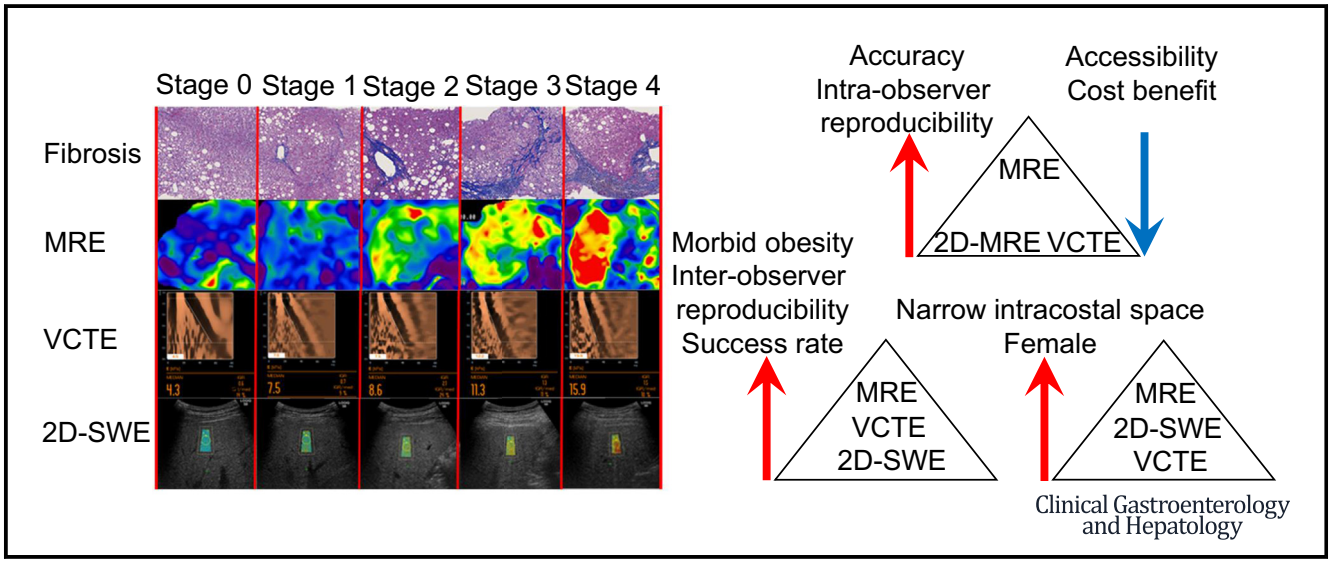
Direct Comparison of US and MR Elastography for Staging Liver Fibrosis in Patients With Nonalcoholic Fatty Liver Disease

Kento Imajo,[*] Yasushi Honda,[*] Takashi Kobayashi,[*] Koki Nagai,[*] Anna Ozaki,[*] Michihiro Iwaki,[*] Takaomi Kessoku,[*] Yuji Ogawa,[*] Hirokazu Takahashi,[‡] Yusuke Saigusa,[§] Masato Yoneda,[*] Hiroyuki Kirikoshi,[‖] Daisuke Utsunomiya,[¶] Shinichi Aishima,[#] Satoru Saito,[*] and Atsushi Nakajima[*]

[*]Department of Gastroenterology, Yokohama City University Graduate School of Medicine, Yokohama; [‡]Liver Center, Saga University Hospital, Faculty of Medicine, Saga University, Saga-shi; [§]Department of Biostatistics, Yokohama City University School of Medicine, Yokohama; [‖]Department of Clinical Laboratory, Yokohama City University Hospital, Yokohama; [¶]Department of Radiology, Yokohama City University Graduate School of Medicine, Yokohama; and [#]Department of Pathology & Microbiology, Faculty of Medicine, Saga University, Saga, Japan



BACKGROUND & AIMS: As alternatives to the expensive liver biopsy for assessing liver fibrosis stage in patients with nonalcoholic fatty liver disease (NAFLD), we directly compared the diagnostic abilities of magnetic resonance elastography (MRE), vibration-controlled transient elastography (VCTE), and two-dimensional shear wave elastography (2D-SWE).

METHODS: Overall, 231 patients with biopsy-proven NAFLD were included. Intra- and inter-observer reproducibility was analyzed using intraclass correlation coefficient in a sub-group of 70 participants, in whom liver stiffness measurement (LSM) was performed by an elastography expert and an ultrasound expert who was an elastography trainee on the same day.

Abbreviations used in this paper: ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under the receiver operating characteristic; BMI, body mass index; FIB-4, fibrosis 4; HbA1c, glycosylated hemoglobin; ICC, intraclass correlation coefficient; IQR/M, the ratio of the interquartile range of liver stiffness to the median; LSM, liver stiffness measurement; MRE, magnetic resonance elastography; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NFS, NAFLD fibrosis score; NPV, negative predictive value; PPV, positive predictive value; ROI, region of interest; SCD, skin-capsule distance; 2D-SWE, two-dimensional shear wave elastography; US, ultrasound; VCTE, vibration-controlled transient elastography.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2020.12.016

RESULTS: **Valid LSMs were obtained for 227, 220, 204, and 201 patients using MRE, VCTE, 2D-SWE, and all three modalities combined, respectively. Although the area under the curve did not differ between the modalities for detecting stage ≥1, ≥2, and ≥3 liver fibrosis, it was higher for MRE than VCTE and 2D-SWE for stage 4. Sex was a significant predictor of discordance between VCTE and liver fibrosis stage. Skin-capsule distance and the ratio of the interquartile range of liver stiffness to the median were significantly associated with discordance between 2D-SWE and liver fibrosis stage. However, no factors were associated with discordance between MRE and liver fibrosis stage. Intra- and inter-observer reproducibility in detecting liver fibrosis was higher for MRE than VCTE and 2D-SWE.**

CONCLUSIONS: **MRE, VCTE, and 2D-SWE demonstrated excellent diagnostic accuracy in detecting liver fibrosis in patients with NAFLD. MRE demonstrated the highest diagnostic accuracy for stage 4 detection and intra- and inter-observer reproducibility. UMIN Clinical Trials Registry No. UMIN000031491.**

Keywords: Nonalcoholic Fatty Liver Disease; Magnetic Resonance Elastography; Two-Dimensional Shear Wave Elastography; Vibration-Controlled Transient Elastography; Intra- and Inter-Observer Reproducibility.

Nonalcoholic fatty liver disease (NAFLD) is the leading cause of chronic liver disease worldwide. The estimated prevalence of NAFLD is 25% in the general population, 90% in those with obesity, and 60% in those with type 2 diabetic mellitus.[1] NAFLD ranges from benign nonalcoholic fatty liver to nonalcoholic steatohepatitis (NASH), which includes advanced fibrosis and is associated with hepatocellular carcinoma.[2] In addition, severe liver fibrosis is associated with vascular events and extrahepatic malignancies.[3] Therefore, early diagnosis of liver fibrosis and early interventions in NAFLD are important to reduce the risk of delayed complications.

Liver biopsy is the gold standard for assessing liver fibrosis stage in patients with NAFLD.[4] However, because of high costs, potential risks, and use of medical resources, it is not a suitable diagnostic modality.[5] Liver stiffness measurement (LSM) is a promising alternative surrogate marker for the severity of liver fibrosis using elastography such as magnetic resonance elastography (MRE), vibration-controlled transient elastography (VCTE), and two-dimensional shear wave elastography (2D-SWE).[6,7] Despite the interest in alternative approaches, there is no consensus on the optimal modality for noninvasive clinical evaluation of liver fibrosis stage in NAFLD.[8] Previously, we reported that MRE has higher diagnostic ability than VCTE in noninvasively assessing liver fibrosis stage in patients with NAFLD.[9] Although MRE appears to be superior to VCTE in diagnosing liver fibrosis in these patients, the costs of VCTE are lower than those of MRE.[9,10]

2D-SWE is a novel technique that enables real-time quantification of LSM using B-mode imaging. However, few studies have evaluated the accuracy of 2D-SWE in patients with NAFLD.[11] Therefore, the aim of this study was to directly compare the ability of MRE, VCTE, and 2D-SWE in the noninvasive assessment of liver fibrosis stage in NAFLD.

Materials and Methods

Patients

This prospective study was performed at Yokohama City University Hospital, Yokohama, Japan. The study protocol is presented in Supplementary Figure 1. Between March 2018 and April 2020, 231 patients with NAFLD who underwent liver biopsy were enrolled. The study protocol was approved by the ethics review committee, and all patients provided written informed consent. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki 2013, approved by the Ethics Committee of Yokohama City University Hospital, and registered in the UMIN Clinical Trials Registry (UMIN000031491).

Histopathologic and Immunohistochemical Evaluations

Biopsy samples were assessed by an experienced pathologist who specialized in liver pathology. Steatosis, lobular inflammation, ballooning, and fibrosis were histologically scored. NAFLD was defined as macrovesicular steatosis in >5% of hepatocytes. Patients with steatosis, lobular inflammation, ballooned hepatocytes, and perisinusoidal/pericellular fibrosis were diagnosed with NASH.[12] Liver fibrosis stage was classified according to the report by Brunt.[13]

Magnetic Resonance Elastography

MRE was performed by using 3.0T imagers (GE Healthcare, Milwaukee, WI) as previously described.[14] Details are presented in the Supplementary Material. MRE images were interpreted as described previously.[15] All measurements were performed by a hepatologist with 6 years of experience in interpreting MRE (K.I.).

Vibration-Controlled Transient Elastography

LSM was assessed using VCTE (3.5-MHz M and/or 2.5-MHz XL probe, FibroScan; EchoSens, Paris, France). The probe was decided by the automatic probe selection tool in the FibroScan software. All examinations were initiated with M probe, and XL probe was used only when prompted by the automatic probe selection tool. The procedure has been described previously.[16] VCTE methods are detailed in the Supplementary Material.

Two-Dimensional Shear Wave Elastography

2D-SWE was performed by using Logic S8 system (GE Healthcare). This new technique uses comb-push and time-aligned sequential tracking for the generation of large elasticity maps superimposed on the grayscale image obtained by using conventional ultrasound (US). 2D-SWE methods are detailed in the Supplementary Material.

Scoring Systems

The following scores were calculated in each patient: Fibrosis 4 (FIB-4) index[17] and NAFLD fibrosis score (NFS).[18] These scores are detailed in the Supplementary Material.

Interobserver and Intraobserver Reproducibility Analysis

We analyzed intraobserver and interobserver reproducibility in a subgroup of 70 patients. The protocol is presented in Supplementary Figure 2. This analysis was performed by an elastography expert (K.I.) with more than 6 years of experience in VCTE and 2 years in 2D-SWE and an elastography trainee (no experience in elastography; received a single day of training in VCTE and 2D-SWE) with US experience (Yasushi Honda; performed more than 1000 examinations). Agreement between the elastographies was evaluated by using intraclass correlation coefficient (ICC). In addition, we analyzed 70 patients in the first (n = 35) and second (n = 35) halves of the study to examine whether the interobserver reproducibility varies with the experience. These methods are detailed in the Supplementary Material.

Statistical Analysis

Continuous and categorical variables are summarized as median and interquartile ranges and frequencies and percentages, respectively. Analysis of variance with Scheffe multiple testing correction was used for univariate comparisons between groups. Kruskal–Wallis test was used for comparisons of nonparametric data of more than 2 independent groups. The z test was used for comparisons of the area under the receiver operating characteristic (AUROC) curve between 2 groups.[19] We used ICC to assess the interobserver and intraobserver reproducibility of MRE, VCTE, and 2D-SWE. The ICCs were obtained by using two-way random-effects model for absolute agreement and single rater.[20] Statistical analyses were performed by using SPSS v12.0 (SPSS Inc, Chicago, IL). All authors had access to the study data and reviewed and approved the final manuscript.

What You Need to Know

Background

There are no head-to-head previous studies directly comparing the diagnostic ability of MRE, VCTE, and 2D-SWE in the noninvasive assessment of liver fibrosis stage in a large cohort of biopsy-proven NAFLD.

Findings

MRE, VCTE, and 2D-SWE exhibited excellent diagnostic accuracy in detecting liver fibrosis stage in patients with NAFLD, with MRE demonstrating the highest diagnostic accuracy for stage 4 detection and intra/interobserver reproducibility.

Implications for patient care

MRE has several advantages over VCTE and 2D-SWE, including excellent diagnostic accuracy and a larger sampling area, whereas the main disadvantage of MRE compared with VCTE and 2D-SWE is the higher cost.

Results

Patient Characteristics

In 231 patients with liver biopsy-proven NAFLD, LSM was evaluated by using MRE, VCTE, and 2D-SWE in 227 (98.3%), 220 (95.2%), and 204 (88.3%) patients, respectively (Supplementary Figure 1). The reasons for failure in these elastographies are presented in Supplementary Table 1. Finally, 201 patients in whom all 3 elastography modalities were successfully performed were evaluated. Their characteristics are presented in Table 1.

Direct Comparison of the Diagnostic Accuracy of Magnetic Resonance Elastography, Vibration-Controlled Transient Elastography, and Two-Dimensional Shear Wave Elastography

The mean LSM values of MRE were 2.19, 2.80, 3.53, 4.78, and 6.44 for stages 0, 1, 2, 3, and 4, respectively; the corresponding values of VCTE were 4.11, 6.95, 8.14, 13.8,

Table 1. Clinical, Serologic, and Histologic Characteristics of Patients With Nonalcoholic Fatty Liver Disease (NAFLD) in Whom All Elastography Modalities Were Successfully Performed

Number (n)	201
Age (*y*)	61.0 (51.0–71.0)
Sex (male:female)	106:95
Body mass index (kg/m^2)	27.1 (25.2–30.8)
Platelets ($/10^4\ \mu L$)	19.4 (15.2–23.6)
AST (*IU/L*)	42.0 (32.0–60.0)
ALT (*IU/L*)	48.0 (33.0–75.5)
GGT (*IU/L*)	55.0 (37.0–96.0)
CRP (*mg/L*)	0.12 (0.07–0.3)
Cr (*mg/dL*)	0.79 (0.56–1.15)
FBS (*mg/dL*)	115 (102–133.3)
Fasting insulin ($\mu U/mL$)	16.0 (11.3–24.4)
HbA1c (%)	6.2 (5.7–7.0)
DM (%)	127 (61.7)
HT (%)	97 (48.3)
DLP (%)	152 (75.6)
Steatosis grade (n)	
5%–33%	100
33%–66%	66
>66%	35
Lobular inflammation (n)	
None	3
<2 foci per ×200 field	120
2–4 foci per ×200 field	70
>4 foci per ×200 field	8
Liver cell ballooning (n)	
None	60
Few balloon cells	114
Many balloon cells	27
NAFL/NASH (n)	63/138
NAS (n)	
1/2/3/4/5/6/7	4/31/47/55/28/21/10/5
Fibrosis stage (n)	
None	9
Perisinusoidal or periportal	50
Perisinusoidal and portal/periportal	36
Bridging fibrosis	69
Cirrhosis	37

NOTE. Values are median (interquartile ranges) or n (%).
ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; Cr, creatinine; CRP, C-reactive protein; DLP, dyslipidemia; DM, diabetes mellitus; FBS, fasting blood glucose; GGT, gamma-glutamyl transferase; HbA1c, glycosylated hemoglobin; HT, hypertension; NAFLD, nonalcoholic fatty liver disease; NAS, NAFLD activity score; NASH, nonalcoholic steatohepatitis.

and 19.5 kPa, respectively, and those for 2D-SWE were 5.59, 6.65, 8.04, 10.69, and 12.37 kPa, respectively (Figure 1*A*). AUROC curves, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are presented in Figure 1*B* and Supplementary Table 2. Table 2 summarizes the direct comparisons of the diagnostic accuracy of LSM and scoring systems (FIB-4 index and NFS). There were no differences in the diagnostic ability between MRE and VCTE or 2D-SWE in detecting liver fibrosis stage ≥1, ≥2, and >3. However, the diagnostic accuracy for liver fibrosis stage 4 was significantly greater with MRE than VCTE and 2D-SWE (Figure 1*B*, Table 2). There was no difference in the diagnostic accuracy between VCTE and 2D-SWE in detecting liver fibrosis stages. All modalities demonstrated good diagnostic accuracy in evaluating liver fibrosis stages relative to clinical scoring systems. In addition, MRE, VCTE, and 2D-SWE images of 5 representative patients with various fibrosis stages that were based on the optimal threshold and average are presented in Figure 2. One patient with cirrhosis and impaired liver dysfunction (Child–Pugh class 6) was included in this study. Interestingly, MRE, VCTE, and 2D-SWE results are all severe at first glance in this patient (Supplementary Figure 3). The relationship between steatosis, lobular inflammation, or ballooning and LSM is presented in Supplementary Table 3.

Factors Associated With Discordance Between Elastography and Pathology

Univariable and multivariable regression analysis used to investigate background factors related to discordance between fibrosis stages predicted using MRE and those predicted by pathology for at least 2 stages revealed substantial discordance in 16 of 201 patients (8.0%). Of them, 7 were upstaged (elastography predicted higher stage than histology), and 9 were downstaged (elastography predicted lower stage than histology) (Table 3). No significant differences were found in demographic or serologic profiles between patients with and without discordance. Substantial discordance between VCTE and histologic staging for at least 2 stages was observed in 23 of 201 patients (11.4%); of them, 10 were upstaged, and 13 were downstaged. We found a significant difference in sex between patients with concordance and those with discordance only in the VCTE upstaged group. Finally, substantial discordance between 2D-SWE and histologic staging for at least 2 stages was observed in 21 of 201 patients (10.4%); of them, 10 were upstaged, and 11 were downstaged. On univariable analysis, in 2D-SWE, body mass index (BMI), skin-capsule distance (SCD), aspartate aminotransferase (AST), and the ratio of the interquartile range of liver stiffness to the median (IQR/M) were associated with discordance in the upstaged group, whereas no significant differences were identified between patients with concordance and those with discordance in the downstaged group. On multivariable analysis, SCD and IQR/M were significantly more likely to be associated with

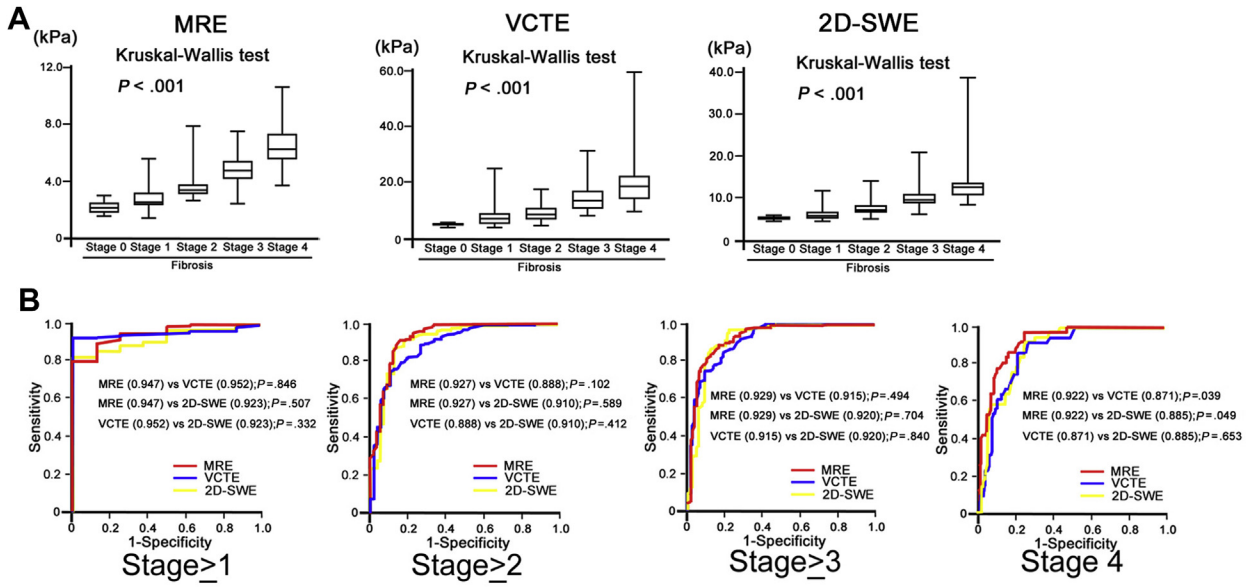


Figure 1. Direct comparison of area under the receiver operating characteristic (AUROC) curves between magnetic resonance elastography (MRE), vibration-controlled transient elastography (VCTE), and two-dimensional shear wave elastography (2D-SWE) in patients with nonalcoholic fatty liver disease (NAFLD) for each fibrosis stage (n = 201). There was no difference between MRE and either VCTE or 2D-SWE in the diagnostic performance for detecting liver fibrosis stage ≥1, ≥2, and >3. The diagnostic accuracy for detecting liver fibrosis stage 4 was significantly greater with MRE than with VCTE and 2D-SWE.

discordance between LSM- and pathology-predicted liver fibrosis stages in the 2D-SWE upstaged group.

Intraobserver and Interobserver Reproducibility

We examined the diagnostic abilities of MRE, VCTE, and 2D-SWE in detecting liver fibrosis in a subgroup of patients with biopsy-proven NAFLD who underwent elastographies twice by an expert and once by a trainee (n = 70; Supplementary Table 4). We did not find any significant differences in the detection of liver fibrosis stage between the 3 modalities in this cohort (Supplementary Table 5). The intraobserver reproducibility of the 3 methods using different scales is demonstrated in Table 4. The lowest intraobserver ICC was observed with VCTE and the highest with MRE compared with 2D-SWE. In contrast, the lowest interobserver ICC was observed with 2D-SWE and the highest with MRE compared with VCTE. Interestingly, the interobserver reproducibility of VCTE and 2D-SWE was higher in patients (n = 35) in the second half than those (n = 35) in the first half (Supplementary Table 6).

Discussion

This study provided a head-to-head comparison of LSM on a large scale; factors associated with discordance between LSM and fibrosis stage; and intra/interobserver reproducibility using MRE, VCTE, and 2D-SWE in patients with biopsy-proven NAFLD.

Our previous head-to-head study directly comparing MRE and VCTE (only M probe) indicated no significant difference in AUROC of the 2 modalities in distinguishing between stages 0 and 1–4 fibrosis and stages 0–2 and 3–4 fibrosis. However, the diagnostic accuracy was better with MRE than with VCTE for stage >2 (AUROC, 0.91 vs 0.82; $P = .001$) and stage 4 fibrosis (AUROC, 0.97 vs 0.92; $P = .049$).[9] In another study, MRE was more accurate than VCTE (M and XL probe) for the diagnosis of stage ≥2 (AUROC, 0.82 vs 0.67; $P = .01$) and stage 4 fibrosis (AUROC value, 0.87 vs 0.69; $P = .05$).[10] However, there was no significant difference between the 2 elastography methods in diagnosing any other dichotomized stage of fibrosis. Consistent with the results of these studies, we found that MRE was more accurate than VCTE (M and XL probe) in diagnosing stage 4 fibrosis. However, in contrast to these studies, the present study demonstrated no difference in the diagnostic ability between MRE and VCTE in distinguishing stages 0–1 from stages 2–4 fibrosis. It is possible that our findings may have been affected by the ordinary use of XL probe and the larger sample size and smaller number of patients with fibrosis stages 0–1 than the respective numbers in previous reports.

Although prior studies have compared the diagnostic ability of MRE and VCTE, there are limited studies on the diagnostic ability of 2D-SWE in NAFLD. A previous study reported that 2D-SWE demonstrates good diagnostic accuracy in the detection of stage ≥2 and ≥3 fibrosis in NAFLD, but it was not different from MRE or VCTE.[21] In contrast, our findings indicated that MRE is more accurate than 2D-SWE in diagnosing stage 4 despite no differences for stages ≥1, ≥2, and ≥3. It is possible that our findings may have been affected by the larger sample size than that in previous reports.

Table 2. Direct Comparisons of the Diagnostic Accuracy Between MRE, VCTE, 2D-SWE, and Scoring Systems in Detecting Liver Fibrosis Stage in Patients With Nonalcoholic Fatty Liver Disease (NAFLD)

Fibrosis stage (n = 201)	Stage 0 vs stage 1–4						Stage 0–1 vs stage 2–4						Stage 0–2 vs stage 3–4						Stage 0–3 vs stage 4					
Modality	Cutoff	AUROC	95% CI	*P* value	*P* value vs MRE	*P* value vs VCTE	*P* value vs SWE	Cutoff	AUROC	95% CI	*P* value	*P* value vs MRE	*P* value vs VCTE	*P* value vs SWE	Cutoff	AUROC	95% CI	*P* value	*P* value vs MRE	*P* value vs VCTE	*P* value vs SWE	Cutoff	AUROC	95% CI	*P* value	*P* value vs MRE	*P* value vs VCTE	*P* value vs SWE
MRE	2.92	0.947	0.863–0.980	<.001[a]	—	.846	.507	3.19	0.927	0.866–0.961	<.001[a]	—	.102	.589	3.90	0.929	0.882–0.958	<.001[a]	—	.494	.704	4.62	0.922	0.871–0.955	<.001[a]	—	.039[a]	.049[a]
VCTE	5.00	0.952	0.910–0.974	<.001[a]	.846	—	.332	8.40	0.888	0.823–0.931	<.001[a]	.102	—	.412	9.70	0.915	0.867–0.947	<.001[a]	.494	—	.840	12.40	0.871	0.807–0.917	<.001[a]	.039[a]	—	.653
2D-SWE	6.35	0.923	0.851–0.962	<.001[a]	.507	.332	—	7.55	0.910	0.843–0.951	<.001[a]	.589	.412	—	8.88	0.920	0.865–0.953	<.001[a]	.704	.840	—	9.88	0.885	0.836–0.925	<.001[a]	.049[a]	.653	—
FIB-4 index	1.28	0.814	0.716–0.884	.001[a]	.006[a]	.002[a]	.005[a]	1.55	0.829	0.756–0.884	<.001[a]	.003[a]	.104	.028[a]	2.16	0.831	0.768–0.880	<.001[a]	<.001[a]	.005[a]	.004[a]	2.28	0.828	0.748–0.887	<.001[a]	.015[a]	.290	.144
NFS	–0.59	0.824	0.702–0.904	.003[a]	.058[a]	.022[a]	.034[a]	–0.59	0.844	0.771–0.894	<.001[a]	.009[a]	.216	.069[a]	–0.11	0.838	0.776–0.885	<.001[a]	.002[a]	.012[a]	.008[a]	0.30	0.849	0.777–0.901	<.001[a]	.036[a]	.586	.265

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; FIB-4 index, fibrosis 4 index; MRE, magnetic resonance elastography; NFS, NAFLD fibrosis score; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.
[a]Significant at $P < .05$.

A previous report using pairwise comparisons also revealed similar diagnostic accuracy between VCTE and 2D-SWE in liver fibrosis staging in patients with NAFLD.[21] Although VCTE and 2D-SWE are very useful in diagnosing the severity of liver fibrosis in NAFLD, their success rate depends on the operator's experience and other factors such as age, ascites, width of the intercostal space, BMI, and visceral fat. Sporea et al[22] reported a rate of reliable measurements of 81.6% in VCTE, which is similar to that reported by Castéra et al.[23] Other authors have suggested that a high BMI or severe ascites may explain the failure of VCTE.[23–25] Indeed, we previously demonstrated that VCTE with M probe is unsuccessful in assessing LSM in approximately 10% of patients with NAFLD.[9] In the present study, the failure rate was lower (approximately 4.7%) because of the combined use of M and XL probes. However, VCTE XL probes generate lower LSM values than M probes, and validated XL thresholds for fibrosis stage in viral hepatitis and NAFLD have not yet been established. Regarding 2D-SWE, LSM was unsuccessful in approximately 11.7% of patients because of the low success rate and unreliable measurements associated with morbid obesity and narrow intercostal space.

Sex was a significant predictor of discordance between histology and VCTE for at least 2 fibrosis stages, whereas SCD was a significant predictor of discordance between histology and 2D-SWE. Because intercostal space tends to be narrower in women than men with NAFLD, the discordance between VCTE and histologic staging may be higher in women. Severe obesity is associated with higher SCD, which is the main reason for unreliable LSM using standard VCTE M probes. However, XL probe reduces the failure rate of LSM in obese patients.[26,27] No discordance associated with BMI and SCD between VCTE and liver biopsy-proven fibrosis stage may be due to the use of XL probe. In contrast, the optimal positioning of the probe for 2D-SWE in patients with morbid obesity may be more difficult and less reproducible. Therefore, the range of variation and quartiles of LSM significantly increased with increasing SCD. Although 2D-SWE has excellent diagnostic ability in patients with NAFLD, data regarding sampling variability in those with higher SCD are required.

Several studies have investigated intraobserver and interobserver variability of MRE, VCTE, and 2D-SWE; however, there are limited studies in NAFLD. Furthermore, no study has directly compared these variabilities between MRE, VCTE, and 2D-SWE in NAFLD. In MRE, intraobserver and interobserver repeatability in patients with liver fibrosis was excellent and superior to those of VCTE and 2D-SWE. Lower ICC for interobserver repeatability between the expert and trainee in VCTE and 2D-SWE suggests

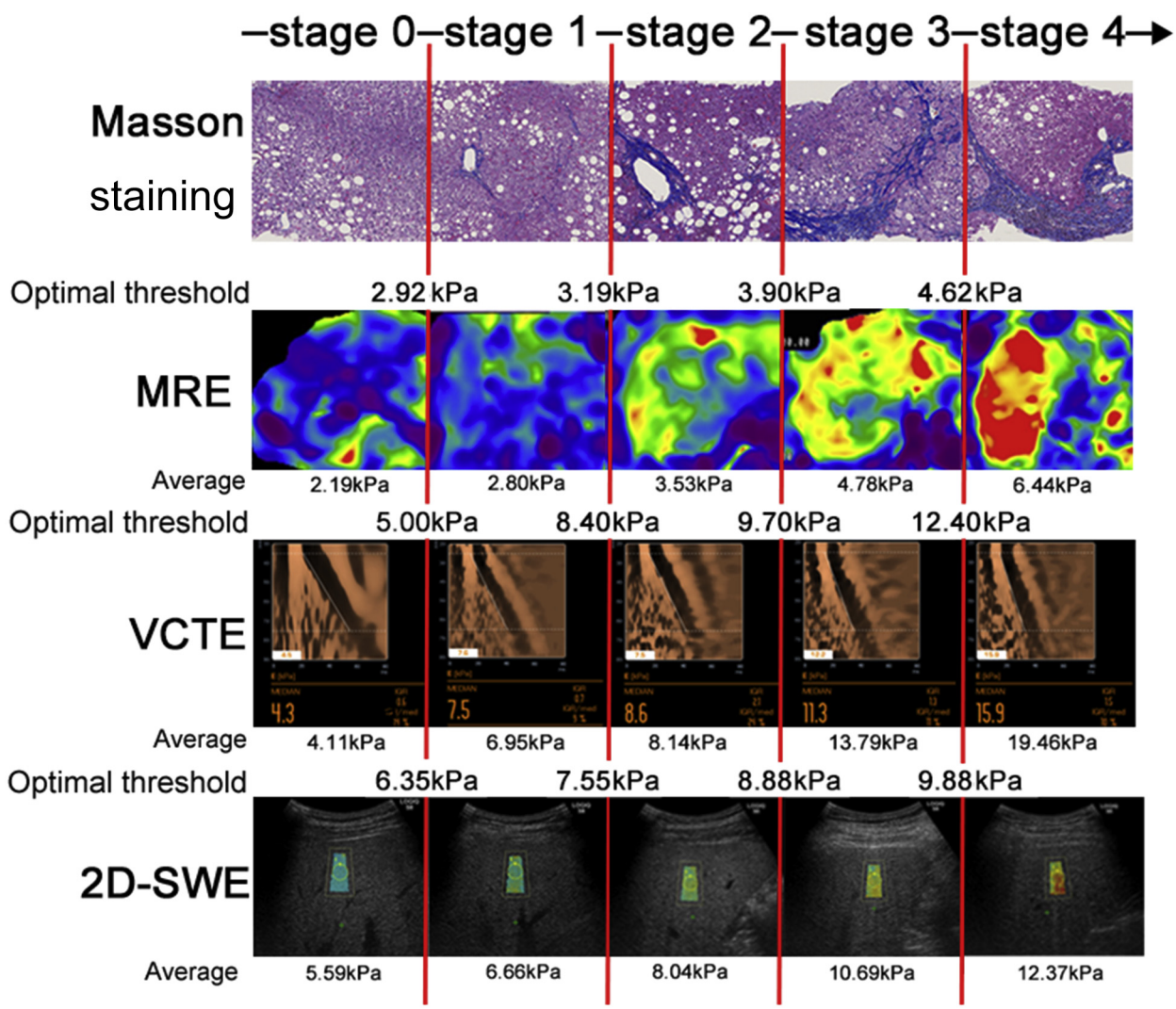


Figure 2. Pathologic fibrosis; magnetic resonance elastography (MRE), vibration-controlled transient elastography (VCTE), and two-dimensional shear wave elastography (2D-SWE) images; average values; and optimal threshold values from representative patients with stage 0, 1, 2, 3, and 4 fibrosis. The optimal MRE threshold was 3.03, 3.19, 4.12, and 4.62 kPa for stages ≥1, 2, 3, and 4, respectively. The optimal VCTE threshold was 5.00, 7.80, 9.80, and 12.4 kPa for stages ≥1, 2, 3, and 4, respectively. The optimal 2D-SWE threshold was 6.35, 7.55, 8.88, and 9.98 kPa for stages ≥1, 2, 3, and 4, respectively.

that experience in US elastography is necessary for reliable results in NAFLD. European Federation of Societies for Ultrasound in Medicine and Biology guidelines on the clinical use of US elastography suggest that VCTE can be readily learned by a trainee after approximately 100 examinations, whereas an experienced operator for 2D-SWE should have performed >300 abdominal US scans or >50 supervised 2D-SWE examinations.[28] In fact, the ICC of interobserver repeatability between an expert and trainee increased with increasing number of cases for the trainee (Supplementary Table 6). Castéra et al[23] reported that experience of 500 examinations resulted in significant differences in VCTE results, probably because of the lack of B-mode visibility. A previous report demonstrated that a trainee of 2D-SWE needed to acquire the necessary skills to perform elastogram accurately.[29] In our study, the intraobserver and interobserver reproducibility of VCTE and 2D-SWE was lower than that of MRE, suggesting that the MRE findings were homogeneous between the trainee and expert in contrast to VCTE and 2D-SWE findings. Extensive experience with US elastographies is needed to determine the best angle to avoid the ribs and large vessels.

On the basis of the present results, MRE appears to be advantageous over VCTE and 2D-SWE (Supplementary Table 7). First, it provides more accurate diagnosis of stage 4 fibrosis. Second, it allows for less discrepancy between LSM and histologic staging. Third, it may be easier to perform in patients with narrow intercostal space. Fourth, it allows for the measurement of larger liver area, avoiding sampling variability caused by the heterogeneity of advanced fibrosis.[30] However, because

Table 3. Factors Associated With Discordance Between LSM and Fibrosis Staging

	MRE (fibrosis stage < stiffness; upstaged group)				VCTE (fibrosis stage < stiffness; upstaged group)				2D-SWE (fibrosis stage < stiffness; upstaged group)			
	Concordance (n = 194)	Discordance (n = 7)	Univariate *P* value	Multiple *P* value	Concordance (n = 191)	Discordance (n = 10)	Univariate *P* value	Multiple *P* value	Concordance (n = 191)	Discordance (n = 10)	Univariate *P* value	Multiple *P* value
Age, *y*	60.7 ± 12.6	64.8 ± 9.53	.426	—	60.8 ± 12.3	58.4 ± 16.1	.590	—	61.0 ± 12.3	56.0 ± 16.0	.238	—
Sex (male/female)	103/91	3/4	.426	—	106/85	2/8	.050[a]	.045[a]	104/87	4/6	.231	—
BMI (kg/m^2)	27.9 ± 4.16	26.6 ± 1.99	.436	—	27.8 ± 4.11	29.1 ± 3.49	.092	—	27.8 ± 4.09	30.2 ± 3.87	.049[a]	.085
SCD	22.0 ± 4.34	20.8 ± 3.28	.503	—	21.9 ± 4.20	24.1 ± 4.85	.099	—	21.8 ± 4.23	25.9 ± 4.58	.005[a]	.025[a]
Platelets ($/10^4$ μL)	19.3 ± 7.04	19.1 ± 4.70	.951	—	19.2 ± 6.94	23.0 ± 6.02	.107	—	19.1 ± 6.99	22.3 ± 6.10	.196	—
AST (*IU/L*)	49.9 ± 28.2	50.2 ± 24.6	.978	—	49.0 ± 26.1	47.6 ± 19.2	.866	—	49.7 ± 28.5	56.7 ± 28.5	.050[a]	.260
ALT (*IU/L*)	58.9 ± 40.8	55.8 ± 24.1	.854	—	57.7 ± 39.1	66.7 ± 36.3	.501	—	58.9 ± 41.1	66.3 ± 31.2	.188	—
IQR/M	—	—	—	—	14.4 ± 9.42	13.6 ± 10.2	.796	—	19.1 ± 8.10	27.4 ± 30.1	.002[a]	.049[a]

	MRE (fibrosis stage > stiffness; downstaged group)				VCTE (fibrosis stage > stiffness; downstaged group)				SWE (fibrosis stage > stiffness; downstaged group)			
	Concordance (n = 192)	Discordance (n = 9)	Univariate *P* value	Multiple *P* value	Concordance (n = 188)	Discordance (n = 13)	Univariate *P* value	Multiple *P* value	Concordance (n = 190)	Discordance (n = 11)	Univariate *P* value	Multiple *P* value
Age, *y*	61.0 ± 12.6	57.1 ± 10.6	.394	—	60.4 ± 12.5	62.9 ± 12.8	.499	—	60.7 ± 12.6	62.8 ± 11.1	.632	—
Sex (male/female)	103/89	4/5	.883	—	100/88	6/7	.811	—	100/90	6/5	.853	—
BMI (kg/m^2)	27.8 ± 4.17	28.5 ± 2.45	.806	—	27.8 ± 4.18	28.4 ± 2.92	.635	—	27.8 ± 4.05	28.7 ± 5.53	.543	—
SCD	21.9 ± 4.35	23.6 ± 2.53	.326	—	21.8 ± 4.25	23.1 ± 3.83	.326	—	21.8 ± 4.04	23.7 ± 8.27	.217	—
Platelets ($/10^4$ μL)	19.1 ± 6.95	22.9 ± 6.79	.135	—	19.1 ± 6.91	19.1 ± 5.76	.990	—	19.2 ± 7.09	21.0 ± 3.50	.446	—
AST (*IU/L*)	50.3 ± 28.5	39.0 ± 7.05	.265	—	49.8 ± 23.8	51.8 ± 50.7	.801	—	49.2 ± 27.5	64.2 ± 37.3	.116	—
ALT (*IU/L*)	58.9 ± 40.9	56.0 ± 26.9	.840	—	59.0 ± 35.3	67.6 ± 85.7	.478	—	58.0 ± 39.6	75.3 ± 54.6	.210	—
IQR/M	—	—	—	—	14.1 ± 8.24	16.1 ± 18.8	.495	—	19.7 ± 11.4	18.3 ± 4.86	.758	—

NOTE. Values are mean ± standard deviation.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; IQR/M, interquartile range/median; NAFLD, nonalcoholic fatty liver disease; NAS, NAFLD activity score; NASH, nonalcoholic steatohepatitis; SCD, skin-capsule distance.

[a]Significant at $P < .05$.

Table 4. Intraobserver and Interobserver Reproducibility of MRE, VCTE, and 2D-SWE in Detecting Liver Fibrosis in Patients With NAFLD

	Intraobserver variability (operator 1)		Interobserver variability (between operators 1 and 2)	
Elastography (n = 70)	ICC	95% CI	ICC	95% CI
MRE	0.9197	0.8739–0.9493	0.9481	0.9134–0.9685
VCTE	0.8369	0.7500–0.8955	0.7900	0.6824–0.8640
2D-SWE	0.8469	0.7649–0.9019	0.7053	0.5645–0.8061

CI, confidence interval; ICC, intraclass correlation coefficient; MRE, magnetic resonance elastography; NAFLD, nonalcoholic fatty liver disease; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.

of the variability in the availability, expertise, and costs of these modalities, no single approach is clearly superior to another. Clinicians should consider all available modalities for initial risk stratification.

Our study had several limitations. First, liver biopsy was used as the gold standard for assessing liver pathology, which can be limited by sampling errors. Second, a different type of 2D-SWE, such as Supersonic Imagine Aixplorer, could have led to different results. Third, the identification of regions of interest (ROIs) on MRE and 2D-SWE are not yet standardized. We measured the mean of large ROIs in the right liver lobe to reduce measurement bias. Fourth, type 2 errors due to insufficient statistical power may be present in the detection of various fibrosis stages between the modalities. Fifth, because only Japanese patients with NAFLD were included, the results may not be applicable to the general population (higher BMI) and populations of other countries. Finally, selection bias may be present partly because of the increased likelihood of undergoing liver biopsies in patients with NAFLD at risk for NASH with advanced fibrosis. Therefore, the enrolled patients were older (mean age, 60.8 years) than the general population. It is likely that the thresholds may be different in primary care population.

Conclusions

Our direct comparisons of the 3 modalities revealed no difference in their ability in diagnosing stage $\geq$1, $\geq$2, and $\geq$3; however, MRE was better than VCTE and 2D-SWE in detecting stage 4. Sex and SCD/IQR/M were significantly associated with discordance between both VCTE and 2D-SWE and pathology findings. However, no factors were associated with discordance in MRE. MRE demonstrated excellent intraobserver and interobserver repeatability, which was better than that of VCTE and 2D-SWE. A graphic abstract about the differences in MRE, VCTE, and 2D-SWE for assessing liver fibrosis stage is shown in Supplementary Figure 4. Further research is needed to explore the prognostic value of these diagnostic techniques in patients with NAFLD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of Clinical Gastroenterology and Hepatology at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2020.12.016.

References

1. Hamaguchi M, Kojima T, Takeda N, et al. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. Ann Intern Med 2005;143:722–728.
2. Harrison SA, Torgerson S, Hayashi PH. The natural history of nonalcoholic fatty liver disease: a clinical histopathological study. Am J Gastroenterol 2003;98:2042–2047.
3. Angulo P. Nonalcoholic fatty liver disease. N Engl J Med 2002;18:1221–1231.
4. Cadranel JF. Good clinical practice guidelines for fine needle aspiration biopsy of the liver: past, present and future. Gastroenterol Clin Biol 2002;26:823–824.
5. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128:1898–1906.
6. Yoneda M, Imajo K, Takahashi H, et al. Clinical strategy of diagnosing and following patients with nonalcoholic fatty liver disease based on invasive and noninvasive methods. J Gastroenterol 2018;53:181–196.
7. 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.
8. Tapper EB, Loomba R. Noninvasive imaging biomarker assessment of liver fibrosis by elastography in NAFLD. Nat Rev Gastroenterol Hepatol 2018;15:274–282.
9. Imajo K, Kessoku T, Honda Y, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. Gastroenterology 2016;150:626–637.e627.
10. Park CC, Nguyen P, Hernandez C, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. Gastroenterology 2017;152:598–607.e2.
11. Herrmann E, de Lédinghen V, Cassinotto C, et al. Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: an individual patient data-based meta-analysis. Hepatology 2018;67:260–272.

12. Younossi ZM, Stepanova M, Rafiq N, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011; 53:1874–1882.
13. Brunt EM. Nonalcoholic steatohepatitis: definition and pathology. Semin Liver Dis 2001;21:3–16.
14. Yin M, Talwalkar JA, Glaser KJ, et al. Assessment of hepatic fibrosis with magnetic resonance elastography. Clin Gastroenterol Hepatol 2007;5:1207–1213.
15. Chen J, Talwalkar JA, Yin M, et al. Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. Radiology 2011;259:749–756.
16. Sandrin L, Fourquet B, Hasquenoph JM, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705–1713.
17. Sumida Y, Yoneda M, Hyogo H, et al. Validation of the FIB4 index in a Japanese nonalcoholic fatty liver disease population. BMC Gastroenterol 2012;12:2.
18. Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007;45:846–854.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837–845.
20. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1:30–46.
21. Furlan A, Tublin ME, Yu L, et al. Comparison of 2D shear wave elastography, transient elastography, and MR elastography for the diagnosis of fibrosis in patients with nonalcoholic fatty liver disease. AJR Am J Roentgenol 2020;214:W20–W26.
22. Sporea I, Jurchiş A, Şirli R. Can transient elastography be a reliable method for assessing liver fibrosis in nonalcoholic steatohepatitis (NASH)? Med Ultrason 2013;15:106–110.
23. Castéra L, Foucher J, Bernard PH, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. Hepatology 2010;51:828–835.
24. de Lédinghen V, Vergniol J, Foucher J, et al. Feasibility of liver transient elastography with FibroScan using a new probe for obese patients. Liver Int 2010;30:1043–1048.
25. Chang PE, Lui HF, Chau YP, et al. Prospective evaluation of transient elastography for the diagnosis of hepatic fibrosis in Asians: comparison with liver biopsy and aspartate transaminase platelet ratio index. Aliment Pharmacol Ther 2008; 28:51–61.
26. Wong VW, Vergniol J, Wong GL, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862–1871.
27. Myers RP, Pomier-Layrargues G, Kirsch R, et al. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. Hepatology 2012;55:199–208.
28. Dietrich CF, Bamber J, Berzigotti A, et al. EFSUMB guidelines and recommendations on the clinical use of liver ultrasound elastography, update 2017 (long version). Ultraschall Med 2017; 38:e16–e47.
29. Moga TV, Stepan AM, Pienar C, et al. Intra- and inter-observer reproducibility of a 2-D shear wave elastography technique and the impact of ultrasound experience in achieving reliable data. Ultrasound Med Biol 2018;44:1627–1637.
30. Hines CD, Bley TA, Lindstrom MJ, et al. Repeatability of magnetic resonance elastography for quantification of hepatic stiffness. J Magn Reson Imaging 2010;31:725–731.

Reprint requests
Address requests for reprints to: Atsushi Nakajima, MD, PhD, Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan. e-mail: nakajima-tky@umin.ac.jp; fax: +81-45-784-3546.

CRediT Authorship Contributions

Kento Imajo (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Project administration: Lead; Resources: Lead; Visualization: Lead; Writing – original draft: Lead)

Yasushi Honda (Funding acquisition: Supporting; Investigation: Equal; Methodology: Equal; Resources: Equal)

Takashi Kobayashi (Formal analysis: Supporting; Investigation: Supporting; Resources: Supporting)

Koki Nagai (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting)

Anna Ozaki (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting)

Michihiro Iwaki (Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting)

Takaomi Kessoku (Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Equal; Investigation: Supporting; Methodology: Supporting)

Yuji Ogawa (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Supporting; Investigation: Equal; Methodology: Equal; Writing – review & editing: Supporting)

Hirokazu Takahashi (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Writing – review & editing: Equal)

Yusuke Saigusa (Formal analysis: Equal; Investigation: Lead; Methodology: Equal)

Masato Yoneda (Data curation: Equal; Formal analysis: Supporting; Funding acquisition: Equal; Investigation: Equal; Supervision: Supporting; Writing – review & editing: Equal)

Hiroyuki Kirikoshi (Data curation: Supporting; Formal analysis: Supporting; Supervision: Supporting; Writing – review & editing: Supporting)

Daisuke Utsunomiya (Conceptualization: Supporting; Supervision: Equal; Writing – review & editing: Equal)

Shinichi Aishima (Conceptualization: Supporting; Project administration: Equal; Supervision: Equal; Writing – review & editing: Equal)

Satoru Saito (Conceptualization: Equal; Data curation: Equal; Funding acquisition: Equal; Supervision: Supporting; Writing – review & editing: Equal)

Atsushi Nakajima, PhD, MD (Data curation: Equal; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Project administration: Lead; Supervision: Lead; Writing – review & editing: Lead)

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by the “Step A” program of the Japan Science and Technology Agency (J.S.T.) and Kiban-B, Shingakujuturyouiki and, in part, by Grants-in-Aid from the Japanese Ministry of Health, Labour and Welfare.

Supplementary Material

Supplementary Methods

Patients. VCTE and 2D-SWE were performed on the same day. MRE was performed within 1 month of US elastography. Finally, liver biopsy was performed within 4 months of all elastographies. Patients with a history of significant alcohol intake, chronic hepatitis including viral and autoimmune hepatitis, and use of medications associated with fatty liver were excluded. Basic demographic data, including the age and sex of participants and their relevant medical history, including history of diabetes mellitus, hypertension, or dyslipidemia, were obtained from the medical records. BMI was calculated as weight in kilograms divided by the square of height in meters (kg/m^2). Platelet count and serum levels of AST, ALT, gamma-glutamyl transpeptidase, C-reactive protein, creatinine, fasting blood glucose, fasting insulin, and glycosylated hemoglobin were measured. Diabetes mellitus was defined as fasting plasma glucose concentration $\geq$126 mg/dL, self-reported history of diabetes mellitus, treatment with dietary modifications, or use of antidiabetic medications. Hypertension was defined as blood pressure $\geq$140/90 mmHg, self-reported history of hypertension, or use of antihypertensive medications. Dyslipidemia was defined as total serum cholesterol $\geq$220 mg/dL, triglycerides $\geq$150 mg/dL, or the use of antilipidemic medications.

Histopathologic evaluations. Ultrasonography-guided percutaneous liver biopsy was performed using a 16-gauge needle biopsy kit. An adequate liver biopsy sample was defined as >20 mm in length and/or with >10 portal tracts. Steatosis affecting <5%, 5%–33%, 33%–66%, and >66% of hepatocytes was classified as grade 0, 1, 2, and 3, respectively. Lobular inflammation was graded according to the number of inflammatory foci per field of view at magnification of ×200, with 0, <2, 2–4, and >4 foci per field classified as grade 0, 1, 2, and 3, respectively. Hepatocellular ballooning involving no, few, and many cells was classified as grade 0, 1, and 2, respectively. Fibrosis was classified as follows: stage 0, no fibrosis; stage 1, mild fibrosis; stage 2, moderate fibrosis; stage 3, severe fibrosis; and stage 4, cirrhosis.

Magnetic resonance elastography. MRE was performed after fasting for 12 hours. Continuous longitudinal mechanical waves (60 Hz) were generated by using a passive acoustic driver placed against the anterior chest wall. A two-dimensional spin-echo planar MRE sequence was used to acquire axial wave images with the following parameters: repetition time ms/echo time ms, 50/23; continuous sinusoidal vibration, 60 Hz; field of view, 32–42 cm; matrix size, 256 × 64; flip angle, 30°; section thickness, 10 mm; 4 evenly spaced phase offsets; and 4 pairs of 60-Hz trapezoidal motion encoding gradients with zeroth and first moment nulling along the through-plane direction. All processing steps were applied automatically without manual intervention to yield quantitative images of tissue shear stiffness in kilopascals. On each section of the image on MRE, the ROIs were drawn to include only the parenchyma of the right lobe, while avoiding the edges of the liver and large blood vessels. The mean of measurements on 4 slices was used. If there was no liver parenchyma that could be measured by elastograms using reliability maps, the study was considered invalid. ROIs also excluded regions where the phase signal-to-noise ratio (the ratio of wave amplitude to the noise in the wave images) was ˂5.

Vibration-controlled transient elastography. VCTE was also performed after fasting for 12 hours. The system is equipped with a probe and an ultrasonic transducer mounted on the axis of a vibrator. A vibration of mild amplitude and low frequency is transmitted from the vibrator to the tissue by the transducer, which induces the propagation of an elastic shear wave through the tissue. The speed of the propagating wave is estimated using a one-dimensional US technique and is automatically converted to a measurement in terms of Young's modulus in kilopascals. VCTE-based measurements of Young's modulus can be approximately compared with MRE-based measurements of shear modulus by dividing them by a factor of 3. Each patient was placed in the supine position with the right arm raised behind the head and was asked to remain still during the procedure. In this study, VCTE measurements with at least 10 valid shots and success rate of $\geq$60% with IQR/M $\leq$30% were considered reliable and used in the statistical analysis.

Two-dimensional shear wave elastography. 2D-SWE was also performed after fasting for 12 hours. Measurements were obtained via the intercostal spaces with the patients in the dorsal decubitus position and their right arm maximally abducted. All sonographers received 2D-SWE application training. The technique used to perform 2D-SWE has been previously described.[1] It combines an acoustic radiation force-induced supersonic push in the tissue by focused ultrasonic beams followed by an ultrafast ultrasound imaging sequence. Elasticity values are displayed as a real-time, color-coded, 2D quantitative SWE map of tissue stiffness expressed in kilopascals or shear wave velocity expressed in meters per second. This map is displayed in color over a conventional grayscale B-mode image. The positions of SWE map and circular ROI used for stiffness/velocity measurement are operator-adjustable. All measurements were obtained by using 10-mm diameter circular ROIs. The acquisition box size was 2–4 cm in the lateral direction and 3–5 cm in the axial direction. As suggested by the manufacturer, the measurement was performed 1–2 cm under the liver capsule in an area of the parenchyma. Circular ROIs were placed in areas of homogeneous portion of the speed image, while avoiding vascular and biliary structures. Parallel propagation contours were observed in the region of homogenous color. The mean and standard deviation within the ROI were recorded.

2D-SWE measurements with at least 10 valid shots and success rate ≥60% with IQR/M ≤30% were considered reliable and used in the statistical analysis.

Scoring systems. The values for the upper limit of normal were set according to the International Federation of Clinical Chemistry: AST levels of 35 U/L for men and 30 U/L for women (these levels were comparable with those used in other analyses); FIB-4 index: age (years) × AST (IU/L)/[platelet count (10^9/L) × $ALT^{1/2}$ (IU/L)]; NFS: −1.675+0.037 × age (years) + 0.094 × BMI (kg/m^2) + 1.13 × impaired fasting glycemia/diabetes (yes = 1, no = 0) + 0.99 × AST/ALT ratio − 0.013 × platelet (×10^9/L) − 0.66 × albumin (g/dL).

Interobserver and intraobserver reproducibility analysis. To evaluate the intraobserver reproducibility, an elastography expert performed US elastography twice on the same day. MRE-based LSM was evaluated twice by the elastography expert within a week. In addition, to evaluate the interobserver reproducibility, US elastography was performed by a trainee who was blinded to the findings obtained by the elastography expert on the same day. MRE-based LSM was also evaluated by a trainee who was blinded to the findings obtained by the elastography expert. In addition, we analyzed 70 patients in the first (n = 35) and second (n = 35) halves of the study to examine whether the interobserver reproducibility varies with the experience with all elastographies. In the first half of the study, a trainee had less than 100 examinations of VCTE and 2D-SWE, but in the second half of the study, the number of examinations increased to more than 100 examinations.

Statistical analysis. Cutoff values were calculated by using Youden index. *P* values <.05 were considered statistically significant. Univariable and multivariate logistic regression was performed to analyze the influence of factors that were discordant between the fibrosis stages predicted by elastography versus those predicted by histology. For the multivariate model, we included variables that achieved statistical significance ($P < .05$) on univariate analysis. Reproducibility based on ICC values was defined as follows: slight (ICC, 0–0.20), fair (ICC, 0.21–0.40), moderate (ICC, 0.41–0.60), substantial (ICC, 0.61-0.80), and almost perfect (ICC, 0.81-1.00).[1]

Supplementary Results

Patient characteristics. Four MRE examinations failed because of excessive iron overload (n = 2) and incorrect wave image (n = 2). VCTE and 2D-SWE failed because of low success rate (n = 6 and n = 17, respectively) and unreliable LSMs (n = 5 and n = 10, respectively) (Supplementary Table 1).

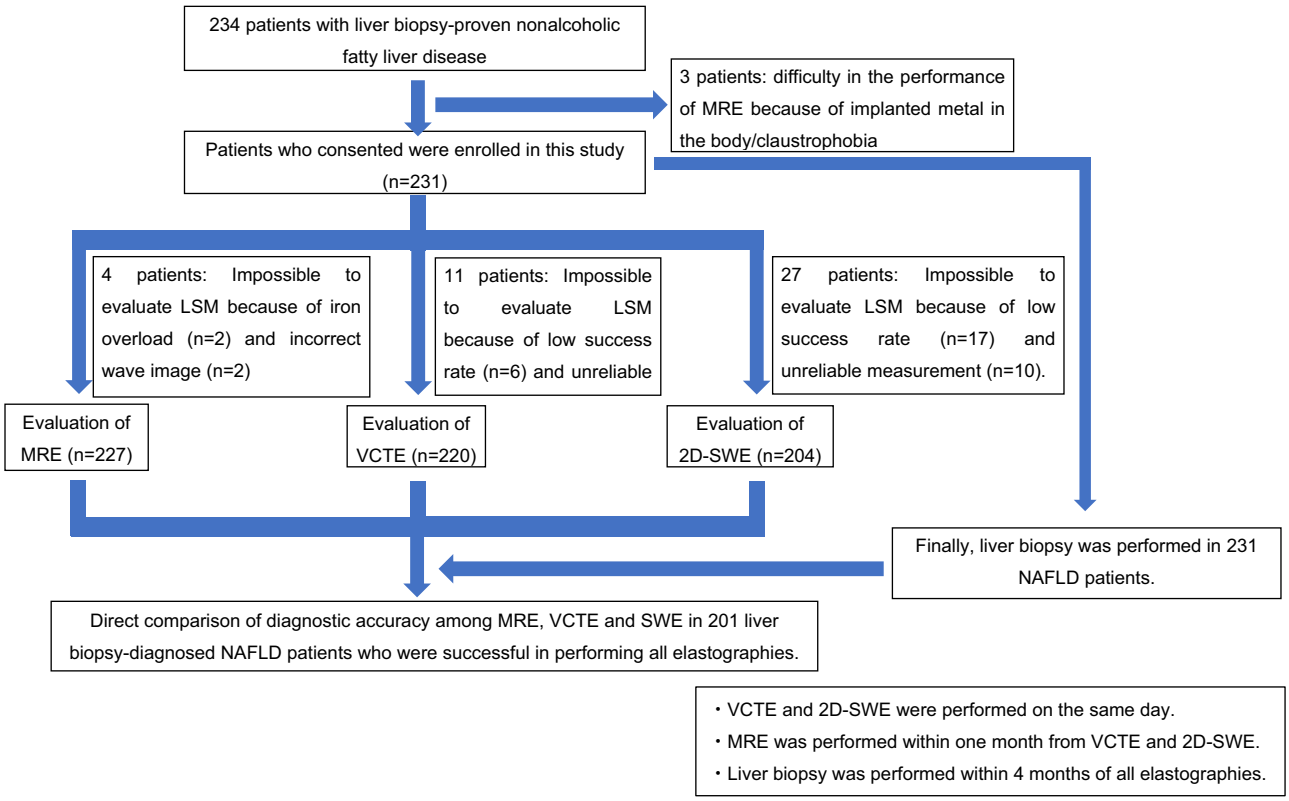
Direct comparison of the diagnostic accuracy of magnetic resonance elastography, vibration-controlled elastography, and two-dimensional shear wave elastography. LSM was assessed by using MRE, VCTE, and 2D-SWE to assess the stage of liver fibrosis in patients with NAFLD. These analyses revealed a stepwise increase in LSM obtained by using MRE, VCTE, and 2D-SWE with increasing histologic severity of liver fibrosis ($P < .001$) (Figure 1*A*). To assess the diagnostic accuracy of MRE, VCTE, and 2D-SWE for liver fibrosis stage, we calculated AUROC curve and cutoff values in patients with NAFLD. The ROC curves to distinguish liver fibrosis stage ≥1, stage ≥2, stage ≥3, and stage 4 on the basis of LSM according to MRE, VCTE, and 2D-SWE are presented in Figure 1*B*. The 95% confidence intervals of AUROC values, cutoff levels, and sensitivity, specificity, PPV, and NPV in diagnosing liver fibrosis stage ≥1, ≥2, ≥3, and ≥4 using MRE, VCTE, and 2D-SWE are presented in Supplementary Table 2.

Next we assessed the relationship between steatosis, lobular inflammation, or ballooning and LSM obtained by using MRE, VCTE, or 2D-SWE in patients with NAFLD. There was a significant association between LSM and the grade of ballooning in all modalities (Supplementary Figure 4). On multivariate regression analysis, the relationship between histologic parameters and LSM values obtained by the 3 elastography modalities was evaluated (Supplementary Table 3). We found that only liver fibrosis stage was significantly associated with LSM.

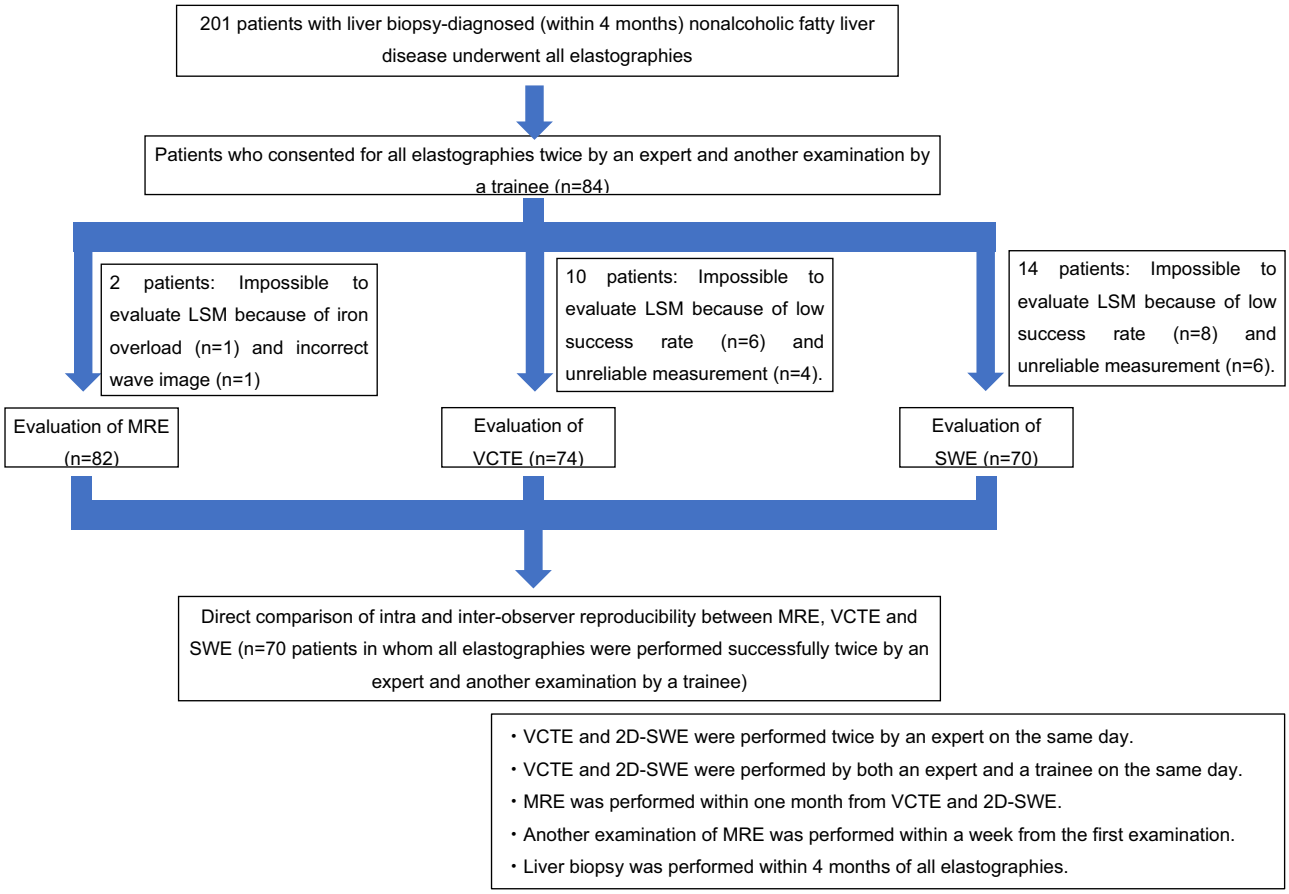
Interobserver reproducibility analysis. There was no difference in ICC for interobserver reproducibility for MRE between the 35 patients in the first half and those in the second half. However, interestingly, ICC of the interobserver reproducibility for VCTE and 2D-SWE was higher in the 35 patients in the second half than in patients in the first half (Supplementary Table 6).

Supplementary Reference

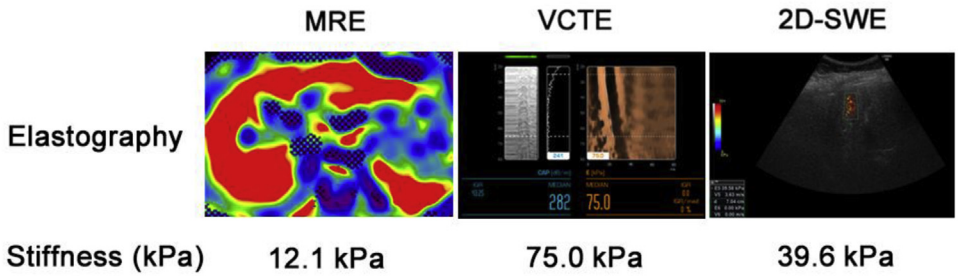
1. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.



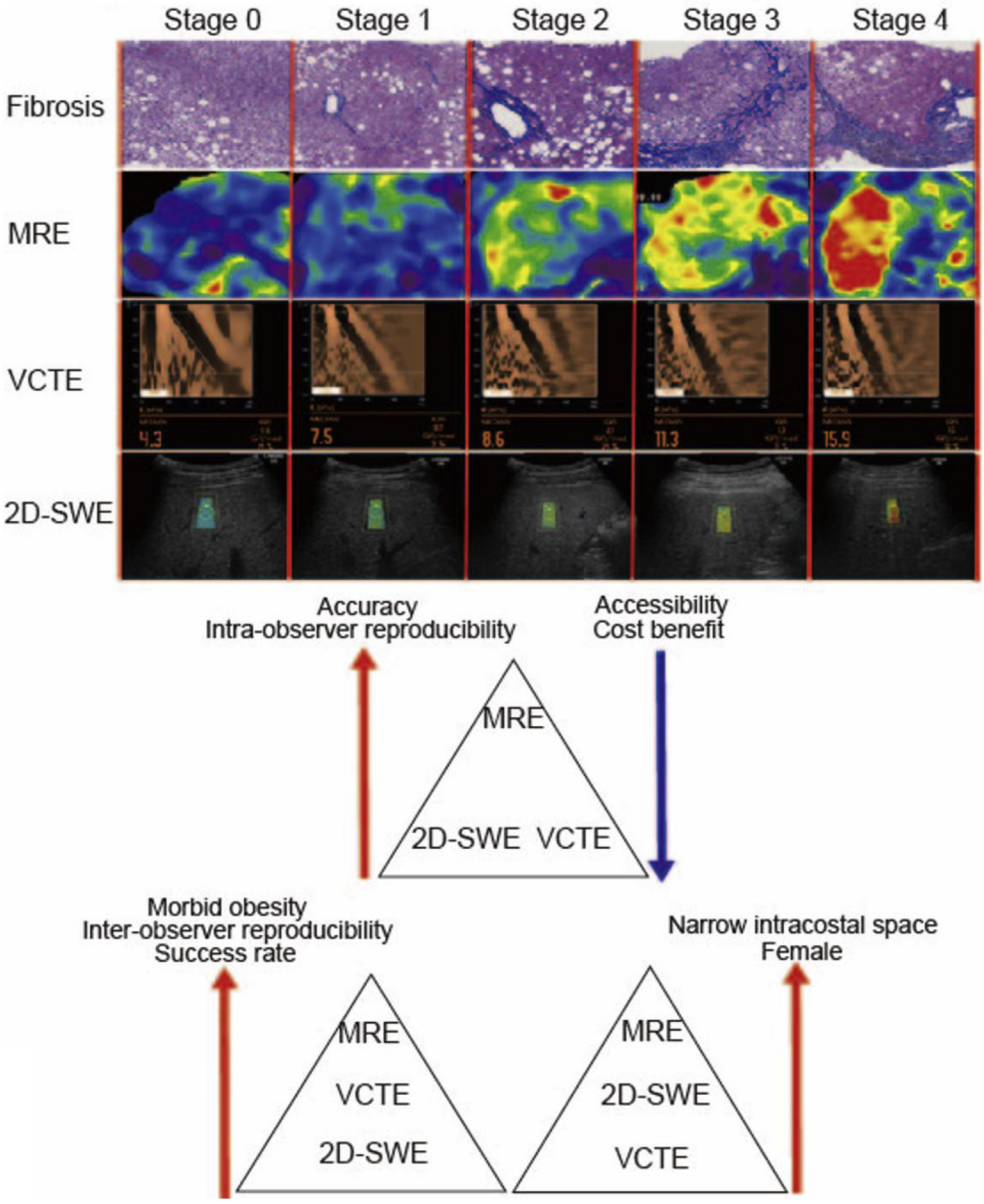
Supplementary Figure 1. Study protocol. LSM, liver stiffness measurement; MRE, magnetic resonance enterography; NAFLD, nonalcoholic liver disease; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.



Supplementary Figure 2. Study protocol for the analysis of intraobserver and interobserver reproducibility. LSM, liver stiffness measurement; MRE, magnetic resonance enterography; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.



Supplementary Figure 3. Patient with cirrhosis with impaired liver dysfunction (Child–Pugh grade 6). Pathologic fibrosis and liver stiffness measurement (LSM) values on magnetic resonance elastography (MRE), vibration-controlled transient elastography (VCTE), and two-dimensional shear wave elastography (2D-SWE) are presented. The consequences are immediately apparent and serious.



Supplementary Figure 4. Graphic abstract. Although magnetic resonance elastography (MRE) is superior to vibration-controlled transient elastography (VCTE) and two-dimensional shear wave elastography (2D-SWE) in diagnosing liver fibrosis, especially stage 4, in patients with nonalcoholic fatty liver disease (NAFLD), the costs and accessibility are better with VCTE and 2D-SWE than MRE (MRE > VCTE = 2D-SWE). Intraobserver reproducibility is also superior to VCTE and 2D-SWE (MRE > VCTE = 2D-SWE). Interobserver reproducibility and success rate for morbid obesity appear to decrease in the order of MRE, VCTE, and 2D-SWE (MRE > VCTE > 2D-SWE). In contrast, the diagnostic ability appears to decrease in the order of MRE, 2D-SWE, and VCTE (MRE > 2D-SWE > VCTE) in female patients and those with narrow intercostal space.

Supplementary Table 1. Cases in Which It Was not Possible to Measure LSM using MRE, VCTE, and 2D-SWE

Case	MRE (n = 4)	VCTE (n = 11)	2D-SWE (n = 27)
1	Iron overload		—
2	Iron overload		—
3	Incorrect wave image (morbid obesity)	Low success rate (morbid obesity)	Unreliable measurement (morbid obesity)
4	Incorrect wave image (morbid obesity)	Low success rate (morbid obesity)	Unreliable measurement (morbid obesity)
5	—	Low success rate (morbid obesity)	Unreliable measurement (morbid obesity)
6	—	Low success rate (narrow intercostal space)	—
7	—	Low success rate (Chilaiditi syndrome)	Low success rate (Chilaiditi syndrome)
8	—	Low success rate (Chilaiditi syndrome)	Low success rate (Chilaiditi syndrome)
9	—	Unreliable measurement (narrow intercostal space)	Unreliable measurement (narrow intercostal space)
10	—	Unreliable measurement (narrow intercostal space)	—
11	—	Unreliable measurement (narrow intercostal space)	—
12	—	Unreliable measurement (narrow intercostal space)	—
13	—	Unreliable measurement (narrow intercostal space)	—
14	—	—	Unreliable measurement (morbid obesity)
15	—	—	Unreliable measurement (morbid obesity)
16	—	—	Unreliable measurement (increase in SCD)
17	—	—	Unreliable measurement (increase in SCD)
18	—	—	Unreliable measurement (increase in SCD)
19	—	—	Unreliable measurement (increase in SCD)
20	—	—	Low success rate (morbid obesity)
21	—	—	Low success rate (morbid obesity)
22	—	—	Low success rate (morbid obesity)
23	—	—	Low success rate (morbid obesity)
24	—	—	Low success rate (morbid obesity)
25	—	—	Low success rate (morbid obesity)
26	—	—	Low success rate (increase in SCD)
27	—	—	Low success rate (increase in SCD)
28	—	—	Low success rate (increase in SCD)
29	—	—	Low success rate (increase in SCD)
30	—	—	Low success rate (increase in SCD)
31	—	—	Low success rate (increase in SCD)
32	—	—	Low success rate (increase in SCD)
33	—	—	Low success rate (increase in SCD)
34	—	—	Low success rate (increase in SCD)

MRE, magnetic resonance elastography; SCD, skin-capsule distance; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.

Supplementary Table 2. Diagnostic Accuracy of MRE, VCTE, and 2D-SWE in Detecting Liver Fibrosis Stages in Patients With NAFLD

	LSM (n = 201)																				
	MRE							VCTE							2D-SWE						
Fibrosis stage	Cutoff level (*kPa*)	AUROC	95% CI	Se	Sp	PPV	NPV	Cutoff level (*kPa*)	AUROC	95% CI	Se	Sp	PPV	NPV	Cutoff level (*kPa*)	AUROC	95% CI	Se	Sp	PPV	NPV
≥1	2.92	0.947	0.863–0.980	78.2	100.0	100.0	82.4	5.00	0.952	0.910–0.974	91.4	100.0	100.0	64.3	6.35	0.923	0.851–0.962	82.5	100.0	100.0	80.5
≥2	3.19	0.927	0.866–0.961	90.1	81.3	92.4	78.5	8.40	0.882	0.823–0.931	86.0	74.2	89.4	67.6	7.55	0.910	0.843–0.951	87.1	85.9	93.8	73.1
≥3	3.90	0.937	0.882–0.958	82.5	91.5	91.7	82.1	9.70	0.924	0.867–0.947	83.6	83.3	85.1	81.7	8.88	0.920	0.865–0.953	87.0	87.8	89.3	84.9
≥4	4.62	0.923	0.871–0.955	95.2	75.0	46.5	98.5	12.40	0.872	0.807–0.917	90.2	74.6	45.1	97.1	9.98	0.886	0.836–0.925	91.9	75.5	46.6	97.6

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; MRE, magnetic resonance elastography; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity; 2D-SWE, two-dimensional elastography; VCTE, vibration-controlled transient elastography.

Supplementary Table 3. Multiple Regression Analysis of Histologic Parameters Associated With LSM for MRE, VCTE, or 2D-SWE

	LSM (n = 201)					
	MRE		VCTE		2D-SWE	
Parameters	Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
Fibrosis	1.09 ± 0.07	<.001[a]	4.12 ± 0.36	<.001[a]	1.97 ± 0.19	<.001
Steatosis	–0.18 ± 0.09	.057	0.54 ± 0.52	.301	0.30 ± 0.26	.255
Inflammation	0.05 ± 0.12	.666	–1.14 ± 0.70	.105	–0.18 ± 0.36	.611
Ballooning	0.04 ± 0.12	.721	0.00 ± 0.69	.997	–0.10 ± 0.36	.785

LSM, liver stiffness measurement; MRE, magnetic resonance elastography; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.
[a]Significant at $P < .05$.

Supplementary Table 4. Clinical, Serologic, and Histologic Characteristics of Patients With NAFLD for Whom the Intraobserver and Interobserver Reproducibility Was Evaluated

	NAFLD
Number (n)	70
Age (*y*)	60.0 (51.5–71.5)
Sex (male; female)	47; 23
Body mass index (kg/m^2)	27.5 (24.8–30.3)
Platelets ($/10^4\ \mu L$)	20.9 (17.7–25.7)
AST (*IU/L*)	43.0 (32.0–52.5)
ALT (*IU/L*)	44.0 (32.5–70.5)
GGT (*IU/L*)	49.0 (35.0–98.5)
CRP (*mg/L*)	0.11 (0.06–0.21)
Cr (*mg/dL*)	0.78 (0.55–1.17)
FBS (*mg/dL*)	108.0 (101.0–131.0)
Fasting insulin ($\mu U/mL$)	15.1 (12.3–18.0)
HbA1c (%)	6.1 (5.6–7.0)
DM (%)	47 (67.1)
HT (%)	36 (51.4)
DLP (%)	53 (75.7)
Steatosis grade (n)	
5%–33%	32
33%–66%	28
>66%	10
Lobular inflammation (n)	
None	1
<2 foci per ×200 field	41
2–4 foci per ×200 field	24
>4 foci per ×200 field	4
Liver cell ballooning (n)	
None	15
Few balloon cells	40
Many balloon cells	15
NAFL/NASH (n)	15/55
NAFLD activity score (NAS) (n)	
1/2/3/4/5/6/7	1/5/11/24/13/9/5/2
Fibrosis stage (n)	
None	3
Perisinusoidal or periportal	19
Perisinusoidal and portal/periportal	15
Bridging fibrosis	23
Cirrhosis	10

NOTE. Values are median (interquartile range) or n (%).
ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; Cr, creatinine; CRP, C-reactive protein; DLP, dyslipidemia; DM, diabetes mellitus; FBS, fasting blood glucose; GGT, gamma-glutamyl transferase; HbA1c, glycosylated hemoglobin; HT, hypertension; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis.

Supplementary Table 5. Diagnostic Accuracy of MRE, VCTE, and 2D-SWE in Detecting Liver Fibrosis Stages by an Elastography Expert and an Elastography Trainee in a Subgroup of 70 Patients

	Operator 1 (experienced)																			
N = 70	MRE							VCTE							2D-SWE					
Fibrosis stage	Cutoff level (kPa)	AUROC	95% CI	*P* value	vs MRE *P* value	vs. VCTE *P* value	vs SWE *P* value	Cutoff level (*kPa*)	AUROC	95% CI	*P* value	vs MRE *P* value	vs. VCTE *P* value	vs SWE *P* value	Cutoff level (*kPa*)	AUROC	95% CI	*P* value	vs MRE *P* value	vs VCTE *P* value	vs SWE *P* value
≥2	2.89	0.947	0.806–0.987	<.001[a]	—	.197	.738	8.50	0.906	0.763–0.967	<.001[a]	.197	—	.482	6.97	0.935	0.819–0.979	<.001[a]	.738	.482	—
≥3	3.98	0.951	0.874–0.982	<.001[a]	—	.151	.364	9.99	0.903	0.882–0.952	<.001[a]	.151	—	.622	8.88	0.920	0.811–0.969	<.001[a]	.364	.622	—
≥4	5.05	0.912	0.815–0.961	<.001[a]	—	.184	.169	15.1	0.884	0.740–0.963	<.001[a]	.184	—	.849	11.1	0.874	0.746–0.950	<.001[a]	.169	.849	—

	Operator 2 (trainee)																			
N = 70	MRE							VCTE							2D—SWE					
Fibrosis stage	Cutoff level (*kPa*)	AUROC	95% CI	*P* value	vs MRE *P* value	vs VCTE *P* value	vs SWE *P* value	Cutoff level (*kPa*)	AUROC	95% CI	*P* value	vs MRE P value	vs VCTE P value	vs SWE P value	Cutoff level (*kPa*)	AUROC	95% CI	*P* value	vs MRE *P* value	vs VCTE *P* value	vs SWE *P* value
≥2	3.10	0.907	0.764–0.967	<.001[a]	—	.949	.790	8.20	0.904	0.759–0.965	<.001[a]	.742	—	.751	8.25	0.918	0.813–0.966	<.001[a]	.790	.751	—
≥3	3.79	0.908	0.813–0.957	<.001[a]	—	.297	.805	11.0	0.857	0.723–0.932	<.001[a]	.297	—	.467	8.48	0.897	0.773–0.957	<.001[a]	.805	.467	—
≥4	5.13	0.917	0.811–0.966	<.001[a]	—	.106	.412	14.3	0.857	0.812–0.914	<.001[a]	.106	—	.833	9.91	0.869	0.743–0.939	<.001[a]	.412	.833	—

AUROC, area under the receiver operating characteristic curve; CI, confidence interval; MRE, magnetic resonance elastography; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.

[a]Significant at $P < .05$.

Supplementary Table 6. Interobserver Reproducibility of MRE, VCTE, and 2D-SWE for Detecting Liver Fibrosis in NAFLD Patients (Total; n = 70) in the First (n = 35) and Second (n = 35) Halves of the Study

	First (n = 35)		Second (n = 35)	
Elastography	ICC	95% CI	ICC	95% CI
MRE	0.9511	0.9202–0.9685	0.9457	0.9134–0.9612
VCTE	0.7285	0.6824–0.7752	0.8443	0.7922–0.8640
2D-SWE	0.6608	0.5645–0.7711	0.7561	0.7029–0.8061

CI, confidence interval; ICC, intraclass correlation coefficient; MRE, magnetic resonance elastography; NAFLD, nonalcoholic fatty liver disease; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.

Supplementary Table 7. Advantages and Disadvantages of MRE, VCTE, and 2D-SWE

	MRE	VCTE	2D-SWE
Sampling volume of liver	Much[a]	Little	Moderate
HCC screening	Good[a]	Impossible	Possible
Convenience of use	Poor	Good	Very good[a]
Interoperator reproducibility	Excellent[a] ICC; 0.9481	Good ICC; 0.7900	Good ICC; 0.7053
Intraoperator reproducibility	Excellent[a] ICC; 0.9197	Good ICC; 0.8369	Good ICC; 0.8469
Evaluation of liver fat accumulation	Available using PDFF Good[a]	Available using VCTE-based CAP But the diagnostic accuracy is insufficient compared with PDFF	Impossible
Ascites	Available if ascites is a little[a]	Impossible	Available if ascites is a little[a]
Morbid obesity (BMI >30 kg/m^2)	Good[a]	Possible by using XL probe	Not good
Measurements of iron deposition	Available[a]	Not available	Not available
Contraindications	Biocompatible metal Pregnancy	No[a]	No[a]
Cost	High	Low[a]	Low[a]
Available institutions	Few	Moderate[a]	A lot in future

BMI, body mass index; CAP, controlled attenuation parameter; HCC, hepatocellular carcinoma; ICC, intraclass correlation coefficient; MRE, magnetic resonance elastography; PDFF, proton density fat fraction; 2D-SWE, two-dimensional shear wave elastography; VCTE, vibration-controlled transient elastography.
[a]Significant at <.05.