Repeatability and Agreement of Shear Wave Speed Measurements in Phantoms and Human Livers Across 6 Ultrasound 2-Dimensional Shear Wave Elastography Systems

Leah A. Gilligan, MD,† Andrew T. Trout, MD,*‡*
Paula Bennett, RDMS, RT(R), and Jonathan R. Dillman, MD, MSc*‡*

Objectives: Ultrasound shear wave elastography (SWE) is an imaging technique that quantifies liver stiffness. However, comparison data across newest ultrasound systems are sparse. The purpose of this study was to assess repeatability and agreement of shear wave speeds (SWSs) across 6 ultrasound 2-dimensional (2D) SWE systems.

Materials and Methods: This cross-sectional, Health Insurance Portability and Accountability Act–compliant study received institutional review board approval. Written informed consent was obtained. Serial 2D SWE examinations were performed with 6 ultrasound systems (Aplio i800, Canon Medical Systems; LOGIQ E10, GE Healthcare; Resona 7, Mindray North America; EPIQ Elite, Philips Healthcare; ACUSON Sequoia, Siemens Medical Solutions; and Aixplorer MACH 30, SuperSonic Imagine) on 4 elastic phantoms (SWS range, 0.82–3.51 m/s) and on livers of 24 adults (healthy volunteers and patients with known liver stiffening). Participants were imaged 2 times per ultrasound system, with 90 to 120 minutes between examinations. Median SWS was calculated from separately acquired SWS measurements per examination (40 phantom measurements and 10 liver measurements per examination).

Results: Overall intraclass correlation coefficient (ICC) for intersystem agreement of median SWS across systems was 0.99 (95% confidence interval, 0.96–1.0) in phantoms and 0.66 to 0.69 (95% confidence interval, 0.47–0.84) in humans across systems. Means of median SWS measurements in humans ranged from 1.24 to 1.56 m/s. Average individual subject-level variance (interquartile range/median SWS) across all examinations was 0.07, with an average coefficient of variation of 6.0%. Pairwise ICCs for intersystem agreement in subjects across systems ranged from 0.41 to 0.91; test-retest repeatability in subjects was excellent for all systems, with ICCs ranging from 0.87 to 0.97.

Conclusions: There is good to excellent intersystem agreement of measured SWS in elastic phantoms and in vivo livers across 6 ultrasound 2D SWE systems. Test-retest repeatability was excellent for all systems.

Key Words: imaging biomarker, liver fibrosis, comparative research, variability, diagnostic

(*Invest Radiol* 2020;00: 00–00)

Ultrasound shear wave elastography (SWE) is a technology that measures the speed of shear waves as they propagate through a tissue of interest upon the application of an acoustic radiation force.[1,2] Shear wave elastography has been used to evaluate various human tissues, including the liver.[3–9] Liver shear wave speed (SWS) estimates Young modulus, a physical measurement of tissue stiffness, which has been shown to positively correlate with histopathologic hepatic fibrosis.[10–15] Importantly, hepatic fibrosis is the common pathophysiological pathway of chronic liver disease, and the stage of hepatic fibrosis impacts patient management and prognosis.[16] Although liver biopsy with histopathologic evaluation remains the diagnostic reference standard for diagnosing and staging fibrosis, ultrasound SWE offers the advantages of portability, cost-efficiency, increased sampling, and noninvasiveness.[10,17,18] Ultrasound SWE technology is now commercially available on all major ultrasound systems used in radiology departments in the United States.

The diagnostic performance of ultrasound 2-dimensional (2D) SWE for predicting the degree of histopathologic liver fibrosis is generally good to excellent, with areas under the receiver operating characteristic curve ranging from 0.85 to 0.94.[19,20] Diagnostic performance of SWE is best for discriminating no or mild from moderate or severe liver fibrosis and decreases when discriminating specific histopathologic stages. This imperfect performance can be attributed in part to sources of measurement variability, including hardware/software differences between ultrasound systems, operator-related factors (eg, measurement location/depth, breath-holding instructions), disease-specific factors (eg, sampling errors related to disease heterogeneity), and patient-specific factors (eg, body habitus, nil per os status).[10,12]

To date, a small number of studies have measured the variability, including intersystem absolute agreement and test-retest repeatability, of SWS measurements in phantoms (elastic and viscoelastic) and human livers across multiple ultrasound systems.[21–27] Notably, most of these comparisons are of point SWE systems (which provide small, typically fixed square or rectangular regions of interest [ROIs]), as opposed to the newer 2D SWE systems (which provide color images [elastograms] of tissue stiffness and that typically have larger, adjustable ROIs). As of early 2019, the major ultrasound manufacturers in the United States have all released systems that are 2D SWE capable (Supplemental Table 1, http://links.lww.com/RLI/A499). To date, no study has assessed the agreement and repeatability of SWS measurements by 2D SWE in elastic phantoms and in vivo human livers across these recently approved ultrasound platforms.

The purpose of our study was to assess the intersystem agreement and test-retest repeatability of 2D SWE SWS measurements in elastic phantoms and in vivo human livers across 6 ultrasound systems, with each included ultrasound system representing the most up-to-date US Food and Drug Administration–approved hardware and software. We hypothesized that intersystem agreement and test-retest repeatability of SWS measurements in phantoms and human livers would be good to excellent and comparable to that described for magnetic resonance elastography (MRE).

MATERIALS AND METHODS

This cross-sectional study was institutional review board (IRB) approved and Health Insurance Portability and Accountability Act (United States) compliant. Written informed consent was obtained

Received for publication August 4, 2019; and accepted for publication, after revision, September 6, 2019.
From the *Department of Radiology, Cincinnati Children's Hospital Medical Center; †Department of Graduate Medical Education, Mount Carmel Health System, Columbus; and ‡Department of Radiology, University of Cincinnati College of Medicine, Ohio.
Conflicts of interest and sources of funding: A.T.T. and J.R.D. received unrelated grant support from Canon Medical Systems and Siemens Healthineers.
Correspondence to: Jonathan R. Dillman, MD, MSc, Department of Radiology, MLC 5031, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH 45229. E-mail: jonathan.dillman@cchmc.org.
Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.investigativeradiology.com).

ISSN: 0020-9996/20/0000–0000
DOI: 10.1097/RLI.0000000000000627

from all participants. Each of the following ultrasound system manufacturers provided support for this investigation in the form of an ultrasound system with 2D SWE (commercial system name in parentheticals) and an expert operator (applications specialist): Canon Medical Systems USA (Aplio i800), GE Healthcare (LOGIQ E10), Mindray North America (Resona 7), Philips Healthcare (EPIQ Elite), Siemens Medical Solutions USA (ACUSON Sequoia), and Supersonic Imagine (Aixplorer MACH 30) (Supplemental Table 1, http://links.lww.com/RLI/A499).

Overall Study Timeline

Ultrasound systems were delivered to the Cincinnati Children's Hospital Medical Center in March 2019. All phantom imaging was performed as described below on a single day (Friday) in Department of Radiology research space. Human imaging was performed over the next 2 days (Saturday and Sunday) in 6 adjacent research rooms of an institutionally designated clinical research space.

2D SWE Protocol—Phantom Imaging

Ultrasound 2D SWE examinations were performed on 4 elastic (Zerdine gel) phantoms with known stiffnesses (0.82, 1.52, 2.48, and 3.51 m/s) (Model 039; Computerized Imaging Reference Systems, Inc, Norfolk, VA) with the 6 ultrasound systems under investigation. The stiffness of each phantom had been characterized by the Nightingale laboratory at Duke University on January 14, 2019, with a Verasonics Vantage ultrasound system using previously reported methodology.[21,28] Phantom imaging was performed using curvilinear low-frequency transducers that are commonly used for liver imaging (Supplemental Table 1, http://links.lww.com/RLI/A499). Transducers were coupled to phantoms using a 45 parts per thousand saline solution (45 g of sodium chloride in 1 L of deionized water) at room temperature.[21,29]

For a given ultrasound system, 2 operators acquired a total of 40 measurements from each phantom (20 measurements per operator), with each measurement obtained as a separate image acquisition. The operator was changed every 10 measurements. Operator 1 was the expert operator provided by the manufacturer for each ultrasound system, whereas operator 2 was the same individual for all 6 ultrasound systems. Specifically, operator 2 was a pediatric sonographer from our institution with 5 years of clinical and research ultrasound 2D SWE experience (P.B.). Shear wave speed measurements were acquired from the centers of the respective phantoms, using a circular (1 cm diameter) ROI centered 4 cm below the phantom surface.

2D SWE Protocol—Human Imaging

On the basis of an a priori sample size calculation ($\alpha = 0.05$; $\beta = 0.2$; correlation, $H_0 = 0.4$, $H_a = 0.8$), 24 adults, 18 years or older, were recruited to undergo serial liver 2D SWE examinations across the 6 ultrasound systems under investigation. Fourteen participants were enrolled via an IRB-approved e-mail sent to employees of the Cincinnati Children's Hospital Medical Center Department of Radiology. Participants were recruited in order of response, while excluding individuals with a known personal history of liver disease. The remaining 10 participants were patients with known liver stiffening based on previous ultrasound 2D SWE (median SWS >1.8 m/s) or MRE (shear stiffness >2.7 kPa) performed at our institution between March 2014 and January 2019. This latter cohort of participants was identified through a query of Department of Radiology records (Illuminate InSight; Softek Illuminate, Overland Park, KS) and was recruited consecutively via telephone using an IRB-approved script. Additional exclusion criteria included the inability to provide informed consent or fast before imaging. Participants were instructed to fast (nil per os) for at least 4 hours before, and throughout, research 2D SWE imaging.

Participants were placed into 4 groups of 6 individuals and imaged over a single weekend, as mentioned above. Participants within a group underwent serial liver ultrasound 2D SWE examinations over a 4-hour period with the same 6 ultrasound systems (hardware and software) and transducers used for phantom imaging. Examinations were performed by expert operators provided by the ultrasound system manufacturers. All 6 participants in a group underwent 2D SWE simultaneously, each imaged with a different ultrasound system. Participants then rotated through the remaining 5 ultrasound systems in series, over 15-minute time blocks. This entire rotation was repeated once more so that each participant was scanned twice by each ultrasound system and underwent a total of 12 separate 2D SWE examinations. First and second examinations for a given ultrasound system were separated by approximately 90 to 120 minutes, with both examinations performed by the same operator.

All research ultrasound 2D SWE examinations were performed using a standardized protocol. Participants were positioned supine on an examination table with the right arm abducted and hand placed above or under the head. The liver was accessed using a right intercostal approach. Patients were instructed to suspend respiration during SWS measurement, avoiding deep inspiration or expiration. Using the respective manufacturers' 2D SWE approaches, 10 consecutive SWS measurements were acquired from the central right hepatic lobe using a circular (1 cm diameter) ROI placed at least 1.5 cm deep but no more than 5 cm deep to the liver capsule, while avoiding visible blood vessels and areas of artifact (Fig. 1). Each SWS measurement was obtained from a separate image acquisition (and breath hold). No SWS measurements were excluded for any subject or ultrasound system.

Demographic and anthropometric data, including age, sex, liver-related medical history, height, and weight, were collected from all participants on the day of imaging by a single investigator (L.A.G.).

Statistical Analysis

Unless otherwise specified, continuous measures were summarized as means and standard deviations. Categorical measures were summarized as counts and percentages. Median SWS (in m/s), interquartile range (IQR), IQR/median SWS, and coefficient of variation of SWS were calculated for all phantom (n = 40 measurements) and human (n = 10 measurements) ultrasound 2D SWE examinations. A single ultrasound system (Resona 7; Mindray North America) only reported liver stiffness in kPa (Young modulus). Measurements from this system were converted to m/s by dividing each measurement by 3 and then taking the square root (Young modulus [kPa] = $3 \times SWS^2$, assuming an isotropic tissue density of 1 g/mL).[30,31]

Intraclass correlation coefficients (ICCs) were used to assess the intersystem absolute agreement of median SWS measurements across ultrasound systems for (1) all phantom examinations (n = 24, across 6 ultrasound systems and 4 phantoms), (2) first participant examinations (n = 144, across 6 ultrasound systems and 24 participants), and (3) second (repeat) participant examinations (n = 144, across 6 ultrasound systems and 24 participants). Intraclass correlation coefficients and Lin concordance coefficients (r_c) were used to assess the agreement between median SWS measurements from first participant examinations for all possible paired ultrasound system combinations. To assess test-retest repeatability of SWS in human livers, the mean difference in median SWS measurements, ICCs, and Lin's concordance coefficients were calculated using first and second examination results. Mann-Whitney U (MWU) and F tests were used to measure differences and compare variances, respectively, in SWS measurements obtained in phantoms between operators.

P values < 0.05 were considered statistically significant, unless otherwise noted. Concordance (r_c) was classified with the following definitions: lower than 0.90, poor; 0.90 to 0.95, moderate; greater than 0.95 to 0.99, substantial; and greater than 0.99, almost perfect.[32] Absolute agreement (ICC) was classified with the following definitions: 0 to 0.39, poor; 0.40 to 0.59, fair; 0.60 to 0.74, good; and 0.75 to 1.0, excellent.[33] Moreover, 95% confidence intervals (CIs) were calculated, as

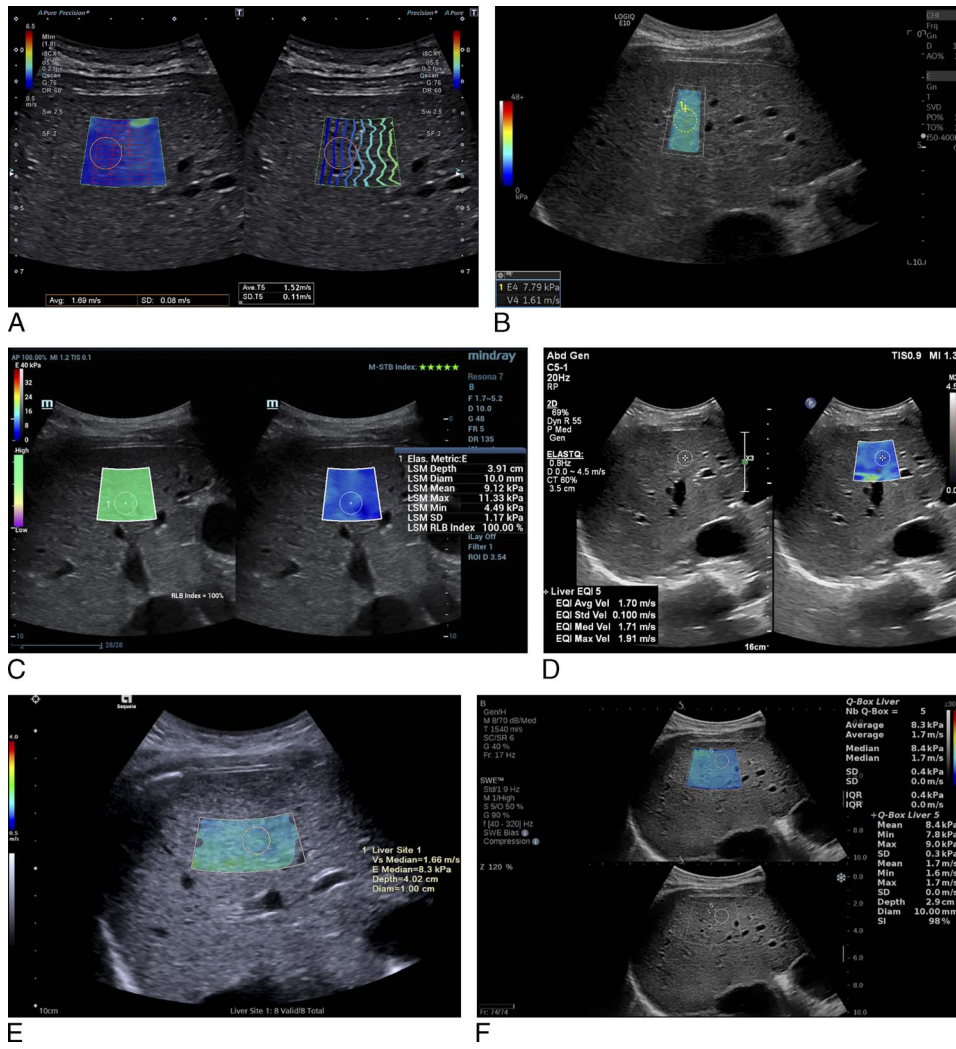

FIGURE 1. Examples of in vivo human liver imaging for all 6 ultrasound systems: Canon Medical Solutions (a), GE Healthcare (b), Mindray North America (c), Philips Healthcare (d), Siemens Medical Solutions (e), and SuperSonic Imagine (f). Images are from first examinations in a 20-year-old woman with known liver stiffening related to autoimmune sclerosing cholangitis. The mean of median shear wave speed values across the 12 examinations in this patient was 1.68 m/s.

appropriate. All statistical analyses were performed using MedCalc Statistical Software version 19 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2019).

RESULTS

Phantom Imaging

Ultrasound 2D SWE summary measures (n = 40 measurements) for each of the 4 elastic phantoms using each of the 6 ultrasound systems, including median SWS, IQR, IQR/median SWS, and coefficient of variation, are presented in Table 1. Tukey box plots (Fig. 2) depict the distribution of SWS measurements for each ultrasound system and phantom. Across ultrasound systems, median SWS ranged from 0.84 to 0.90 m/s for the 0.82 m/s phantom, from 1.48 to 1.66 m/s for the 1.52 m/s phantom, from 2.31 to 2.84 m/s for the 2.48 m/s phantom, and from 3.23 to 3.63 m/s for the 3.51 m/s phantom. Overall, there was excellent intersystem absolute agreement of median SWS values across the 4 phantoms and 6 ultrasound systems (ICC, 0.99; 95% CI, 0.96–1.0).

Median SWS values for operators 1 and 2 for each elasticity phantom and ultrasound system are reported in Table 2. For most phantoms and ultrasound systems, there was no difference in median SWS or SWS variance between operators. There were 5 (of 24) phantom-ultrasound system combinations that showed significant differences in median SWS between operators by MWU tests following Bonferroni correction, with differences in median SWS ranging from 0.01 to 0.11 m/s. There were also 5 (of 24) phantom-ultrasound system combinations that showed significant differences in SWS variance between operations by F tests following Bonferroni correction.

Human Imaging

Of 24 participants (50%), 12 were women. Mean age was 28 ± 11 years (range, 18–58 years). Ten participants (42%) had a history of elevated liver stiffness on previous 2D SWE or MRE examinations, 3 by ultrasound (mean SWS, 2.07 ± 0.24 m/s) and 7 by MRE (mean shear stiffness, 3.66 ± 0.65 kPa). Causes of liver stiffening were diverse and included Fontan circulation (n = 3), autoimmune hepatitis (n = 2), primary sclerosing cholangitis (n = 2), nonalcoholic fatty liver disease (n = 1), alpha-1 antitrypsin deficiency (n = 1), and Fanconi anemia (n = 1). Mean participant height, weight, and body mass index were 172 ± 9 cm, 79.2 ± 18.1 kg, and 26.7 ± 4.3 kg/m^2, respectively. Of note,

TABLE 1. Summary of Shear Wave Speed Measurements for 4 Elastic Phantoms Using 6 Ultrasound Systems and 2D Shear Wave Elastography

	Phantom 1 (0.82 m/s)	Phantom 2 (1.52 m/s)	Phantom 3 (2.48 m/s)	Phantom 4 (3.51 m/s)
Aplio i800 (Canon)	0.85	1.52	2.51	3.63
	(0.06)	(0.09)	(0.16)	(0.36)
	[0.07]	[0.06]	[0.06]	[0.10]
	<5.5>	<4.3>	<5.7>	<9.2>
LOGIQ E10 (GE Healthcare)	0.87	1.48	2.31	3.23
	(0.01)	(0.02)	(0.17)	(0.32)
	[0.01]	[0.01]	[0.07]	[0.10]
	<0.9>	<1.7>	<4.6>	<5.6>
Resona 7 (Mindray)	0.86	1.53	2.52	3.39
	(0.01)	(0.02)	(0.04)	(0.04)
	[0.01]	[0.01]	[0.01]	[0.01]
	<0.8>	<1.0>	<1.6>	<1.3>
EPIQ Elite (Philips)	0.84	1.66	2.84	3.57
	(0.02)	(0.06)	(0.11)	(0.32)
	[0.03]	[0.04]	[0.04]	[0.09]
	<3.4>	<2.2>	<3.5>	<6.6>
ACUSON Sequoia (Siemens)	0.85	1.57	2.63	3.40
	(0.03)	(0.05)	(0.10)	(0.14)
	[0.04]	[0.03]	[0.04]	[0.04]
	<2.7>	<2.0>	<2.8>	<2.7>
Aixplorer MACH 30 (SuperSonic Imagine)*	0.9	1.5	2.5	3.5
	(0.0)	(0.0)	(0.1)	(0.0)
	[0.0]	(0.0)	[0.0]	[0.0]
	<0.0>	<2.0>	<2.0>	<1.2>
Mean of Medians (n = 6)	0.86	1.54	2.55	3.45
Mean coefficient of variation (%) between systems	2.2	2.2	3.4	4.4

Median of 40 measurements (20 each by 2 operators) in m/s, IQR in parentheses, IQR/median in brackets, and coefficient of variation in percentage between < >.

*Shear wave speed measurements only have a single decimal place.

Abbreviation: IQR, interquartile range.

upon post hoc review of imaging acquired for this study by 2 radiologists (J.R.D, A.T.T), 5 of 14 (36%) of the healthy participants had overt, previously unknown hepatic steatosis based on gray-scale ultrasound imaging.

Ultrasound 2D SWE summary measures for each of the 6 ultrasound systems for the first and second human examinations are presented in Table 3, including the means of median SWS, IQR, IQR/median SWS, and coefficient of variation across all 24 participants. The overall mean of median SWS was 1.46 and 1.43 m/s for the first and second examinations, respectively. Intersystem means of median SWS ranged from 1.24 to 1.56 m/s (including both the first and second examinations). Individual participant median SWS values of the first examinations are presented in Figure 3. Tukey box plots (Fig. 4) depict the distribution of SWS measurements for each ultrasound system for the first and second examinations. Mean IQR/median SWS across all ultrasound systems was 0.07 (including the first and second examinations), and overall mean coefficient of variation across all ultrasound systems was 6.0%. There was good intersystem absolute agreement of median SWS values across the 24 participants and 6 ultrasound systems on both the first (ICC, 0.66; 95% CI, 0.47–0.81) and second (ICC, 0.69; 95% CI, 0.50–0.84) examinations.

Pairwise assessment of intersystem median SWS absolute agreement using ICCs and first examinations (n = 24) was fair to excellent (ICC range, 0.41–0.91), while evaluation of agreement using Lin's concordance coefficients was mostly poor (r_c range, 0.40–0.90) (Table 4). There was excellent intrasystem test-retest repeatability when comparing SWS measurements from the first and second examinations using ICCs (ICC range, 0.87–0.97). However, test-retest repeatability was considered poor to substantial using Lin's concordance coefficients (r_c range, 0.86–0.97) (Table 3, Fig. 5).

DISCUSSION

Multiple studies have evaluated the diagnostic performance of ultrasound SWE technology, including both point and 2D SWE methods, for the assessment of liver fibrosis and have found it to be a valid imaging biomarker for detection and, to some degree, staging.[15,17,19,34] However, despite such literature, the reproducibility of SWS measurements between ultrasound systems (intersystem agreement) and test-rest repeatability remain topics of interest. This is particularly true as there are multiple ultrasound systems that have only recently become commercially available, all of which are 2D SWE-capable. Although data regarding agreement and repeatability exist for previous generation ultrasound systems and SWE methods, there is a paucity of data regarding the degree of SWS measurement variability between and within newer 2D SWE systems.[21–23]

In the first part of our study, we have demonstrated excellent overall agreement (ICC, 0.99) between 6 different ultrasound systems on 2D SWS measurements obtained in elastic phantoms of varying stiffness. Differences in the mean of median SWS between systems was generally small, particularly in the softest 2 phantoms (up to 0.06 m/s for the softest phantom [0.82 m/s] and 0.18 m/s for the next softest phantom [1.52 m/s]). Notably, the differences in median SWS between the 6 systems were greater in the stiffer phantoms (up to 0.40 m/s for the stiffest phantom [3.51 m/s]). However, the difference between a median SWS

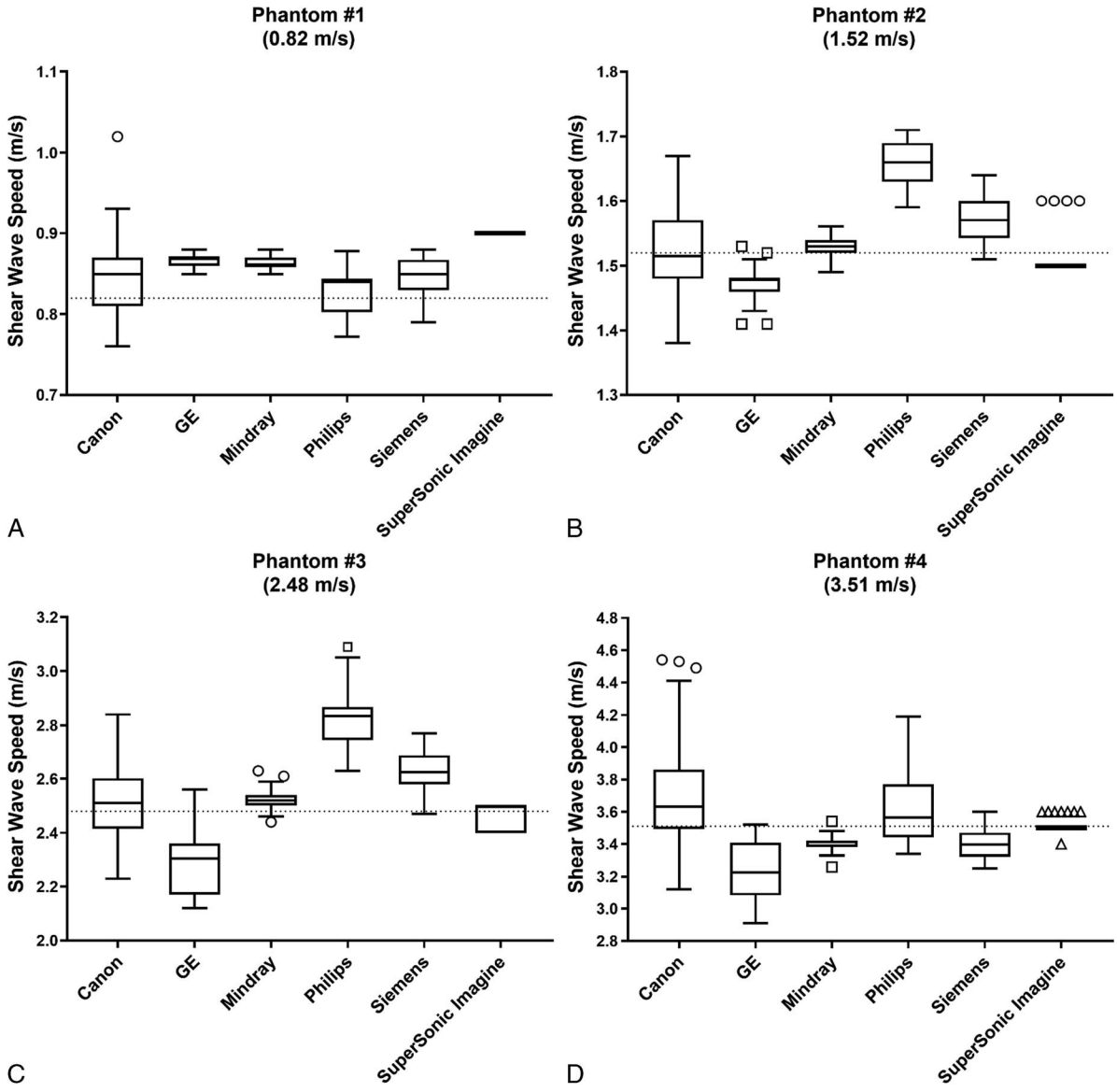

FIGURE 2. Tukey box plots showing the distribution of shear wave speed measurements (n = 40) for 6 different ultrasound systems using elastic phantoms with known stiffnesses of (a) 0.82 m/s, (b) 1.52 m/s, (c) 2.48 m/s, and (d) 3.51 m/s. The horizontal dotted line on each plot represents known phantom stiffness. Note that SuperSonic Imagine reports values with only 1 decimal place.

of 3.23 and 3.63 m/s is unlikely to be clinically important when evaluating a patient at a single time point.

We also found minimal phantom measurement variability within ultrasound systems, with the mean coefficient of variation for the 6 ultrasound systems ranging from 2.2% to 4.4% for the 4 phantoms, increasing with increasing phantom stiffness. Most IQR/median values were less than 0.05 and ranged from 0.0 to 0.10. The phenomenon of increased variability of SWS measurements within and between systems in stiffer material has previously been demonstrated in human livers.[23,35] Previous work by the Quantitative Imaging Biomarker Alliance using elastic phantoms showed 6% to 12% variation in SWS across ultrasound systems and another study demonstrated 12.7% to 17.6% variation for the greatest outlier ultrasound system using viscoelastic phantoms of varying stiffnesses.[21,22] The smaller degree of measurement variability in our study likely relates to the implementation of new, more robust ultrasound hardware and software, including updated 2D SWE algorithms.

We also demonstrated minimal interobserver variability of SWS in the phantom arm of our study. There were only small differences in median SWS for the various ultrasound system-phantom combinations when comparing 2 operators. Of the 5 (of 24) combinations that showed statistically significant differences between operators, differences were 0.03 m/s or lower for 3 combinations, which is unlikely to be clinically significant. Significant differences in measurement variance between operators also were uncommon and observed with the softest (0.82 m/s) and stiffest (3.51 m/s) phantoms.

In the second portion of our study, we found good intersystem agreement across the 6 ultrasound systems of SWS measurements obtain in human livers, with overall ICCs of 0.66 and 0.69 for first and second examinations, respectively. Interestingly, this level of agreement

TABLE 2. Median Shear Wave Speed Measurements (in m/s) for 2 Operators Using 4 Elastic Phantoms, 6 Ultrasound Systems, and 2-Dimensional Shear Wave Elastography*,†

	Phantom 1 (0.82 m/s)		Phantom 2 (1.52 m/s)		Phantom 3 (2.48 m/s)		Phantom 4 (3.51 m/s)	
	Operator							
	1	2	1	2	1	2	1	2
Aplio i800 (Canon)	0.86	0.85	1.50	1.54	2.48	2.57	3.57	3.85
P (MWU/*F* test)	0.92/0.28		0.03/0.38		**0.003**/0.16		0.02/**<0.0001**	
LOGIQ E10 (GE)	0.87	0.86	1.48	1.46	2.27	2.31	3.24	3.16
P (MWU/*F* test)	**0.0005**/0.06		0.03/**<0.0001**		0.17/0.52		0.34/0.02	
Resona 7 (Mindray)	0.87	0.86	1.54	1.52	2.53	2.52	3.39	3.42
P (MWU/*F* test)	0.04/0.53		**<0.0001**/0.41		0.56/0.02		0.06/**0.005**	
EPIQ Elite (Philips)	0.82	0.84	1.68	1.65	2.75	2.86	3.68	3.45
P (MWU/*F* test)	0.56/**<0.0001**		0.53/0.72		**<0.0001**/0.37		0.16/**<0.0001**	
ACUSON Sequoia (Siemens)	0.85	0.85	1.56	1.59	2.62	2.64	3.38	3.41
P (MWU/*F* test)	0.68/0.75		**0.008**/0.17		0.75/0.57		0.88/0.26	
Aixplorer MACH 30 (SuperSonic Imagine) ‡	0.9	0.9	1.5	1.5	2.4	2.5	3.5	3.5
P (MWU/*F* test)	>0.99/§		>0.99/>0.99		0.02/0.38		0.06/0.11	

P values indicative of statistically significant differences are in bold.

*20 measurements acquired by each operator during the same imaging session.

†Mann-Whitney *U* (MWU) test used to compare median phantom shear wave speed between operators; *F* test used to compare variances in phantom shear wave speed measurements between operators. Note that significant $P = 0.0083$ due to Bonferroni correction for multiple comparisons.

‡Shear wave speed measurements only have a single decimal place.

§*F* test *P* value not calculable because of perfect interoperator agreement.

is similar to a study assessing MRE using multiple magnetic resonance imaging scanner manufacturers, field strengths, and pulse sequences (eg, gradient recalled echo versus spin-echo echo-planar) by Serai et al. Their study also demonstrated good overall absolute agreement of liver stiffness measurements, with an ICC of 0.69.[36] Another interesting finding of the human arm of our study was that the overall mean liver SWS for 1 ultrasound system (Siemens) was ~0.2 m/s below the mean of the 6 systems combined at both first and second examinations. This finding suggests that SWS measurements of the viscoelastic human liver obtained with this system are systematically lower than the other 5 systems. Interestingly, our purely elastic phantoms, which do not simulate tissue viscosity, did not show this difference.

We also found only slight variability between SWS measurements for human examinations, with a mean coefficient of variation of 6.0% and a mean IQR/median SWS of 0.07. Comparing these values to previous work, Serai et al[36] found, for liver stiffness by MRE and across multiple scanners, field strengths, and pulse sequences, a mean coefficient of variation of 10.7% in 24 healthy volunteer subjects. Another previous study by Ferraioli et al[23] assessed liver SWS in 26 adults across 7 ultrasound elastography systems, only 2 of which were enabled with 2D SWE: Aixplorer (SuperSonic Imagine) and Aplio 500 (Canon Medical Systems USA). This study found, between the two 2D SWE systems, a mean bias of 1.54 kPa, or ~0.72 m/s. To our knowledge, no other studies have evaluated agreement of liver SWS measurements across all of the currently commercially available 2D SWE systems used in our investigation.

Another important observation of our study was that the intersystem agreement of median liver SWS values between pairs of ultrasound systems was considerably variable, with ICCs ranging from fair

TABLE 3. Summary of Liver Shear Wave Speed Measurements and Test-Retest Repeatability for 24 Research Participants Using 6 Ultrasound Systems and 2-Dimensional Shear Wave Elastography

	Mean of Medians (m/s)	Mean IQR	Mean IQR/Median	Mean Coefficient of Variation (%)	ICC (95% CI)	r_c (95% CI)
Aplio i800 (Canon)	1.44/1.40	0.13/0.13	0.09/0.09	7.7/7.3	0.87 (0.71–0.94)	0.86 (0.72–0.93)
LOGIQ E10 (GE)	1.44/1.42	0.08/0.09	0.06/0.06	4.9/5.4	0.94 (0.86–0.97)	0.94 (0.86–0.97)
Resona 7 (Mindray)	1.56/1.55	0.12/0.09	0.08/0.05	5.7/4.9	0.95 (0.89–0.98)	0.95 (0.89–0.98)
EPIQ Elite (Philips)	1.52/1.50	0.11/0.12	0.07/0.08	6.3/6.7	0.93 (0.83–0.97)	0.92 (0.83–0.97)
ACUSON Sequoia (Siemens)	1.26/1.24	0.06/0.04	0.05/0.04	4.5/3.5	0.97 (0.93–0.99)	0.97 (0.93–0.99)
Aixplorer MACH 30 (SuperSonic Imagine)*	1.5/1.5	0.1/0.1	0.1/0.1	6.9/8.1	0.94 (0.87–0.98)	0.94 (0.88–0.97)
Mean of Means (n = 6)	1.46/1.43	0.10/0.10	0.07/0.07	6.0/6.0	N/A	N/A

Median of 10 liver shear wave speed measurements from the first and second examinations (first/second), performed during 1 research encounter approximately 90 to 120 minutes apart.

*Shear wave speed measurements only have a single decimal place.

Abbreviations: IQR, interquartile range; ICC, intraclass correlation coefficient; CI, confidence interval.

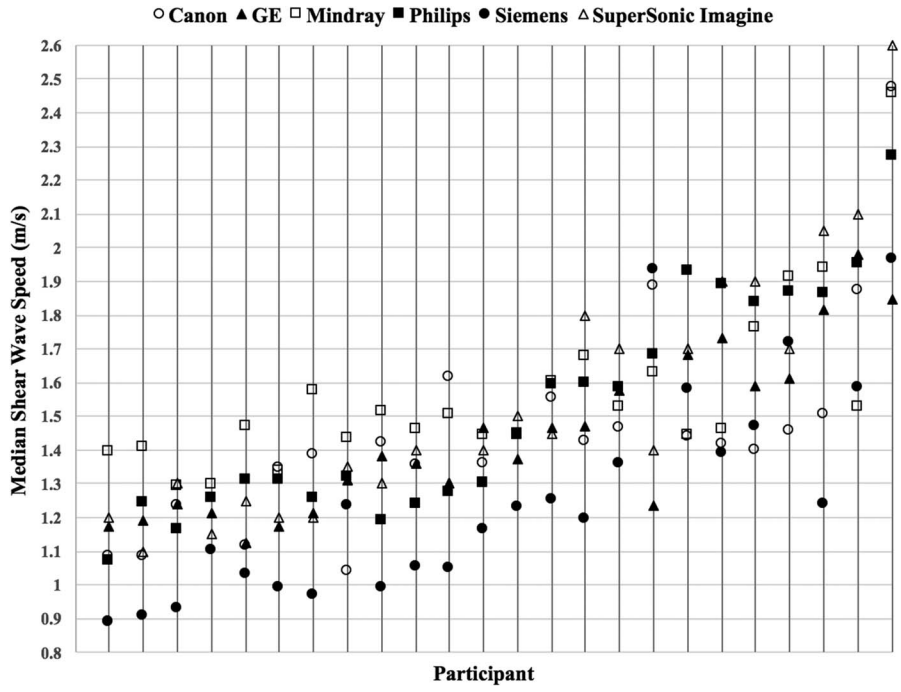

FIGURE 3. Dot plot of median shear wave speed (SWS) values from first ultrasound shear wave elastography examinations by participant, arranged along the *x* axis in ascending order of mean of median SWS (across 6 vendors).

to excellent. Using the Lin concordance correlation coefficients, a more rigorous method for assessing absolute agreement, intersystem agreement was mostly poor. Furthermore, our Lin concordance correlation coefficients are lower than those reported by Ferraioli et al,[23] who found an overall concordance correlation coefficient of 0.95, which may be explained by the smaller range of participant median SWS values observed in our population, despite the inclusion of 10 participants with known liver stiffening. However, given the observed differences in SWS measurements between some ultrasound systems and less than excellent interagreement, it may be best practice to use the same ultrasound system to serially follow liver stiffness over time in a given patient. Similarly, specific cutoff values for liver fibrosis (presence/absence and grade) likely still need to be established for each ultrasound system.

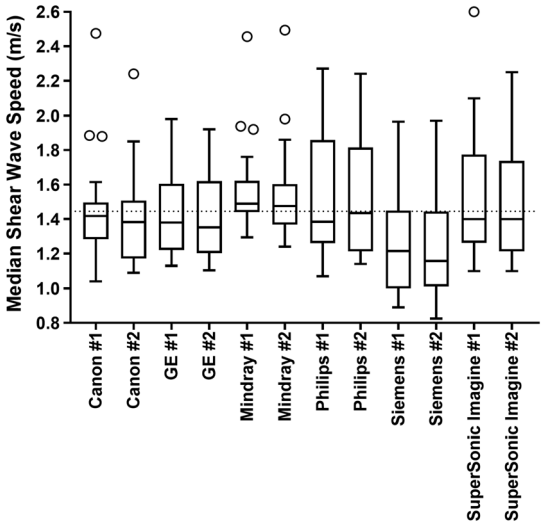

FIGURE 4. Tukey box plots showing the distribution of (a) first and (b) second examination median shear wave speed measurements for 6 different ultrasound systems in 24 participants. The horizontal dotted line on plot represents the overall mean of medians across all 6 ultrasound systems.

Finally, we observed excellent test-retest repeatability of SWS measurements in human livers for all 6 ultrasound systems. Mean bias for median SWS values was trivial for all systems (−0.02 to −0.05 m/s), and all but a single system had ICCs greater than 0.9. A previous study of 11 subjects who underwent 2D SWE examinations with a single ultrasound system (LOGIQ E9; GE Healthcare) and 4 operators showed lower test-retest agreement than observed in our study (ICC of 0.83).[37] The level of test-retest agreement observed in our study is slightly better than that demonstrated for MRE by Serai et al,[36] which showed ICCs ranging from 0.77 to 0.94, depending on magnetic resonance imaging system manufacturer, field strength, and pulse sequence. Given that test-retest repeatability is generally better than intersystem agreement (by ICCs), our results again suggest that it may be most appropriate to longitudinally follow liver stiffness in patients with chronic liver disease using the same ultrasound system.

Our study has several unique aspects. First, we have been able to assess intersystem reproducibility and repeatability of 2D SWE using all major ultrasound systems used in the United States. All of these systems have only recently become commercially available and have updated shear wave algorithms. Second, our study used state-of-the art Zerdine gel elastic phantoms of 4 different stiffnesses, with the stiffness of each phantom characterized using rigorous methods. However, elastic phantoms have limitations, as they do not exactly reflect the exact viscoelastic properties of the in vivo human liver. Third, the human arm of our study included both healthy participants and individuals with known liver stiffening from chronic liver disease in attempt to include a wide range of liver stiffnesses. Finally, all research imaging was performed over a 3-day period, with imaging performed by expert operators provided by the individual ultrasound system manufacturers.

Our study also has limitations. First, the number of subjects is relatively small. However, unlike several previous studies, we included

TABLE 4. Summary of Pairwise Agreement Among 6 Ultrasound Systems Using 2-Dimensional Shear Wave Elastography*

	Aplio i800 (Canon)	LOGIQ E10 (GE)	Resona 7 (Mindray)	EPIQ Elite (Philips)	ACUSON Sequoia (Siemens)	Aixplorer MACH 30 (SuperSonic Imagine)†
Aplio i800 (Canon)		0.60 (0.26–0.80)	0.67 (0.32–0.85	0.70 (0.43–0.86)	0.65 (0.13–0.86)	0.69 (0.41–0.86)
		0.59 (0.26–0.79)	0.66 (0.39–0.82)	0.69 (0.42–0.85)	0.64 (0.38–0.80)	0.69 (0.42–0.84)
LOGIQ E10 (GE)			0.53 (0.15–0.77)	0.80 (0.57–0.91)	0.52 (0.08–0.78)	0.80 (0.53–0.91)
			0.52 (0.19–0.74)	0.80 (0.63–0.90)	0.51 (0.23–0.72)	0.79 (0.65–0.99)
Resona 7 (Mindray)				0.70 (0.43–0.86)	0.41 (−0.10–0.74)	0.73 (0.47–0.88)
				0.70 (0.44–0.85)	0.40 (0.16–0.60)	0.72 (0.51–0.85
EPIQ Elite (Philips)					0.65 (−0.08–0.89)	0.91 (0.78–0.96)
					0.64 (0.43–0.79)	0.90 (0.80–0.95)
ACUSON Sequoia (Siemens)						0.53 (−0.04–0.81)
						0.53 (0.27–0.71)

A median of 10 liver shear wave speed measurements from the first examinations (n = 24 participants) was used for these analyses.

*Top values indicate intraclass correlation coefficient (with 95% confidence intervals), and bottom values indicate the *Lin concordance coefficient* (with 95% confidence intervals).

†Shear wave speed measurements only have a single decimal place.

patients with and without known liver disease of varying etiologies to assess reproducibility and repeatability across a range of liver stiffness and across a diverse study population. Second, although we used standardized protocols for both phantom and human imaging, other sources of variability, such as operator-related (eg, exact location of ROI placement in the right lobe of the liver) and patient-related (eg, precise manner of

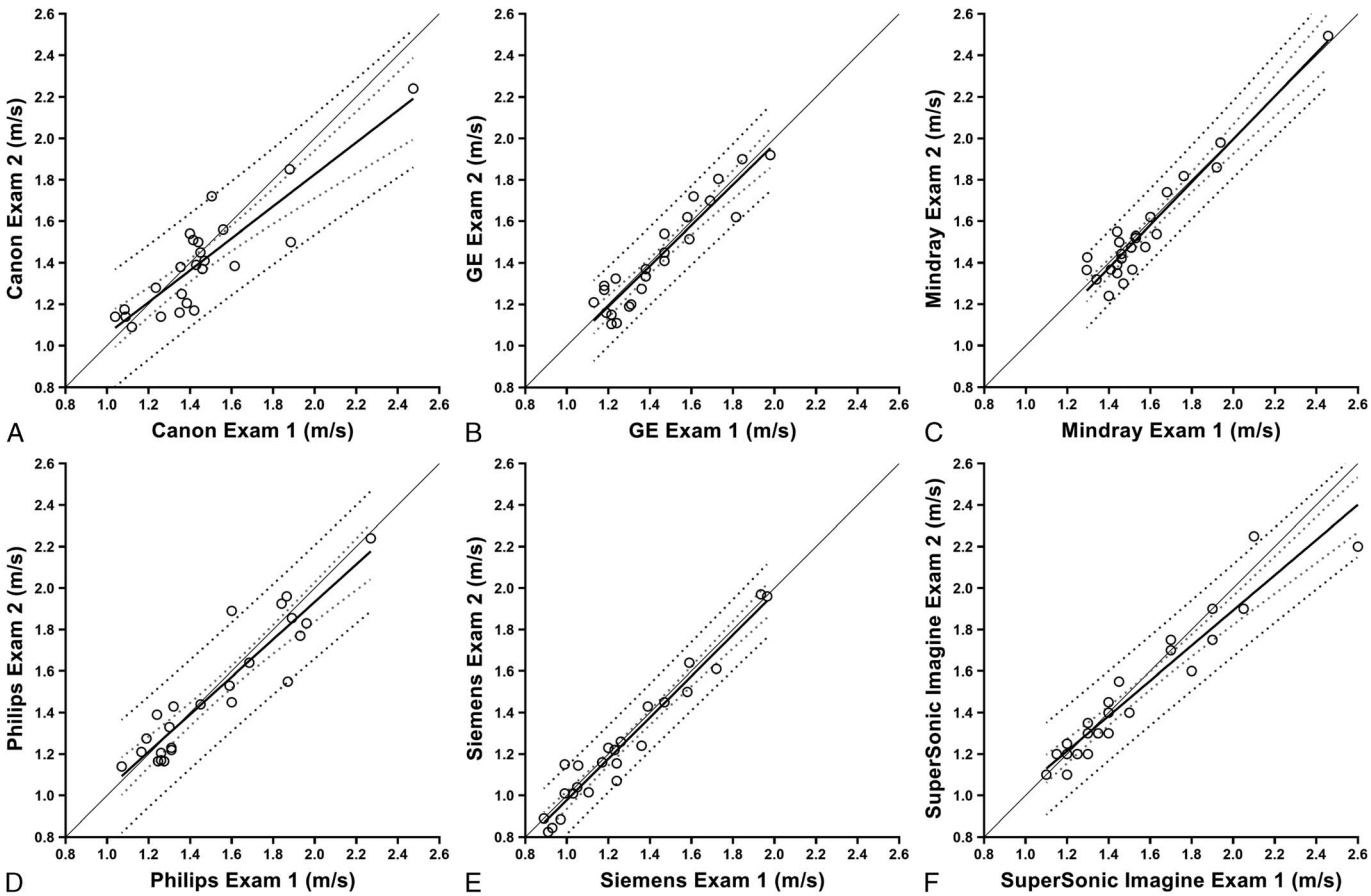

FIGURE 5. Scatterplots demonstrating test-rest repeatability per system: Canon Medical Solutions (a), GE Healthcare (b), Mindray North America (c), Philips Healthcare (d), Siemens Medical Solutions (e), and SuperSonic Imagine (f). The thick black line represents the line of best fit, whereas the thin black line represents the line of identity. The dotted gray lines represent 95% confidence interval for the line of best fit, whereas the dotted black lines represent 95% prediction limits. Exam 1 = first examination; exam 2 = second examination.

breath-holding) factors, are possible. In general, we believe the variability related to these factors is small and unlikely to impact our results or conclusions. Also, it should be noted that 1 ultrasound system reports SWS values with only 1 decimal place, which could slightly underestimate or overestimate estimates of variability and agreement when compared with other ultrasound systems. Finally, we only assessed 2D SWE methods using curved low-frequency transducers; there was no attempt to study point SWE methods or other transducers. However, the methods used in this study are applicable to clinical practice and, in particular, ultrasound 2D SWE of the liver.

In conclusion, we have shown that the overall agreement of 2D SWE measurements among 6 commercially available ultrasound systems is excellent in phantoms and good in in vivo human livers. We also have demonstrated excellent test-retest repeatability for each included ultrasound system. The degree of pairwise intersystem agreement was commonly less than excellent, indicating that it may be best practice to use the same ultrasound system to serially follow liver stiffness over time in a given patient. Similarly, cutoff SWS values used for establishing the presence of liver fibrosis and fibrosis stage should probably be system-specific. The degree of variability between 2D SWE systems and within each system is comparable to previously reported variability for MRE. Continued standardization of 2D SWE methods to improve agreement is warranted, and our results can guide such research endeavors and technological improvements.

ACKNOWLEDGMENTS

We would like to acknowledge the vendors who helped support and make this study possible, including Canon Medical Systems USA, GE Healthcare, Mindray North America, Philips Healthcare, Siemens Medical Solutions USA, and Supersonic Imagine. We would also like to acknowledge that the clinical space in which some of the research activities of this study were performed was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), through Grant 1 UL1 TR001425. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Nightingale K, McAleavey S, Trahey G. Shear-wave generation using acoustic radiation force: in vivo and ex vivo results. *Ultrasound Med Biol*. 2003;29:1715–1723.
2. Tang A, Cloutier G, Szeverenyi NM, et al. Ultrasound elastography and MR elastography for assessing liver fibrosis: part 1, principles and techniques. *AJR Am J Roentgenol*. 2015;205:22–32.
3. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. *Radiology*. 2006;239:341–350.
4. Barr RG, Memo R, Schaub CR. Shear wave ultrasound elastography of the prostate: initial results. *Ultrasound Q*. 2012;28:13–20.
5. Woo S, Suh CH, Kim SY, et al. Shear-wave elastography for detection of prostate cancer: a systematic review and diagnostic meta-analysis. *AJR Am J Roentgenol*. 2017;209:806–814.
6. Moon SK, Kim SY, Cho JY, et al. Quantification of kidney fibrosis using ultrasonic shear wave elastography: experimental study with a rabbit model. *J Ultrasound Med*. 2015;34:869–877.
7. Pawlus A, Inglot MS, Szymanska K, et al. Shear wave elastography of the spleen: evaluation of spleen stiffness in healthy volunteers. *Abdom Radiol (NY)*. 2016;41:2169–2174.
8. Lyshchik A, Higashi T, Asato R, et al. Thyroid gland tumor diagnosis at US elastography. *Radiology*. 2005;237:202–211.
9. Kawada N, Tanaka S, Uehara H, et al. Potential use of point shear wave elastography for the pancreas: a single center prospective study. *Eur J Radiol*. 2014;83:620–624.
10. Barr RG, Ferraioli G, Palmeri ML, et al. Elastography assessment of liver fibrosis: Society of Radiologists in Ultrasound Consensus Conference Statement. *Radiology*. 2015;276:845–861.
11. Ferraioli G, Parekh P, Levitov AB, et al. Shear wave elastography for evaluation of liver fibrosis. *J Ultrasound Med*. 2014;33:197–203.
12. Tang A, Cloutier G, Szeverenyi NM, et al. Ultrasound elastography and MR elastography for assessing liver fibrosis: part 2, diagnostic performance, confounders, and future directions. *AJR Am J Roentgenol*. 2015;205:33–40.
13. Ferraioli G, Tinelli C, Dal Bello B, et al. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology*. 2012;56:2125–2133.
14. Yeh WC, Li PC, Jeng YM, et al. Elastic modulus measurements of human liver and correlation with pathology. *Ultrasound Med Biol*. 2002;28:467–474.
15. Dillman JR, Heider A, Bilhartz JL, et al. Ultrasound shear wave speed measurements correlate with liver fibrosis in children. *Pediatr Radiol*. 2015;45:1480–1488.
16. Friedman SL. Mechanisms of hepatic fibrogenesis. *Gastroenterology*. 2008;134:1655–1669.
17. Ferraioli G, Filice C, Castera L, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: part 3: liver. *Ultrasound Med Biol*. 2015;41:1161–1179.
18. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. *Hepatology*. 2009;49:1017–1044.
19. Jiang T, Tian G, Zhao Q, et al. Diagnostic accuracy of 2D-shear wave elastography for liver fibrosis severity: a meta-analysis. *PLoS One*. 2016;11:e0157219.
20. Herrmann E, de Ledinghen 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.
21. Palmeri M, Nightingale K, Fielding S, et al. RSNA QIBA Ultrasound Shear Wave Speed Phase II Phantom Study in Viscoelastic Media. 2015 IEEE International Ultrasonics Symposium (IUS). Taipei, Taiwan; October 21–24, 2015.
22. Hall TJ, Milkowski A, Garra B, et al. RSNA/QIBA: shear wave speed as a biomarker for liver fibrosis staging. In: *2013 IEEE International Ultrasonics Symposium (IUS)*. Prague, Czech Republic: IEEE; 2013:397–400.
23. Ferraioli G, De Silvestri A, Lissandrin R, et al. Evaluation of inter-system variability in liver stiffness measurements. *Ultraschall Med*. 2019;40:64–75.
24. Mulabecirovic A, Mjelle AB, Gilja OH, et al. Liver elasticity in healthy individuals by two novel shear-wave elastography systems—comparison by age, gender, BMI and number of measurements. *PLoS One*. 2018;13:e0203486.
25. Dillman JR, Chen S, Davenport MS, 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.
26. Gress VS, Glawion EN, Schmidberger J, et al. Comparison of Liver shear wave elastography measurements using Siemens Acuson S3000, GE LOGIQ E9, Philips EPIQ7 and Toshiba Aplio 500 (Software Versions 5.0 and 6.0) in Healthy Volunteers. *Ultraschall Med*. 2019;40:504–512.
27. Piscaglia F, Salvatore V, Mulazzani L, et al. Differences in liver stiffness values obtained with new ultrasound elastography machines and fibroscan: a comparative study. *Dig Liver Dis*. 2017;49:802–808.
28. Rouze NC, Deng Y, Trutna CA, et al. Characterization of viscoelastic materials using group shear wave speeds. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2018;65:780–794.
29. Deng Y, Rouze NC, Palmeri ML, et al. On system-dependent sources of uncertainty and bias in ultrasonic quantitative shear-wave imaging. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2016;63:381–393.
30. Garra BS. Imaging and estimation of tissue elasticity by ultrasound. *Ultrasound Q*. 2007;23:255–268.
31. Palmeri ML, Nightingale KR. Acoustic radiation force-based elasticity imaging methods. *Interface Focus*. 2011;1:553–564.
32. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255–268.
33. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8:23–34.
34. 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:e48.
35. Northern NA, Dillman JR, Trout AT. Frequency of technical success of two-dimensional ultrasound shear wave elastography in a large pediatric and young adult cohort: a clinical effectiveness study. *Pediatr Radiol*. 2019.
36. Trout AT, Serai S, Mahley AD, et al. Liver stiffness measurements with MR elastography: agreement and repeatability across imaging systems, field strengths, and pulse sequences. *Radiology*. 2016;281:793–804.
37. Fang C, Konstantatou E, Romanos O, et al. Reproducibility of 2-dimensional shear wave elastography assessment of the liver: a direct comparison with point shear wave elastography in healthy volunteers. *J Ultrasound Med*. 2017;36:1563–1569.