffness (LS) measurements of the VTQ
Figure 2b

Liver stiffness (LS) measurements of the S-Shearwave
Liver stiffness (LS) measurements of the 2D-SWE
Figure 2d

Liver stiffness (LS) measurements of the TE
Figure 3a

LS measurement correlation between VTQ and TE
Figure 3b

LS measurement correlation between S-Shearwave and TE
Figure 3c

LS measurement correlation between 2D-SWE vs TE
Figure 4a

Bland-Altman plots from VTQ and TE
Figure 4b

Bland-Altman plots from S-shearwave and TE
Figure 4c

Bland-Altman plots from 2D-SWE and TE
Figure 5

Comparison of ROC curves for VTQ, S-Shearwave, and 2D-SWE for the detection of significant fibrosis (F ≥ 2)