



Reproducibility of shear wave elastography among operators, machines, and probes in an elasticity phantom

ULTRA
SONO
GRAPHY

Abdulaziz Ibrahim Alrashed¹, Abdulrahman M. Alfuraih²

¹Radiodiagnostics and Medical Imaging Department, Prince Sultan Military Medical City, Riyadh; ²Radiology and Medical Imaging Department, College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Kharij, Saudi Arabia

ORIGINAL ARTICLE

<https://doi.org/10.14366/usg.20011>
pISSN: 2288-5919 • eISSN: 2288-5943
Ultrasonography 2021;40:158-166

Purpose: This study was aimed to investigate the reproducibility of shear wave elastography (SWE) among operators, machines, and probes in a phantom, and to evaluate the effect of depth of the embedded inclusions and the accuracy of the measurements.

Methods: *In vitro* stiffness measurements were made of six inclusions (10, 40, and 60 kPa) embedded at two depths (1.5 cm and 5 cm) in an elastography phantom. Measurements were obtained by two sonographers using two ultrasound machines (the SuperSonic Imagine Aixplorer with the XC6-1, SL10-2 and SL18-5 probes, and the General Electric LOGIQ E9 with the 9L-D probe). Variability was evaluated using the coefficient of variation. Reproducibility was calculated using intraclass correlation coefficients (ICCs).

Results: For shallow inclusions, low variability was observed between operators (range, 0.9% to 5.4%). However, the variability increased significantly for deep inclusions (range, 2.4% to 80.8%). The measurement difference between the operators was 1%–15% for superficial inclusions and 3%–43% for deep inclusions. Inter-operator reproducibility was almost perfect (ICC>0.90). The measurement difference between machines was 0%–15% for superficial inclusions and 38.6%–82.9% for deep inclusions. For superficial inclusions, the reproducibility among the three probes was excellent (ICC>0.97). The mean stiffness values of the 10 kPa inclusion were overestimated by 16%, while those of the 40 kPa and 60 kPa inclusions were underestimated by 42% and 48%, respectively.

Conclusion: Phantom SWE measurements were only reproducible among operators, machines, and probes at superficial depths. SWE measurements acquired in deep regions should not be used interchangeably among operators, machines, or probes.

Keywords: Phantoms; Elasticity imaging techniques; Ultrasonography; Reproducibility of results; Transducers

Received: January 25, 2020

Revised: April 11, 2020

Accepted: May 9, 2020

Correspondence to:

Abdulrahman M. Alfuraih, PhD,
Radiology and Medical Imaging
Department, College of Applied
Medical Sciences, Prince Sattam bin
Abdulaziz University, Office 146,
Alkharj 6838-16273, Saudi Arabia

Tel. +966530005853

E-mail: a.alfuraih@psau.edu.sa

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2021 Korean Society of
Ultrasound in Medicine (KSUM)



How to cite this article:

Alrashed AI, Alfuraih AM. Reproducibility of shear wave elastography among operators, machines, and probes in an elasticity phantom. Ultrasonography. 2021 Jan;40(1):158-166.

Introduction

Pathological changes can result in alterations in tissue stiffness, as malignant lesions are usually stiffer than the surrounding tissue [1,2]. New methods, such as ultrasound shear wave elastography (SWE), have been sought to help characterise lesions' elasticity. SWE is safe, relatively inexpensive

and provides a non-invasive real-time assessment of tissue stiffness [3]. The principle of SWE is based on measuring the velocity of the propagating shear waves generated using acoustic radiation forces [3]. The shear wave velocity (SWV) is proportional to the tissue's stiffness, as shear waves travel faster in a hard lesion, while they travel slower in a soft lesion. The SWV is then converted into the Young modulus, a physical parameter that corresponds to the elasticity of the tissue [3].

SWE is a useful diagnostic and monitoring tool for various diseases [4,5]. However, its application can be hindered by discrepancies in SWE technologies across different manufacturers' systems [6–8]. Consistent image generation should be achieved for any new imaging technique before it becomes established in clinical practice. The Quantitative Imaging Biomarker Alliance (QIBA) works to achieve better consistency among the values reported by different manufacturer systems [9]. The QIBA also created a processing code and an open-source repository of research sequences to help manufacturers with calibration to achieve more consistent results [10]. However, to date, little is known regarding the reproducibility of SWE among different systems and particularly among different probes used with the same system. Additionally, no studies have combined a thorough investigation among machines, probes, and operators with a focus on the absolute differences between the measurements. To date, only one company (Computerized Imaging Reference Systems [CIRS], Norfolk, VA, USA) manufactures quality assurance (QA) elastography phantoms. Their phantoms have been widely investigated, especially models 049 [11–16], 049A

[17,18], and 039 [19–21] for liver fibrosis. However, no studies have investigated the multi-purpose QA phantom (model 040GSE, CIRS), which is equipped with elasticity targets and B-mode imaging test objects [22]. As a general-purpose QA phantom, this model is expected to be more widely available, making it worthy of investigation.

The primary aim of this study was to investigate the reproducibility of SWE results among different operators, machines, and probes on a phantom. The secondary aims were to evaluate the effect of depth on SWE values and to assess the accuracy of the acquired measurements in comparison with the manufacturer's reported values.

Materials and Methods

Phantom

This study was performed on a QA phantom (multi-purpose, multi-tissue ultrasound phantom, model 040GSE, CIRS) [22], made of Zerdine solid elastic hydrogel with a density of 1,050 kg/m³ and an acoustic speed of 1,540 m/s (Fig. 1). It had six elasticity inclusions situated at two depths, superficial (1.5 cm) and deep (5 cm). Each depth had three spherical isoechoic inclusions with respective stiffnesses of 10, 40, and 60 kPa. The superficial inclusions had a diameter of 6 mm each, and the deep inclusions had a diameter of 8 mm each.

Ultrasonography

Two ultrasound machines equipped with an SWE mode were used:

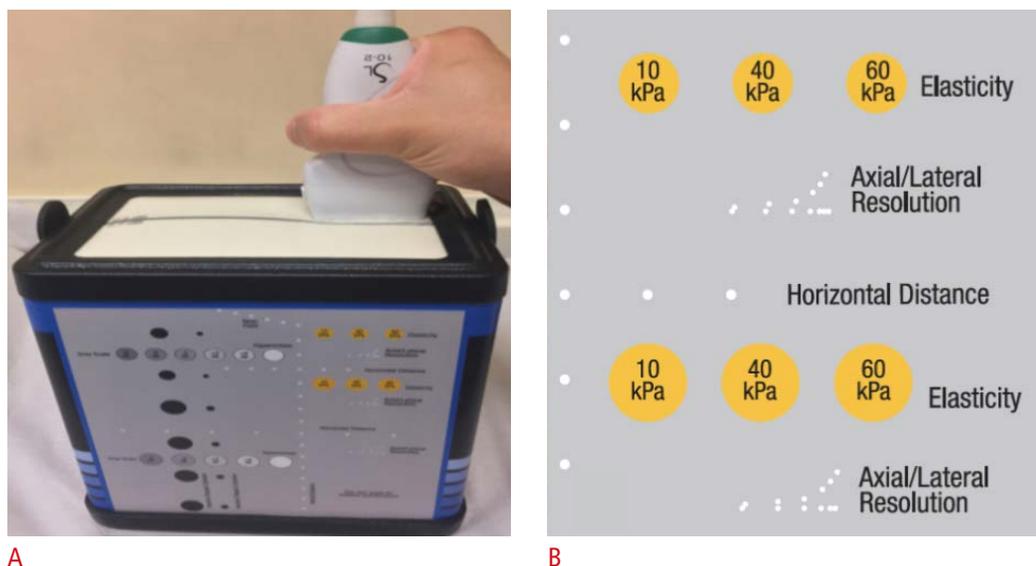


Fig. 1. Elastography phantom.

A. The image shows the external shape of the quality assurance phantom and the position of the ultrasound probe. **B.** The yellow circles represent the position, stiffness, and shape of different elasticity inclusions in the phantom (reprinted with permission from Computerized Imaging Reference Systems [22]).

SuperSonic Imagine Aixplorer (SSI; Aix-en-Provence, France), and General Electric LOGIQ E9 (GE Ultrasound, Milwaukee, IL, USA). Each system uses a unique SWE method. The SSI system emits multiple acoustic radiation force pulses at different depths, forming a Mach cone wave to perturb the tissue. Then, to track the SWV, it sends Ultrafast plane waves into the targeted tissue to insonify the entire imaging plane in a single shot [2]. In contrast, the GE system uses the comb-push excitation technique to emit multiple acoustic radiation pulses simultaneously from different parallel locations before employing time-interleaved tracking to estimate the SWV [23].

The SSI had SWE available for three probes: probe 1 was the single crystal curved XC6-1 MHz, probe 2 was the SuperLinear SL10-2 MHz, and probe 3 was the SuperLinear SL18-5 MHz. The GE device only had the SWE mode available for the linear 9L-D (2–8 MHz) probe. Hence, for inter-machine reproducibility, only the similar linear array probes—the SuperLinear SL10-2 and the 9L-D—were used to compare SWE values. For the inter-probe reproducibility, all

the inclusions were measured using the SSI device with the three different probes to compare the SWE values. The reproducibility of the technique was also assessed in terms of inter-operator reproducibility by comparing the SWE values obtained by two sonographers using the two ultrasound machines.

Data Acquisition

The data were acquired by two certified sonographers, each of whom had more than 7 years of experience. One of the sonographers (operator B) had 2 years of daily SWE experience, and the other (operator A) received SWE training by operator B before acquiring the measurements. The measurements were acquired in the transverse plane and after waiting for the SWE to stabilize (approximately 3 frames). A minimal probe load was applied on the surface [16]. Stiffness measurements of all inclusions were obtained using the penetration mode with high SWE gain (95%) to detect the stiffness of the deep inclusions. The SWE region of interest

Table 1. Descriptive summary of the mean elasticity (kPa) and differences between operators

Machine	Inclusion	Operator A		Operator B			
		Mean±SD (kPa)	CV (%)	Mean±SD (kPa)	CV (%)	Mean difference (95% LOA)	Difference (%)
SSI probe 1	10 kPa at 1.5 cm	18.3±0.5	2.6	18.1±0.2	1.2	0.2 (−0.8 to 1.1)	1.0
	40 kPa at 1.5 cm	27.2±0.7	2.4	24.4±0.6	2.4	2.8 (1.5 to 4.1)	10.2
	60 kPa at 1.5 cm	31.4±0.6	2.0	29.1±0.6	2.0	2.3 (1.1 to 3.6)	7.5
	10 kPa at 5 cm	13.5±0.6	4.6	13.9±0.5	3.4	−0.4 (−1.7 to 0.8)	−3.3
	40 kPa at 5 cm	24.5±0.7	2.8	22.5±0.3	2.9	2.0 (0.6 to 3.3)	8.0
	60 kPa at 5 cm	29.9±0.7	2.4	28.1±0.8	2.9	1.8 (0.4 to 3.2)	6.0
SSI probe 2	10 kPa at 1.5 cm	11.0±0.4	3.3	10.8±0.3	3.1	0.3 (−0.4 to 1.0)	2.5
	40 kPa at 1.5 cm	24.5±0.9	3.7	23.2±0.5	2.1	1.3 (−0.5 to 3.1)	5.4
	60 kPa at 1.5 cm	33.5±0.8	2.3	34.4±1.1	3.3	−1.0 (−2.5 to 0.5)	−2.9
	10 kPa at 5 cm	10.7±1.0	9.1	10.4±1.4	13.6	0.3 (−1.6 to 2.2)	3.0
	40 kPa at 5 cm	37.9±8.1	21.3	26.9±7.0	26.2	11.0 (−4.8 to 26.8)	29.1
	60 kPa at 5 cm	59.6±8.2	13.8	33.9±4.8	14.1	25.6 (9.5 to 41.8)	43.0
SSI probe 3	10 kPa at 1.5 cm	10.3±0.5	4.8	9.2±0.2	2.2	1.1 (0.1 to 2.1)	10.6
	40 kPa at 1.5 cm	24.3±0.4	1.5	23.2±0.5	2.0	1.1 (0.4 to 1.8)	4.6
	60 kPa at 1.5 cm	35.3±0.5	1.5	34.9±0.3	0.9	0.4 (−0.6 to 1.4)	1.2
	10 kPa at 5 cm	–	–	–	–	–	–
	40 kPa at 5 cm	–	–	–	–	–	–
	60 kPa at 5 cm	–	–	–	–	–	–
GE	10 kPa at 1.5 cm	12.7±0.7	5.4	10.8±0.2	1.9	1.9 (0.6 to 3.3)	15.2
	40 kPa at 1.5 cm	26.8±0.5	1.7	26.1±1.3	4.9	0.7 (−0.2 to 1.6)	2.6
	60 kPa at 1.5 cm	36.7±1.4	3.7	35.9±1.5	4.1	0.9 (−1.8 to 3.5)	2.3
	10 kPa at 5 cm	6.6±2.3	35.5	5.9±1.2	20.8	0.7 (−3.9 to 5.3)	10.7
	40 kPa at 5 cm	7.6±2.4	31.3	6.4±2.2	34.5	1.2 (−3.4 to 5.9)	15.9
	60 kPa at 5 cm	10.2±8.2	80.6	6.9±1.8	26.9	3.4 (−12.8 to 19.5)	32.9

(–) indicates failed to acquire shear wave elastography measurements.

SD, standard deviation; CV, coefficient of variation; 95% LOA, Bland-Altman 95% limits of agreement; SSI, Supersonic Imagine machine; GE, General Electric machine.

(ROI) was placed in the center of the inclusion with a diameter of 3 mm to avoid counting background stiffness. Following this, the mean stiffness of the ROI was recorded. Each operator acquired the reading blinded to the other. The probe was lifted and replaced between acquisitions. With the goal of obtaining highly accurate results and assessing measurement variability, each inclusion was measured 10 times [13,14].

Statistics

The data were analyzed using several descriptive and inferential statistical tests. The analysis was performed using SPSS version 23 (IBM Corp., Armonk, NY, USA). The coefficient of variation (CV) was used to measure intra-operator variability to evaluate the precision and repeatability of the measurements. The CV was calculated using the following equation: $CV = (\text{Standard deviation} / \text{Mean}) \times 100\%$. If the CV was close to zero, there was excellent repeatability and minimally dispersed measurements [24]. Additionally, the intraclass correlation coefficient (ICC) was used to assess inter-operator, inter-machine, and inter-probe reproducibility by evaluating measurement consistency. The ICC values were interpreted as follows: 0.00–0.20 indicates poor agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and >0.80 indicates almost perfect agreement [25].

The difference in SWE between the superficial and deep inclusions was tested using the paired-samples t test. The accuracy of the SWE measurements compared to the phantom values was assessed using the one-sample t test. A P-value <0.05 was considered to indicate

a statistically significant difference. To demonstrate the degree of agreement between two observers or measurement methods, Bland-Altman limits of agreement were employed [26]. Error bars were also plotted to display the mean stiffness according to SWE and the 95% confidence interval (CI). Differences between the manufacturer’s reported elasticity value and the experimental value were reported as $\text{Bias} = [(\text{Manufacturer value} - \text{Mean experimental value}) / \text{Manufacturer value}] \times 100\%$.

Results

Overall, 420 measurements were obtained for this study. The two operators acquired 10 measurements on six inclusions using four probes (excluding probe 3 on the SSI machine, which failed to acquire SWE readings for the three deep inclusions). The results are described in Table 1. A sample of the SWE images is shown in Fig. 2.

Comparison of SWE between the Operators

The inter-operator reproducibility (ICC [95% CI]) was 0.98 [0.70–0.99] for SSI probe 1, 0.84 [0.12–0.98] for SSI probe 2, 0.99 [0.78–1.0] for SSI probe 3, and 0.99 [0.81–0.99] for GE, demonstrating almost perfect overall agreement. The percentage difference between the operators was less than 10% for most inclusions, except the stiffer deep inclusions (Table 1). The intra-operator variability of operator A was slightly higher than that of operator B.

Comparison of SWE between Machines

The descriptive statistics for agreement between the machines are

Table 2. Descriptive summary of the mean elasticity (kPa) and differences between machines

Machine	Inclusion	SSI		GE			
		Mean±SD (kPa)	CV (%)	Mean±SD (kPa)	CV (%)	Mean difference (95% LOA)	Percentage difference (%)
Operator A	10 kPa at 1.5 cm	11.0±0.4	3.3	12.7±0.7	5.4	−1.7 (−2.4 to −0.9)	15.0
	40 kPa at 1.5 cm	24.5±0.9	3.7	26.8±0.5	1.7	−2.3 (−4.1 to −0.5)	9.6
	60 kPa at 1.5 cm	33.5±0.8	2.3	36.7±1.4	3.7	−3.3 (−4.7 to −1.8)	9.7
	10 kPa at 5 cm	10.7±1.0	9.1	6.6±2.3	35.5	4.1 (2.2 to 6.0)	−38.6
	40 kPa at 5 cm	37.9±8.1	21.3	7.6±2.4	31.3	30.3 (14.5 to 46.1)	−79.9
	60 kPa at 5 cm	59.6±8.2	13.8	10.2±8.2	80.6	49.3 (33.2 to 65.5)	−82.9
Operator B	10 kPa at 1.5 cm	10.8±0.3	3.1	10.8±0.2	1.9	0 (−0.6 to 0.6)	0
	40 kPa at 1.5 cm	23.2±0.5	2.1	26.1±1.3	4.9	−2.9 (−3.9 to −2.0)	12.7
	60 kPa at 1.5 cm	34.4±1.1	3.3	35.9±1.5	4.1	−1.4 (−3.6 to 0.8)	4.2
	10 kPa at 5 cm	10.4±1.4	13.6	5.9±1.2	20.8	4.5 (1.7 to 7.3)	−43.4
	40 kPa at 5 cm	26.9±7.0	26.2	6.4±2.2	34.5	20.5 (6.7 to 34.3)	−76.2
	60 kPa at 5 cm	33.9±4.8	14.1	6.9±1.8	26.9	27.1 (17.7 to 36.5)	−79.8

(−) indicates failed to acquire shear wave elastography measurements.

SSI, Supersonic Imagine machine; GE, General Electric machine; SD, standard deviation; CV, coefficient of variation; 95% LOA, Bland-Altman 95% limits of agreement.

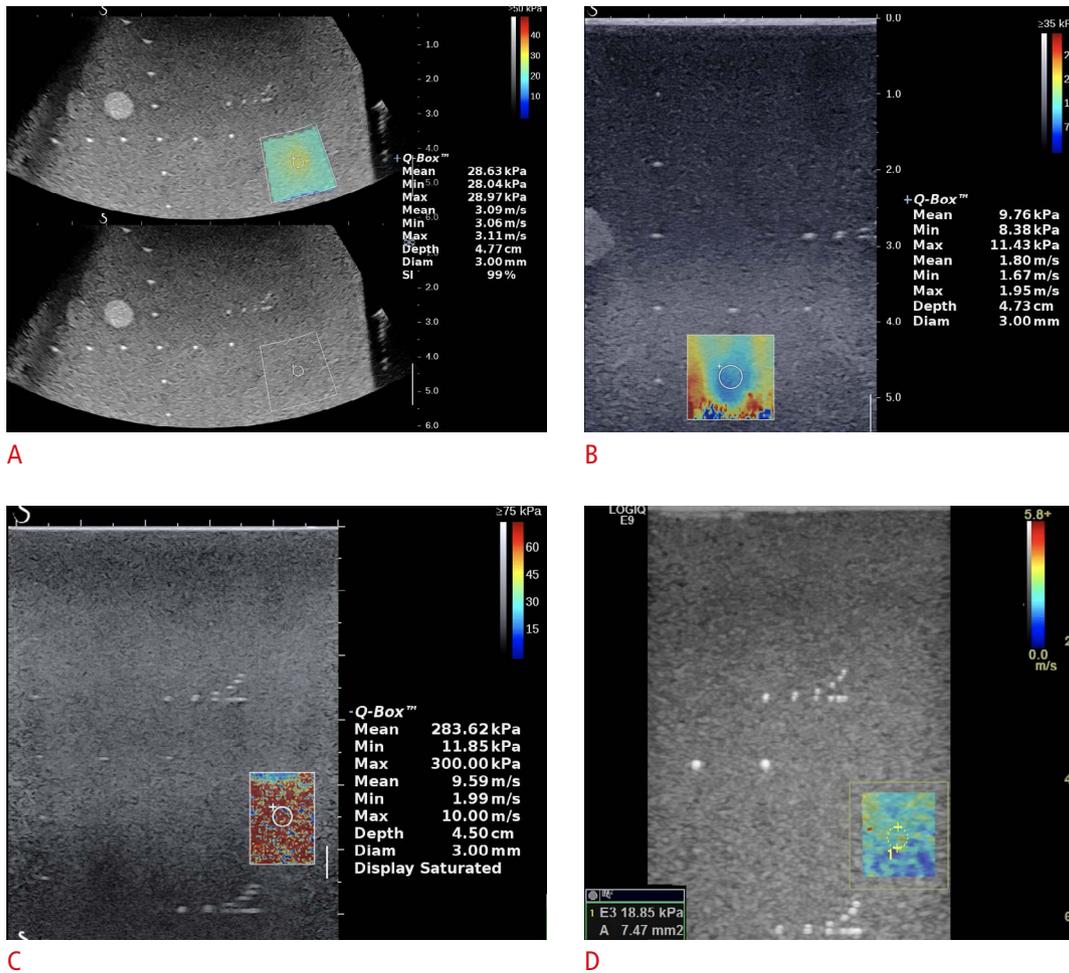


Fig. 2. Shear wave elastography (SWE) measurements obtained using the different systems and probes. All SWE images acquired at a depth of 5 cm. **A.** The image shows SuperSonic Imagine Aixplorer (SSI) probe 1 for the 60 kPa inclusion. **B.** The image shows SSI probe 2 for the 10 kPa inclusion. **C.** The image shows SSI probe 3 for the 40 kPa inclusion which demonstrates a failed SWE acquisition showing a saturation artifact. **D.** The image shows GE probe 9L-D for the 60 kPa inclusion.

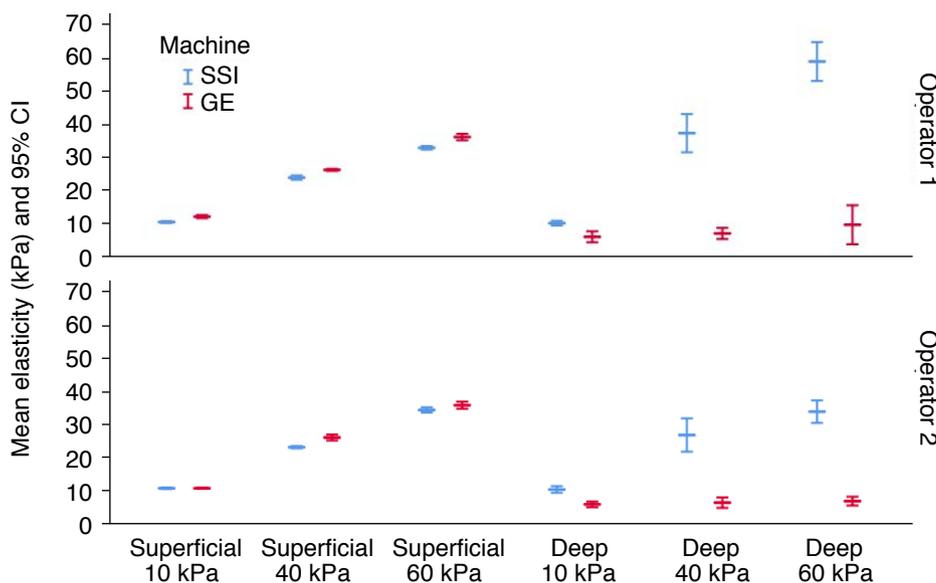


Fig. 3. Comparison of measurements between machines. The error bars represent mean elasticity and 95% confidence intervals (CIs) for both machines.

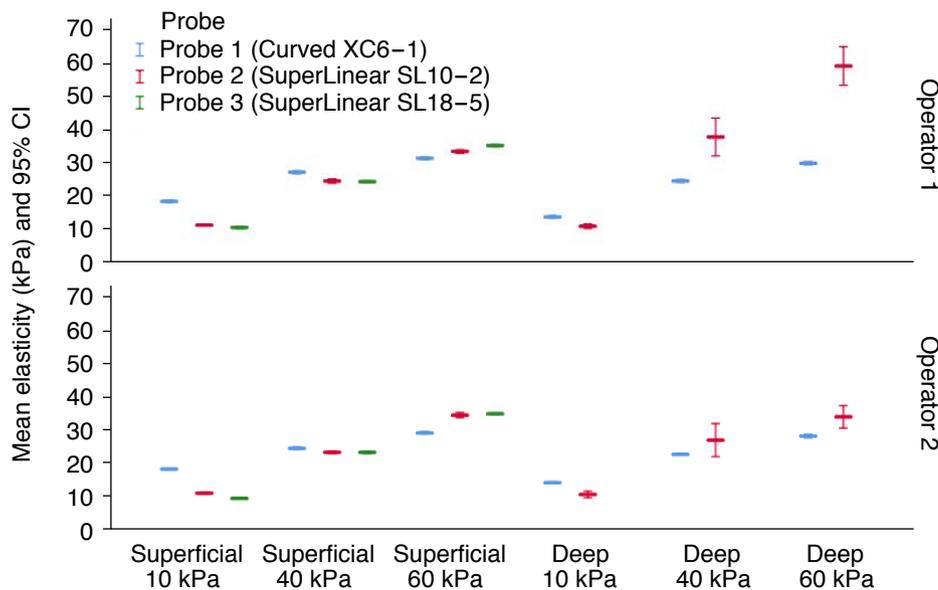


Fig. 4. Comparison of measurements among probes. The error bars represent mean elasticity and 95% confidence intervals (CIs) for the three probes.

presented in Table 2. The ICCs and the error bars for both operators suggested that the two machines only agreed for the superficial inclusions (Fig. 3). The ICC (95% CI) values of operators A and B for these inclusions were 0.99 (0.13 to 1.0) and 0.99 (0.85 to 1.0), respectively. The agreement of both operators for the deep inclusions was extremely poor. The ICCs (95% CIs) for operator A and operator B were 0.134 (−0.745 to 0.959) and 0.056 (−0.486 to 0.939), respectively. The GE machine failed to estimate the increasing stiffness in the deeper inclusions (Fig. 3). It also showed poorer results for measurement variation and inter-operator differences than the SSI device.

Comparison of SWE between the Probes

A summary of the results between probes is illustrated in Fig. 4. Probe 3 demonstrated the lowest overall variability. Probe 1 overestimated the stiffness of the 10 kPa inclusion, while all probes underestimated the stiffness of the 60 kPa inclusion (Table 1, Fig. 4). For the superficial inclusions, there was almost perfect agreement among the three probes for operators A and B (ICC [95% CI], 0.97 [0.79 to 0.99] and 0.97 [0.64 to 0.99], respectively). For all inclusions, there was substantial agreement among the three probes for operator A (0.68 [0.10 to 0.95]) and almost perfect agreement for operator B (0.97 [0.30 to 0.98]).

Effect of Depth on SWE Measurements

As shown in Table 1, both operators had low variability (CV) for the shallow inclusions on both machines. In contrast, the variability significantly increased as the depth increased (except for SSI probe 1). There was a significant difference between superficial and deep inclusions of the same elasticity in most cases ($P \leq 0.001$). The

only exception was for probe 2 on the SSI machine for the 10 kPa inclusion (operator A, $P=0.346$; operator B, $P=0.443$) and the 40 kPa and 60 kPa inclusions for operator B ($P=0.124$ and $P=0.782$, respectively).

Accuracy of the Measurements

The results demonstrated that the measured SWE values differed considerably from those reported by the phantoms' manufacturers. It was found that, on average, the mean SWE values of the stiffness of the 10 kPa inclusions were overestimated by 16%, while those for the 40 kPa and 60 kPa inclusions were underestimated by 42% and 48%, respectively. Only four of the 42 mean elasticity values were not significantly different from those reported by the manufacturer ($P > 0.05$)—the measurements made using probe 2 of 10.3, 10.4, 37.9, and 59.5 kPa (Table 1).

Discussion

To the authors' knowledge, no previously published research has simultaneously analyzed SWE measurements among operators, machines, probes, and imaging depths in a commercially available phantom. The current study found that SWE only seemed to be reproducible among operators, machines, and probes at shallow depths. However, the results suggest that the SWE measurements on both machines were overestimated for the low-stiffness 10 kPa inclusions and significantly underestimated for the higher-stiffness 40 kPa and 60 kPa inclusions. Whether the inclusion has a high or low stiffness did not significantly influence the results, which is in agreement with a previous study [13]. However, depth significantly impacted the reproducibility of the results.

The excellent within-operator and between-operator reproducibility of SWE in phantoms is in agreement with the previous studies [13,17,21,27–29]. Lee et al. [27] recently investigated the same SWE machines using a custom-made Zerdine phantom and reported almost perfect reproducibility ($ICC > 0.90$). Mun et al. [28] assessed the intra- and inter-observer reproducibility of SWE and found that the intra-observer reproducibility was almost perfect for three out of four operators. The low reproducibility of the last operator was attributed to his limited experience. This is in line with the present study, as the variability of operator A (who had limited experience) was higher than that of operator B.

Our results support the consensus in the literature that SWE measurements can be considerably different between different machines [11,13,15,17,19,27,30]. For example, Shin et al. [19] investigated three machines (Aixplorer, ACUSON S3000, and EPIQ 5) and reported inconsistent measurements using curved probes at various depths between 2 cm and 5 cm. Their reported percentage difference ($< 10\%$) between the machines was close to our percentage for superficial inclusions ($< 15\%$). Our high percentage disagreement for the deep inclusions could have been due to the failure of the GE machine to accurately measure the deep inclusions, rather than true disagreement. Thus, the 9L-D probe may not be optimal for imaging deep inclusions in clinical practice. The between-machine differences could be attributed to each system's patent-protected technology of generating and tracking waves. Hence, caution is needed when comparing SWE measurements across vendors in clinical practice.

To our knowledge, this study is the first to describe the SWE readings of three different probes on the SSI machine. The high frequency of probe 3 hindered the ability of its acoustic impulse and tracking waves to reach the deep inclusions (Fig. 2C). Similar findings on high-frequency probes have been reported previously [19]. However, the CVs of probe 3 for the superficial inclusions were lower than the values obtained with the other probes. We found that the low-frequency convex probe (probe 1) maintained good variability ($< 5\%$) regardless of depth. However, it significantly overestimated the stiffness of superficial inclusions. These findings correspond to what Chang et al. [12] have reported. In contrast, Shin et al. [19] reported similar variability between probes.

The measurement depth had a remarkable impact on the reproducibility of the readings and differences in actual elasticity measurements. The literature concurs regarding this finding [12,17,19,31,32]. The two SWE studies by the QIBA also demonstrated an increase in measurement bias from depths of 3 to 7 cm [9,33]. This was attributed to the attenuation of acoustic push pulses and tracking waves as they traverse deeper, which results in a poor signal-to-noise ratio [33]. Two previous studies

reported a limited depth of ≤ 4 cm in the Siemens ACUSON machine [19,34]. This again implies that ultrasound machines and probes have different capabilities. The good agreement and reproducibility found for the superficial inclusions may suggest that the SSI and GE readings can be comparable in superficial organs. However, we strongly recommend against reading the measurements interchangeably in deep structures when using different systems or acquisition protocols.

In terms of measurement bias, our results suggest a significant difference, ranging on average from 16% to 48%, between the measured SWE values and the theoretical values provided by the phantom manufacturer. This finding is consistent across the majority of studies conducted on CIRS phantoms [11,12,14,15,17–21]. The only exception was the study of Mulabecirovic et al. [13], who used the 049 model and reported accurate measurements using the GE machine (9L-D probe) and SSI machine (probe 2). However, their measurements using the SSI (probe 1) and Philips machine (C5-1 probe) suffered the same biases observed in our study. This discrepancy has been surprisingly seldom discussed despite its importance. The underlying reason for the discrepancy is not clear, although it could potentially be due to the different calibration techniques used when building the elasticity phantoms. Indeed, the consistent biases across different vendors in numerous studies cast doubt on the integrity of the phantom elasticity values. An additional plausible explanation is that the phantoms' elastic integrity may change over time through inappropriate storage or aging of the hydrogel material. Long et al. [35] recently suggested using elasticity phantoms for acceptance testing only (i.e., to verify that the new SWE machine is acquiring the same readings as another similar machine from the same manufacturer).

It should be noted that in early 2020, CIRS adjusted their elastography phantoms' reference values (models 049 and 049A) by decreasing them by more than 10% [36]. This major change could render previous publications obsolete. The results from the tested general-purpose QA phantom provided similar results to what has been reported for dedicated elastography phantoms. This may support its use without the need to purchase a dedicated phantom. Our results provide crucial information about the expected results on two machines and multiple probes between two operators. These reported data are highly relevant for the growing number of SWE-related tasks for medical physicists, such as acceptance testing and routine QA assessment.

Our results have several clinical implications. First, clinical follow-up SWE scans should be performed using the same SWE system and probe used in the baseline assessment. This is particularly important in departments with multiple SWE-enabled machines or in multi-center studies. Moreover, operators should be aware of the

effect of depth, especially at 5 cm and deeper, on the reliability of the measurements. It should be noted that there is no universally acceptable depth cutoff for reliable measurements, as it largely depends on the probe's frequency and the attenuation of the tissue [32]. Nevertheless, for linear probes with a frequency of 10 MHz, we recommend not exceeding a depth of 4–5 cm. This recommendation corresponds with the previous literature [19,32,34,37].

The study has several limitations. Phantoms, in general, present an idealized testing environment and do not necessarily mimic clinical cases. However, such *in vitro* investigations are excellent for assessing technical performance, as they minimize potential *in vivo* confounders (e.g., respiration, surrounding tissue stiffness, and aging). To accurately evaluate the role of experience, a large number of operators with different levels of experience would be required. This was unfortunately not possible for our study. Future research should replicate the results on the same phantom and compare them to those obtained using a dedicated SWE phantom on more than two machines. This will shed light on the merit of using this general-purpose phantom model for various applications, including SWE. Moreover, depth should be evaluated as a continuous variable. This could be achieved by testing a phantom that has multiple inclusions situated vertically to accurately estimate the SWE trade-off with depth. Further studies should test the results on different machines and develop means of reliably comparing measurements across different systems and probes.

In conclusion, this phantom study showed that the SWE technique seemed to be reproducible within and between operators and machines only for superficial inclusions (1.5 cm). The SWE readings depended fundamentally on the depth of acquisition, the type of machine, and the probe. Caution is warranted when comparing results across different machines or probes, especially for deep regions. To ensure reproducible results, prospective elasticity scans should be performed using the same system and settings. Finally, our SWE results demonstrated a large difference compared to the manufacturer's reported phantom elasticity values, which is consistent with previous studies.

ORCID: Abdulaziz Ibrahim Alrashed: <https://orcid.org/0000-0003-4219-0482>;
Abdulrahman M. Alfuraih: <https://orcid.org/0000-0002-4655-7248>

Author Contributions

Conceptualization: Alrashed AI, Alfuraih AM. Data acquisition: Alrashed AI, Alfuraih AM. Data analysis or interpretation: Alrashed AI, Alfuraih AM. Drafting of the manuscript: Alrashed AI, Alfuraih AM. Critical revision of the manuscript: Alrashed AI, Alfuraih AM. Approval of the final version of the manuscript: all authors.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Nowicki A, Dobruch-Sobczak K. Introduction to ultrasound elastography. *J Ultrason* 2016;16:113-124.
2. Bercoff J. ShearWave elastography. *Supersonic Imagine: white paper*. Aix en Provence: Supersonic Imagine, 2008.
3. Shiina T, Nightingale KR, Palmeri ML, Hall TJ, Bamber JC, Barr RG, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology. *Ultrasound Med Biol* 2015;41:1126-1147.
4. Sigrüst RM, Liao J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: review of techniques and clinical applications. *Theranostics* 2017;7:1303-1329.
5. Davis LC, Baumer TG, Bey MJ, Holsbeeck MV. Clinical utilization of shear wave elastography in the musculoskeletal system. *Ultrasonography* 2019;38:2-12.
6. Yoo J, Lee JM, Joo I, Yoon JH. Assessment of liver fibrosis using 2-dimensional shear wave elastography: a prospective study of intra- and inter-observer repeatability and comparison with point shear wave elastography. *Ultrasonography* 2020;39:52-59.
7. Ryu H, Ahn SJ, Yoon JH, Lee JM. Reproducibility of liver stiffness measurements made with two different 2-dimensional shear wave elastography systems using the comb-push technique. *Ultrasonography* 2019;38:246-254.
8. Ryu H, Ahn SJ, Yoon JH, Lee JM. Inter-platform reproducibility of liver stiffness measured with two different point shear wave elastography techniques and 2-dimensional shear wave elastography using the comb-push technique. *Ultrasonography* 2019;38:345-354.
9. Hall TJ, Milkowski A, Garra B, Carson P, Palmeri M, Nightingale K, et al. RSNA/QIBA: shear wave speed as a biomarker for liver fibrosis staging. In: 2013 IEEE International Ultrasonics Symposium (IUS); 2013 Jul 21-25; Prague, Czech Republic. New York: IEEE, 2013;397-400.
10. Palmeri M. RSNA/QIBA efforts to standardize shear wave speed as a biomarker for liver fibrosis staging. *Ultrasound Med Biol* 2019;45(Suppl 1):S24.
11. Safonov DV, Rykhtik PI, Shatokhina IV, Romanov SV, Gurbatov SN, Demin IY. Shear wave elastography: comparing the accuracy of ultrasound scanners using calibrated phantoms in experiment. *Med Technol Med* 2017;9:51-58.
12. Chang S, Kim MJ, Kim J, Lee MJ. Variability of shear wave velocity using different frequencies in acoustic radiation force impulse (ARFI) elastography: a phantom and normal liver study. *Ultraschall Med* 2013;34:260-265.
13. Mulabecirovic A, Vesterhus M, Gilja OH, Havre RF. In vitro

- comparison of five different elastography systems for clinical applications, using strain and shear wave technology. *Ultrasound Med Biol* 2016;42:2572-2588.
14. Seliger G, Chaoui K, Kunze C, Dridi Y, Jenderka KV, Wienke A, et al. Intra- and inter-observer variation and accuracy using different shear wave elastography methods to assess circumscribed objects: a phantom study. *Med Ultrason* 2017;19:357-365.
 15. Franchi-Abella S, Elie C, Correas JM. Performances and limitations of several ultrasound-based elastography techniques: a phantom study. *Ultrasound Med Biol* 2017;43:2402-2415.
 16. Hollerlieth K, Gassmann B, Wagenpfeil S, Kemmner S, Heemann U, Stock KF. Does standoff material affect acoustic radiation force impulse elastography? A preclinical study of a modified elastography phantom. *Ultrasonography* 2018;37:140-148.
 17. Hwang J, Yoon HM, Jung AY, Lee JS, Cho YA. Comparison of 2-dimensional shear wave elastographic measurements using ElastQ Imaging and SuperSonic Shear Imaging: phantom study and clinical pilot study. *J Ultrasound Med* 2020;39:311-321.
 18. Carlsen JF, Pedersen MR, Ewertsen C, Saftoiu A, Lonn L, Rafaelsen SR, et al. A comparative study of strain and shear-wave elastography in an elasticity phantom. *AJR Am J Roentgenol* 2015;204:W236-W242.
 19. Shin HJ, Kim MJ, Kim HY, Roh YH, Lee MJ. Comparison of shear wave velocities on ultrasound elastography between different machines, transducers, and acquisition depths: a phantom study. *Eur Radiol* 2016;26:3361-3367.
 20. Ruby L, Mutschler T, Martini K, Klingmuller V, Frauenfelder T, Rominger MB, et al. Which confounders have the largest impact in shear wave elastography of muscle and how can they be minimized? An elasticity phantom, ex vivo porcine muscle and volunteer study using a commercially available system. *Ultrasound Med Biol* 2019;45:2591-2611.
 21. Mulabecirovic A, Mjelle AB, Gilja OH, Vesterhus M, Havre RF. Repeatability of shear wave elastography in liver fibrosis phantoms-Evaluation of five different systems. *PLoS One* 2018;13:e0189671.
 22. Computerized Imaging Reference Systems. Multi-purpose, multi-tissue ultrasound phantom: model 040GSE [Internet]. Norfolk, VA: Computerized Imaging Reference Systems, 2013 [cited 2020 Jan 5]. Available from: <http://www.cirsinc.com/wp-content/uploads/2019/04/040GSE-DS-120418.pdf>.
 23. GE Healthcare. LOGIQ E9 shear wave elastography white paper [Internet]. Wauwatosa, WI: GE Healthcare, 2014 [cited 2020 Apr 8]. Available from: https://www3.gehealthcare.com/~/media/rsna-2016-press-kit-assets/press%20releases/ultrasound/global%20shear%20wave%20whitepaper_october%202014.pdf.
 24. Brown CE. Applied multivariate statistics in geohydrology and related sciences. Berlin: Springer, 1998;155-157.
 25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
 26. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015;25:141-151.
 27. Lee SM, Chang W, Kang HJ, Ahn SJ, Lee JH, Lee JM. Comparison of four different shear wave elastography platforms according to abdominal wall thickness in liver fibrosis evaluation: a phantom study. *Med Ultrason* 2019;21:22-29.
 28. Mun HS, Choi SH, Kook SH, Choi Y, Jeong WK, Kim Y. Validation of intra- and interobserver reproducibility of shearwave elastography: Phantom study. *Ultrasonics* 2013;53:1039-1043.
 29. Dillman JR, Chen S, Davenport MS, Zhao H, Urban MW, Song P, et al. Superficial ultrasound shear wave speed measurements in soft and hard elasticity phantoms: repeatability and reproducibility using two ultrasound systems. *Pediatr Radiol* 2015;45:376-385.
 30. Alfuraih AM, O'Connor P, Tan AL, Hensor E, Emery P, Wakefield RJ. An investigation into the variability between different shear wave elastography systems in muscle. *Med Ultrason* 2017;19:392-400.
 31. Hwang JA, Jeong WK, Song KD, Kang KA, Lim HK. 2-D shear wave elastography for focal lesions in liver phantoms: effects of background stiffness, depth and size of focal lesions on stiffness measurement. *Ultrasound Med Biol* 2019;45:3261-3268.
 32. Alfuraih AM, O'Connor P, Hensor E, Tan AL, Emery P, Wakefield RJ. The effect of unit, depth, and probe load on the reliability of muscle shear wave elastography: Variables affecting reliability of SWE. *J Clin Ultrasound* 2018;46:108-115.
 33. Palmeri M, Nightingale K, Fielding S, Rouze N, Deng Y, Lynch T, et al. RSNA QIBA ultrasound shear wave speed Phase II phantom study in viscoelastic media. In: 2015 IEEE International Ultrasonics Symposium (IUS); 2015 Oct 21-24; Taipei, Taiwan. New York: New York, 2015;1-4.
 34. Yamanaka N, Kaminuma C, Taketomi-Takahashi A, Tsushima Y. Reliable measurement by virtual touch tissue quantification with acoustic radiation force impulse imaging: phantom study. *J Ultrasound Med* 2012;31:1239-1244.
 35. Long Z, Tradup DJ, Song P, Stekel SF, Chen S, Glazebrook KN, et al. Clinical acceptance testing and scanner comparison of ultrasound shear wave elastography. *J Appl Clin Med Phys* 2018;19:336-342.
 36. Computerized Imaging Reference Systems. Elasticity QA Phantom: models 049 & 049A [Internet]. Norfolk, VA: Computerized Imaging Reference Systems, 2020 [cited 2020 Jan 5]. Available from: <http://www.cirsinc.com/wp-content/uploads/2020/03/049-049-DS-030920.pdf>.
 37. Cosgrove D, Barr R, Bojunga J, Cantisani V, Chammas MC, Dighe M, et al. WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. *Ultrasound Med Biol* 2017;43:4-26.