Clinical Gastroenterology and Hepatology 2015;13:772–779

Accuracy of Fibroscan, Compared With Histology, in Analysis of Liver Fibrosis in Patients With Hepatitis B or C: A United States Multicenter Study

Nezam H. Afdhal,[*] Bruce R. Bacon,[‡] Keyur Patel,[§] Eric J. Lawitz,[‖] Stuart C. Gordon,[¶] David R. Nelson,[#] Tracy L. Challies,[*] Imad Nasser,[*] Jyotsna Garg,[§] Lee-Jen Wei,[**] and John G. McHutchison[§]

*Department of Gastroenterology and Hepatology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts; ‡St Louis University Liver Center, St Louis University School of Medicine, St Louis, Missouri; §Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina; ‖The Texas Liver Institute, University of Texas Health Science Center, San Antonio, Texas; ¶Division of Gastroenterology and Hepatology, Henry Ford Hospital, Detroit, Michigan; #Division of Gastroenterology, Hepatology, and Nutrition, University of Florida College of Medicine, Gainesville, Florida; and **Harvard School of Public Health, Boston, Massachusetts

BACKGROUND & AIMS: **Liver biopsy is invasive and associated with complications, sampling errors, and observer variability. Vibration-controlled transient elastography (VCTE) with FibroScan can be used to immediately assess liver stiffness. We aimed to define optimal levels of liver stiffness to identify patients with chronic viral hepatitis and significant fibrosis, advanced fibrosis, or cirrhosis.**

METHODS: **In a prospective, 2-phase study, patients with chronic hepatitis C or B underwent VCTE followed by liver biopsy analysis from January 2005 through May 2008 at 6 centers in the United States. In phase 1 we identified optimal levels of liver stiffness for identification of patients with stage F2−F4 or F4 fibrosis (the development phase, n = 188). In phase 2 we tested these cutoff values in a separate cohort of patients (the validation phase, n = 560). All biopsies were assessed for METAVIR stage by a single pathologist in the phase 1 analysis and by a different pathologist in the phase 2 analysis. Diagnostic performances of VCTE were assessed by area under the receiver operating characteristic curve (AUROC) analyses.**

RESULTS: **In phase 1 of the study, liver stiffness measurements identified patients with ≥F2 fibrosis with AUROC value of 0.89 (95% confidence interval, 0.83–0.92) and identified patients with F4 fibrosis with AUROC value of 0.92 (95% confidence interval, 0.87–0.95). Liver stiffness cutoff values (kPa) in phase 1 were 8.4 for ≥F2 (82% sensitivity, 79% specificity) and 12.8 for F4 (84% sensitivity, 86% specificity). In the phase 2 analysis, the liver stiffness cutoff values identified patients with ≥F2 fibrosis with 58% sensitivity ($P < .0001$ vs phase 1) and 75% specificity (nonsignificant difference vs phase 1); they identified patients with F4 fibrosis with 76% sensitivity ($P < .0001$ vs phase 1) and 85% specificity (nonsignificant differences vs phase 1). VCTE had an interobserver agreement correlation coefficient of 0.98 (n = 26) and an intraobserver agreement correlation coefficient of 0.95 (n = 34).**

CONCLUSIONS: **In a large U.S. multicenter study, we confirmed that VCTE provides an accurate assessment of liver fibrosis in patients with chronic viral hepatitis. Our findings are similar to those from European and Asian cohorts.**

Keywords: Liver Disease; HBV; HCV; Diagnosis; Diagnostic.

See editorial on page 780; see similar article on page 763.

Liver inflammation and cellular injury lead to fibrosis with progression to cirrhosis and complications of decompensated end-stage liver disease such as hepatocellular carcinoma. Currently, liver biopsy is the

Abbreviations used in this paper: APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; AUROC, area under receiver operating characteristic curve; BMI, body mass index; CHC, chronic hepatitis C; ICC, intraclass correlation coefficient; LSM, liver stiffness measurement; VCTE, vibration-controlled transient elastography.

© 2015 by the AGA Institute
1542-3565/$36.00
http://dx.doi.org/10.1016/j.cgh.2014.12.014

reference technique for determining the extent of hepatic fibrosis and inflammation. However, the procedure is invasive and can result in occasional significant complications. Limitations of biopsy also include variability in tissue sampling and interobserver and intraobserver variability that result in incorrect staging of disease.[1] Because of the limitations of several liver biopsy imaging methods that are based on the principle of elastography have been evaluated for staging liver disease, which aims at measuring the stiffness of the liver. Several manufacturers have developed technologies evaluating liver stiffness such as the Virtual Touch Tissue Quantification system[2] (Siemens, Munich, Germany), the Shear Wave Elastography system[3] (SuperSonic Imagine, Aix en Provence, France), or the Vibration-Controlled Transient Elastography system (VCTE) implemented on the FibroScan[4] (Echosens, Paris, France). VCTE using the FibroScan device is the most validated and commonly used elastography method worldwide and was recently approved in the United States by the Food and Drug Administration.[5,6] This technology is based on a rapid measure of shear wave velocity and subsequent calculation of liver stiffness, which correlates with severity of fibrosis. Data suggest VCTE is reliable in diagnosing cirrhosis in patients with chronic liver disease,[7] advanced fibrosis in patients with alcoholic and nonalcoholic fatty liver disease,[8,9] and significant fibrosis in patients with chronic hepatitis C (CHC)[10] and in biliary diseases.[11] However, some factors such as patient body mass index (BMI) were reported to be associated with lower applicability of VCTE, but they caused unreliable measurements or examination failures.[12] Despite this limitation, a meta-analysis of 50 studies evaluating VCTE in comparison with liver biopsy as a reference showed that this technique has good diagnostic accuracy in detecting cirrhosis, regardless of the underlying cause of liver disease.[13]

In addition to these imaging techniques, other noninvasive methods to assess fibrosis are based on a biological approach that uses direct and indirect blood markers.[14–17] Among them, aspartate aminotransferase-to-platelet ratio index (APRI), which is based on aspartate aminotransferase (AST) and platelets, and FIB-4, which is based on age, AST, alanine aminotransferase, and platelets, are commonly used because the required blood parameters are inexpensive and routinely assessed for the management of patients with chronic liver disease. In addition, both APRI and FIB-4 exhibit good diagnostic performance for exclusion of cirrhosis in CHC patients.[18–20]

The primary objective of the study was to (1) identify optimal liver stiffness measurement (LSM) cutoff values for staging significant fibrosis, advanced fibrosis, and cirrhosis in a development cohort of U.S. patients with chronic viral hepatitis and (2) to validate these LSM cutoff points in an independent validation cohort. Secondary objectives were to (1) assess the intraoperator and interoperator reproducibility of LSM performed by VCTE, (2) identify the factors independently associated with LSM, (3) evaluate the potential influence of patient's BMI on the diagnostic performances of VCTE for significant fibrosis and cirrhosis assessment, and (4) to compare the diagnostic performance of VCTE versus the fibrosis biomarkers APRI and FIB-4 in the validation cohort.

Methods

Consecutive adult male or female patients with chronic hepatitis B or CHC who were undergoing liver biopsy were prospectively included in this study. Enrollment was from January 2005 through May 2008 at 6 centers in the United States.

Study Design

The study was conducted in 2 phases. Phase 1 was designed to identify the optimal LSM thresholds to stage significant liver fibrosis ($\geq$ F2), advanced fibrosis ($\geq$F3), and cirrhosis (F4). Phase 2 was designed to validate the selected LSM thresholds from phase 1. Assessment included interobserver and intraobserver variations in LSM, and liver biopsy served as the reference in staging fibrosis or cirrhosis.

The time between the FibroScan reading and the biopsy was not to exceed 6 months for phase 1 and 6 weeks for phase 2. The FibroScan operator was blinded to the fibrosis stage, and only the study pathologist, data center (Duke Clinical Research Institute), and sponsor had access to the centralized liver biopsy results.

Vibration-Controlled Transient Elastography

LSMs were performed by using FibroScan device powered by VCTE (Echosens) as previously described (Supplementary Materials), equipped with the standard M probe.

Intraoperator and Interoperator Variability of Liver Stiffness

Intraoperator and interoperator variability of LSM was performed on a subgroup of patients randomly selected from phase 2. For interoperator variability analysis, 2 LSMs were performed by 2 separate trained operators before liver biopsy on the same day and in the same anatomic location. Subjects enrolled in the intraoperator analysis had a second examination performed within 6 weeks by the same operator. In both interoperator and intraoperator analyses, the initial LSMs were considered the efficacy data, and the second measurements were variability data.

Liver Biopsy

All liver biopsies were evaluated by the central pathology lab at Beth Israel Deaconess Medical Center according

to the METAVIR scoring system.[21] The protocol specified that biopsies should be performed by using an 18-gauge cutting needle (TruCut), and the criteria for adequacy are given in the Supplementary Materials. Biopsies in phase 1 were read by I.N. and in phase 2 by either I.N. or T.C.

Biological Parameters

Basic biochemical measures were taken at a maximum within 3 months of biopsy and FibroScan: alanine aminotransferase, AST, alkaline phosphatase,

Table 1. Characteristics of Patients in Matched Population for Phase 1 And Phase 2

Characteristics	Phase 1	Phase 2	Statistical difference
Total	188	560	—
Gender (male)	130 (69%)	361 (64%)	1.37 (.24)
Age (*y*)	48.1 (9.3)	49.9 (8.8)	0.02
Height (*cm*)	174 (9.7)	173 (9.7)	0.24
Weight (*kg*)	81.1 (16.9)	80 (9.0)	0.40
BMI (kg/m^2)	26.7 (4.2)	26.6 (4.3)	0.96
Waist circumference (*cm*)	94 (13.4)	91 (13.4)	0.005
White	149 (79%)	437 (78%)	0.12 (.73)
Hispanic or Latino	5 (3%)	49 (9%)	7.80 (.005)
Hepatitis C virus	179 (95%)	521 (92%)	1.11 (.29)
Hepatitis B virus	10 (5%)	43 (8%)	1.19 (.28)
Human immunodeficiency virus	9 (5%)	15 (3%)	2.02 (.17)
Diabetes	18 (10%)	66 (12%)	0.69 (.41)
Arterial hypertension	57 (30%)	169 (30%)	0.001 (.97)
Rheumatoid arthritis	6 (3%)	30 (5%)	1.44 (.23)
Osteoarthritis	18 (10%)	49 (9%)	0.12 (.73)
Psoriasis	4 (2%)	11 (2%)	0.02 (.89)
Gout	5 (3%)	10 (2%)	0.55 (.46)
Hepatomegaly	34 (18%)	24 (4%)	40.70 (<.001)
Splenomegaly	7 (4%)	3 (1%)	10.01 (.007)
Spider nevi	17 (9%)	7 (1%)	28.18 (<.001)
History of alcohol usage	162 (86%)	452 (81%)	3.46 (.18)
Albumin (*g/dL*)	4.6 (3.1)	4.9 (5.0)	0.40
Alkaline phosphatase (*IU/L*)	85.7 (35.0)	91.6 (47.6)	0.13
Alpha fetoprotein (*ng/mL*)	15.4 (69.1)	8.5 (21.5)	0.11
Alanine aminotransferase/serum glutamic pyruvic transaminase (*IU/L*)	90.2 (94.5)	86.8 (79.8)	0.63
AST/serum glutamic oxaloacetic transaminase (*IU/L*)	69.7 (65.7)	65.9 (59.8)	0.46
Platelets ($10^3/mm^3$)	214 (70)	222 (78)	0.24
International normalized ratio	1.07 (0.47)	1.35 (2.00)	0.06
Total bilirubin (*mg/dL*)	0.65 (0.50)	0.80 (3.02)	0.49
METAVIR fibrosis stages			47.7 (<.001)
F0	20 (10.6%)	28 (5%)	
F1	85 (45.2%)	159 (28.4%)	
F2	23 (12.2%)	185 (33%)	
F3	22 (11.7%)	105 (18.8%)	
F4	38 (20.3%)	83 (14.8%)	
METAVIR activity			17.1 (.001)
A0	6 (3%)	5 (1%)	
A1	122 (65%)	294 (53%)	
A2	57 (30%)	255 (45%)	
A3	3 (2%)	6 (1%)	
Steatosis stages			7.9 (.05)
None (S0)	146 (77.7%)	398 (71.1%)	
5%–30% (S1)	39 (30.7%)	124 (22.1%)	
30%–60% (S2)	3 (1.6%)	37 (6.6%)	
>60% (S3)	0 (0.0%)	1 (0.2%)	
Study sites (%)			
1	102 (54.3%)	144 (25.7%)	
2	30 (16.0%)	48 (8.6%)	
3	56 (30.0%)	11 (19.6%)	
4		80 (14.3%)	
5		95 (17.0%)	
6		83 (14.8%)	

NOTE. Data expressed as mean (standard deviation) or N (percentage). Statistical difference displays the result of *t* test for quantitative variables (*P* value) or of χ^2 test for proportion (χ^2 (*P* value)).

γ-glutamyltransferase, prothrombin time, albumin, total bilirubin, platelet count, and α-fetoprotein (if available).

In patients with complete biochemical data, the performance of VCTE was compared with that of APRI and FIB-4. The APRI was calculated as follows: ASTx (upper limit of normal)/platelet count ($\times 10^9$/L) × 100.[19] FIB-4 was calculated as age [y] × AST [U/L])/((PLT [10^9/L]) × (ALT [U/L])$^{1/2}$).[18]

Statistics

All statistical analyses were performed by using R software (R Development Core Team 2008) and IBM SPSS (Armonk, NY), and P value $< .05$ was considered significant. Details of the statistical analysis are given in the Supplementary Materials.

Results

Overall, 907 patients were enrolled in this study. The study flow sheet is shown in Supplementary Figure 1. Patient baseline characteristics are listed in Table 1. Average time between LSM and biopsy was 19.1 ± 34.7 days. Both phase 1 and phase 2 populations exhibited similar clinical and biochemical characteristics, except patients in phase 1 were younger ($P = .02$), had a larger waist circumference ($P = .005$), and were more likely to have hepatomegaly ($P < .001$), splenomegaly ($P = .007$), or spider nevi ($P < .001$). Unreliable LSM rate was also significantly different between phase 1 and phase 2 (4.2% vs 12.3%, respectively, $P < .001$).

Figure 1 displays the box plots of LSM by METAVIR fibrosis stage for both cohorts.

In phase 1, LSM could discriminate F3 versus F4 patients ($P = .0014$) but not F1 versus F2 patients and F2 versus F3 patients ($P > .05$ for both). In phase 2, LSM could discriminate F2 versus F3 patients ($P = .015$) and F3 versus F4 patients ($P < .001$) but not F1 versus F2 patients ($P > .05$). According to Kruskal–Wallis test, LSM was significantly different across fibrosis stages in both populations ($P < .001$). Mean LSM was significantly different between phase 1 and phase 2 for METAVIR stage F2 (12.0 vs 8.4 kPa, $P < .001$) and when taking all fibrosis stages together (13.6 vs 11.3 kPa, $P = .025$).

Reproducibility of Vibration-Controlled Transient Elastography

FibroScan VCTE examination was performed on 2 occasions within 6 weeks in a subgroup of 34 patients;

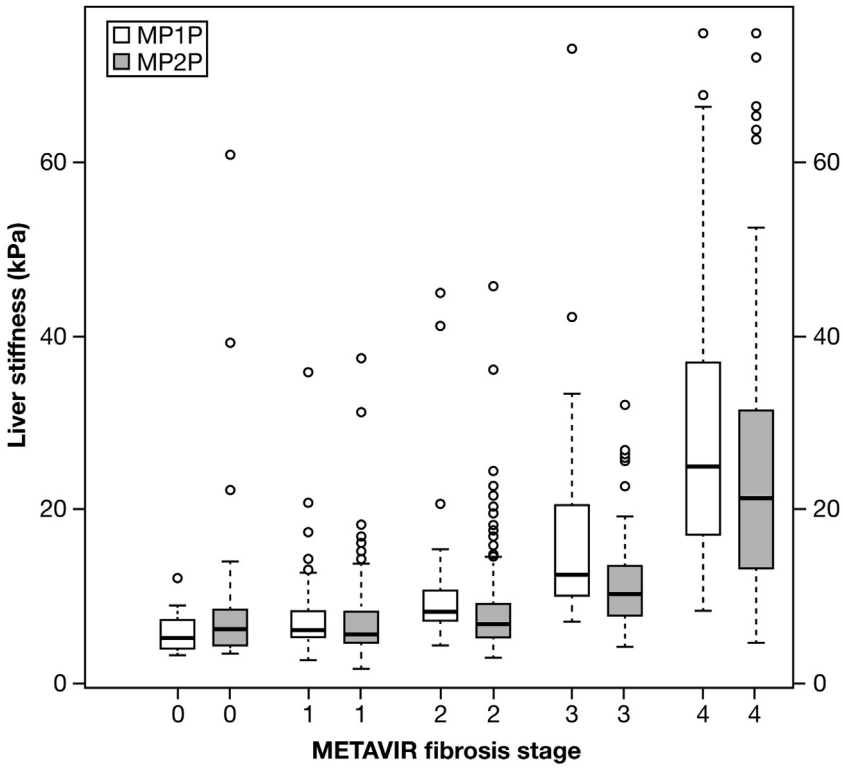


Figure 1. Box plot of LSM per METAVIR fibrosis stage, with associated P values for statistical difference of LSM between each consecutive stage. MP1P, matched population phase 1; MP2P, matched population phase 2.

Differences of LSM between consecutive fibrosis stage				
METAVIR stage	**F0-F1**	**F1-F2**	**F2-F3**	**F3-F4**
***P*-value (phase 1)**	.97 (ns)	.38 (ns)	.25 (ns)	**.0014**
***P*-value (phase 2)**	.35 (ns)	.55 (ns)	**.0148**	**<.0001**

the intraoperator intraclass correlation coefficient (ICC) was 0.95 (95% confidence interval, 0.89–0.97). For LSM performed by 2 different operators in a subgroup of 26 patients, the interoperator ICC was 0.98 (95% confidence interval, 0.95–0.99).

Diagnostic Performances of Vibration-Controlled Transient Elastography and Optimal Thresholds (Phase 1)

Diagnostic performances (areas under receiver operating characteristic curves [AUROCs]) of VCTE for diagnosing F≥2 were good (95% confidence interval, 0.89; 95% confidence interval, 0.83–0.92) and excellent for both diagnosis of F≥3 and F4 with AUROCs of 0.92 (95% confidence interval, 0.87–0.95) and 0.92 (95% confidence interval, 0.87–0.95), respectively. Optimal stiffness cutoff values (kPa) for maximizing the sum of sensitivity and specificity were 8.4 kPa for F≥2 (sensitivity, 82%; specificity, 79%), 9.6 kPa for F≥3 (sensitivity, 88%; specificity, 82%), and 12.8 kPa for F4 (sensitivity, 84%, specificity, 86%). AUROCs of VCTE and cutoffs for discriminating between each METAVIR fibrosis stage are given in Table 2. Adjusted AUROC by using the Obuchowski method (0.89 ± 0.01) confirmed this excellent performance.

Application of Predefined Thresholds on the Validation Cohort (Phase 2)

LSM thresholds defined in phase 1 (N = 188) applied to the phase 2 cohort (N = 560) exhibited sensitivity and specificity of 58% and 75% for diagnosing F≥2, 72% and 80% for diagnosing F≥3, and 76% and 85% for diagnosing F4, respectively (Table 2).

There were no statistically significant differences in sensitivity and specificity in phase 2 patients when applying the predefined cutoffs from phase 1, except for the diagnosis of F≥2, for which sensitivity decreased from 81.9% to 57.9% ($P = .042$, Table 2). Similarly, there was no significant difference in terms of well-classified patients between both phases 1 and 2, except for the diagnosis of F≥2 (diagnostic accuracy of 80.3% vs 70.5%, $P = .012$).

Factors Independently Associated With Liver Stiffness Measurement in Phase 1 and Phase 2

Multiple regression analysis in the phase 1 cohort (N = 188) indicated that LSM was independently associated with METAVIR fibrosis stages ($P < .001$), total bilirubin ($P = .003$), and alkaline phosphatase ($P = .014$).

In the phase 2 cohort (N = 560), LSM was independently associated with METAVIR fibrosis stages ($P < .001$), alkaline phosphatase ($P < .001$), BMI ($P = .017$), and platelet count ($P = .033$) (Table 3).

Table 2. Diagnostic Performances (AUROCs) of VCTE for Prediction of Liver Fibrosis in Comparison With Histology

METAVIR fibrosis stage	Study phase	N (%)	AUROC	LSM cutoff (*kPa*)	Sensitivity (%)		Specificity (%)		PPV (%)		NPV (%)		DA (%)		LR+	LR–
F≥2	Phase 1	83 (44.1)	0.89	8.4	81.9	.042	79.0	.741	75.6	.155	84.7	.012	80.3	.012	3.91	0.23
	Phase 2	373 (66.7)	0.73		57.9		74.9		80.8		55.0		70.5		2.11	0.41
F≥3	Phase 1	50 (26.6)	0.92	9.6	88.3	.192	81.9	.888	68.8	.112	93.7	.062	83.5	.075	4.71	0.14
	Phase 2	193 (34.5)	0.83		71.8		80.1		62.0		88.6		77.0		3.23	0.26
F4	Phase 1	38 (20.3)	0.92	12.8	84.2	.512	86.0	.945	60.4	.009	95.6	.245	85.6	.098	6.02	0.18
	Phase 2	83 (14.8)	0.90		75.9		85.1		41.6		97.6		79.8		4.09	0.14

NOTE. Comparison performed in the phase 1 cohort (N = 188) with associated cutoffs and application of these cutoffs in the phase 2 cohort (N = 560), with associated sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV), diagnostic accuracy (DA), positive likelihood ratio (LR+), and negative likelihood ratio (LR–) and their related *P* values for statistical significance between both phases. Cutoffs were chosen to maximize the sum of Sensitivity + Specificity (Youden Index).

Table 3. Predictors of Liver Stiffness Values by Multiple Regression Analysis for Each Phase of the Study

Phase 1 (N = 188)			Phase 2 (N = 560)		
Factor	R^2 (SD)	*P* value	Factor	R^2 (SD)	*P* value
METAVIR fibrosis stage	31.7% (9.6)	<.05	METAVIR fibrosis stage	44.4% (11.6)	<.001
Alkaline phosphatase		.014	Alkaline phosphatase		<.001
Total bilirubin		.003	BMI		.017
			Platelets		.033

R^2, regression coefficient; SD, standard deviation.

Influence of Body Mass Index on Performance of Vibration-Controlled Transient Elastography

In the phase 2 validation cohort, performances of VCTE (AUROC) to diagnose F≥2 and cirrhosis were similar in patients with low or high BMI. However, for diagnosis of F≥3, AUROC was 0.88 in low BMI subgroup compared with 0.79 in high BMI subgroup (Table 4).

These results were confirmed by evaluating the overall performance of VCTE by using the Obuchowski method; AUROC of the BMI+ group was 0.78 ± 0.80, whereas AUROC of the BMI– group was 0.81 ± 0.22.

Vibration-Controlled Transient Elastography in Comparison With Biochemical Scores Aspartate Aminotransferase-to-Platelet Ratio Index and FIB-4

Comparison of diagnostic performances between LSM, APRI, and FIB-4 was performed on patients from the phase 2 cohort with available results (n = 546). LSM had an AUROC of 0.73 for the diagnosis of F≥2, which was higher than APRI (0.67, $P = .015$) and FIB-4 (0.67, $P = .016$). Similar differences were observed for predicting cirrhosis F4, with higher AUROC of 0.90 for LSM compared with 0.78 for APRI ($P < .001$) and 0.78 ($P < .001$) for FIB-4 (Table 5).

Table 4. Diagnostic Performances of VCTE as Function of Patient BMI in the Validation Cohort (N = 560)

METAVIR	BMI group	AUROC
F≥2	BMI+	0.73
	BMI–	0.75
	Whole population	0.73
F≥3	BMI+	0.79
	BMI–	0.88
	Whole population	0.83
F4	BMI+	0.91
	BMI–	0.90
	Whole population	0.90

Discussion

This study is one of the few published biopsy-controlled studies defining stiffness thresholds for stages ≥F2, F3, and F4 in a development cohort and then evaluating their performance in an independent validation cohort. Moreover, this is a study in a large cohort of U.S. patients with chronic viral hepatitis. We confirmed good performance of VCTE to stage significant fibrosis F≥2, with AUROC of 0.89. Performance was also excellent for detecting advanced fibrosis F≥3 and cirrhosis F4, with AUROC of 0.92 for both. The new American Association for the Study of Liver Diseases guidelines for prioritization of direct acting antiviral treatment have suggested that F3 and greater is an important group to treat.[22] Our findings would suggest that VCTE represents an important tool for the diagnosis of F3 and greater and that a liver stiffness >9.6 kPa can be used to help guide treatment decisions. These diagnostic performances were confirmed by the excellent adjusted AUROC of the VCTE overall diagnostic accuracy calculated by using the Obuchowski method (0.89) and are consistent with previous studies performed in hepatitis C virus patients or in other viral hepatitis cohorts[10,23] with the M probe. LSM cutoffs were also equivalent to those of previous studies:[24,25] Castera et al[24] reported similar cutoffs of 7.1, 9.5, and 12.5 kPa for the diagnostic of F≥2, F≥3, and F4, respectively, in a comparable hepatitis C virus population.

LSM performance characteristics were markedly better when compared with controlled liver biopsy in phase 1 compared with the validation cohort with uncontrolled biopsy in phase 2. Nonetheless, the application of our cutoffs in the validation cohort (phase 2) showed lower

Table 5. Diagnostic Performances of VCTE and Fibrosis Biochemical Scores APRI and FIB-4 for Diagnosis of F≥2 and for F4 in the Validation Cohort (Phase 2, N = 546)

	AUROC F≥2	*P* value vs VCTE	AUROC F4	*P* value vs VCTE
VCTE	0.735		0.897	
APRI	0.673	.015	0.785	<.001
FIB-4	0.672	.016	0.781	<.001

sensitivities, specificities, and diagnostic accuracies. However, these differences in diagnostic performance between the development and validation cohorts were only statistically significant for the diagnosis of significant fibrosis.

The relatively poor performance of the selected thresholds in the validation group could be explained by several factors. First, the mean liver stiffness in phase 1 was significantly different between the 2 cohorts for METAVIR stage $\geq$F2 ($P = .007$), causing a comparison bias and potentially explaining the low values of sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic accuracy in the phase 2 cohort. Second, biopsy represents an imperfect gold standard to evaluate fibrosis because of non-homogenous distribution of fibrosis, and sampling variability between cohorts could also explain differences in LSM diagnostic performance. Prior studies have shown that limited diagnostic accuracy of fibrosis markers might be due to error of the biopsy itself.[26]

Interestingly, we also noted poor diagnostic performance for indirect serum biomarkers such as APRI and FIB-4 for cirrhosis. The diagnostic performance for these markers was lower than that of prior studies. Zarski et al[27] reported AUROCs of 0.86 for APRI and 0.83 for FIB-4 to diagnose cirrhosis in a cohort of CHC patients, compared with 0.78 and 0.79 in our study. However, our results confirm that VCTE is more accurate for fibrosis assessment than the studied biomarkers APRI and FIB-4 and as reported in previous studies in both hepatitis C virus[27] and hepatitis B virus patients.[28]

In our overall study population, 93 of 907 patients had to be excluded because of unreliable LSM (less than 8 valid measurements within 20 attempts). Thus, the applicability rate of VCTE was 89.7% in the U.S. population, demonstrating good feasibility of the technique with better applicability rates than previously published.[29,30] However, the reliability criteria for a reliable VCTE examination in these studies were more restrictive (minimum of 10 valid measurements, success rate $\geq$60%, and interquartile range/median stiffness ratio $\leq$30% were required) compared with our study, potentially explaining these differences. It is important to point out that success rate of 60% has not been validated as an independent variable that predicts accuracy and that the interquartile range and obtaining 8 or more valid shots are probably the best criteria for a valid scan.[31,32]

We further investigated the influence of patient BMI on LSM performance and report a positive correlation between stiffness and patient BMI in the phase 2 cohort. However, the diagnostic performance of LSM was not significantly affected by BMI except for patients with advanced fibrosis F3, for which the AUROC for patients with elevated BMI was lower than for patients with lower BMI (0.79 vs 0.88, respectively). This finding is consistent with prior studies because patient BMI is known to be associated with LSM failure and unreliable measurements.[12] However, the limitation of the standard M probe in terms of diagnostic performance in overweight or obese patients may be overcome by using the newer XL probe.[33]

We also confirmed the good reproducibility of VCTE, which exhibited excellent intraobserver agreement (ICC, 0.95) and interobserver agreement (ICC, 0.98) similar to that reported by Boursier et al[34] and Neukam et al.[35]

Our study has some limitations. First, the quality of the biopsy specimens (1.5-cm minimum length with at least 6 portal tracts and 1-cm minimum for cirrhosis) was suboptimal in comparison with recommended standards.[1] This could have affected the accuracy of the biopsy, which is our reference method, thus potentially causing bias in the evaluation of the diagnostic performance of VCTE.

Second, the VCTE measurements were not taken in fasting conditions. It has recently been demonstrated that VCTE measurements can be overestimated because of food intake.[29] The manufacturer now recommends waiting at least 2 hours after a meal before doing the VCTE examination. Nonetheless, such recommendations were not published at the time of the study, and this also could have affected and potentially underestimated the diagnostic accuracy of VCTE.

Third, we did not exclude hepatitis B virus patients from our analysis, although different cutoffs have been reported in this indication.[30] Nevertheless, because of the very small number of hepatitis B virus patients in our cohort, it did not affect the cutoffs and diagnostic performances of VCTE (data not shown). In summary, we can conclude that there is a strong correlation between liver stiffness and fibrosis in a heterogeneous but representative U.S. population of hepatitis C virus/hepatitis B virus patients and similar to those reported in European and Asian populations, and that useful thresholds can be implied from this study. However, because of the variability of histologic fibrosis and the accuracy of biopsy, we believe that VCTE should be used in conjunction with other clinical data, including simple fibrosis markers, to provide information on fibrosis stage and risk of cirrhosis. Application of VCTE to clinical care has been recently reviewed and is a new point-of-care evaluation for assessing liver stiffness as a noninvasive measure of fibrosis evaluation in U.S. patients with chronic viral hepatitis.

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 http://dx.doi.org/10.1016/j.cgh.2014.12.014.

References

1. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;38:1449–1457.
2. Friedrich-Rust M, Wunder K, Kriener S, et al. Liver fibrosis in viral hepatitis: noninvasive assessment with acoustic radiation force

impulse imaging versus transient elastography. Radiology 2009; 252:595–604.
3. Bavu E, Gennisson JL, Couade M, et al. Noninvasive in vivo liver fibrosis evaluation using supersonic shear imaging: a clinical study on 113 hepatitis C virus patients. Ultrasound Med Biol 2011;37:1361–1373.
4. Sandrin L, Fourquet B, Hasquenoph J-M, et al. Transient elastography: a new non-invasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705–1713.
5. Bonder A, Afdhal N. Utilization of FibroScan in clinical practice. Curr Gastroenterol Rep 2014;16:372.
6. Tapper EB, Castera L, Afdhal NH. FibroScan (vibration controlled transient elastography): where does it stand in the US practice. Clin Gastroenterol Hepatol 2015;13:27–36.
7. Ganne-Carrie N, Ziol M, De Ledinghen V, et al. Accuracy of liver stiffness measurement for the diagnosis of cirrhosis in patients with chronic liver diseases. Hepatology 2006;44:1511–1517.
8. Nguyen-Khac E, Chatelain D, Tramier B, et al. Assessment of asymptomatic liver fibrosis in alcoholic patients using fibroscan: prospective comparison with seven non-invasive laboratory tests. Aliment Pharmacol Ther 2008;28:1188–1198.
9. Nahon P, Kettaneh A, Tengher-Barna I, et al. Assessment of liver fibrosis using transient elastography in patients with alcoholic liver disease. J Hepatol 2008;49:1062–1068.
10. Ziol M, Handra-Luca A, Kettaneh A, et al. Non-invasive assessment of liver fibrosis by stiffness measurement: a prospective multicentre study in patients with chronic hepatitis C. Hepatology 2005;41:48–54.
11. Wong VW, Vergniol J, Wong G, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454–462.
12. Castera 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.
13. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology 2008;134:960–974.
14. Parkes J, Guha IN, Roderick P, et al. Performance of serum marker panels for liver fibrosis in chronic hepatitis C. J Hepatol 2006;44:462–474.
15. Patel K, Gordon SC, Jacobson I, et al. Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients. J Hepatol 2004;41:935–942.
16. Rosenberg WM, Voelker M, Thiel R, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004;127:1704–1713.
17. Wai CT, Cheng CL, Wee A, et al. Non-invasive models for predicting histology in patients with chronic hepatitis B. Liver Int 2006;26:666–672.
18. Vallet-Pichard A, Mallet V, Pol S. FIB-4: a simple, inexpensive and accurate marker of fibrosis in HCV-infected patients. Hepatology 2006;44:769; author reply 769–770.
19. Wai CT, Greenson JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38 :518–526.
20. Le Calvez S, Thabut D, Messous D, et al. The predictive value of Fibrotest vs APRI for the diagnosis of fibrosis in chronic hepatitis C. Hepatology 2004;39:862–863; author reply 863.
21. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. Hepatology 1996;24:289–293.
22. AALSD AAftSoLD-, IDSA IDSoA-. Recommendations for testing, managing, and treating hepatitis C. HCV Guidelines 2014.
23. Cardoso AC, Carvalho-Filho RJ, Stern C, et al. Direct comparison of diagnostic performance of transient elastography in patients with chronic hepatitis B and chronic hepatitis C. Liver Int 2012;32:612–621.
24. Castera L, Vergniol J, Foucher J, et al. Prospective comparison of transient elastography, Fibrotest, APRI and liver biopsy for the assessement of fibrosis in chronic hepatitis C. Gastroenterology 2005;128:343–350.
25. Lupsor Platon M, Stefanescu H, Feier D, et al. Performance of unidimensional transient elastography in staging chronic hepatitis C: results from a cohort of 1,202 biopsied patients from one single center. J Gastrointest Liver Dis 2013;22:157–166.
26. Mehta SH, Lau B, Afdhal NH, et al. Exceeding the limits of liver histology markers. J Hepatol 2009;50:36–41.
27. Zarski JP, Sturm N, Guechot J, et al. Comparison of 9 blood tests and transient elastography for liver fibrosis in chronic hepatitis C: the ANRS HCEP-23 study. J Hepatol 2012; 56:55–62.
28. Zhu X, Wang LC, Chen EQ, et al. Prospective evaluation of FibroScan for the diagnosis of hepatic fibrosis compared with liver biopsy/AST platelet ratio index and FIB-4 in patients with chronic HBV infection. Dig Dis Sci 2011;56:2742–2749.
29. Berzigotti A, De Gottardi A, Vukotic R, et al. Effect of meal ingestion on liver stiffness in patients with cirrhosis and portal hypertension. PLoS One 2013;8:e58742.
30. Verveer C, Zondervan PE, Ten Kate FJ, et al. Evaluation of transient elastography for fibrosis assessment compared with large biopsies in chronic hepatitis B and C. Liver Int 2012;32:622–628.
31. Lucidarme D, Foucher J, Le Bail B, et al. Factors of accuracy of transient elastography (FibroScan®) for the diagnosis of liver fibrosis in chronic hepatitis C. Hepatology 2009;49:1–7.
32. Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria of liver stiffness evaluation by transient elastography. Hepatology 2013;57:1182–1191.
33. 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;35:199–208.
34. Boursier J, Konate A, Gorea G, et al. Reproducibility of liver stiffness measurement by ultrasonographic elastometry. Clin Gastroenterol Hepatol 2008;6:1263–1269.
35. Neukam K, Recio E, Camacho A, et al. Interobserver concordance in the assessment of liver fibrosis in HIV/HCV-coinfected patients using transient elastometry. Eur J Gastroenterol Hepatol 2010;22:801–807.

Reprint requests
Address requests for reprints to: Nezam H. Afdhal, MD, Liver Center, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215. e-mail: nafdhal@bidmc.harvard.edu; fax: (617) 632-1054.

Conflicts of interest
The authors disclose the following: all investigators received a FibroScan device as part of the study. Dr Afdhal was the IND holder for FibroScan and has been a consultant for EchoSens.

Funding
Grant support was provided to the investigators by EchoSens.

Supplementary Materials

Methods

Patients with any of the following were excluded from study participation: active malignancy, any antiviral treatment within the last 6 months, uninterpretable biopsy specimen, other chronic liver disease, clinical ascites, BMI $\geq$40 kg/m^2, pregnancy, or any implantable cardiac device. All patients provided written informed consent, and the study was approved by institutional review board at each center.

Vibration-Controlled Transient Elastography

FibroScan VCTE examinations were conducted by using the conventional standard M probe (ultrasound central frequency of 3.5 MHz, with measurement depth from 2.5 to 6.5 cm below skin surface).

The tip of the probe transducer was covered with coupling gel and placed on the skin between the ribs anterior to the right lobe of the liver. By using an ultrasonic image, the operator located a portion of the liver that was at least 7 cm in depth and free of large vascular structures. During acquisition, patients lay on their back and had their right arm behind their head. LSM was considered valid if there were at least 8 valid measurements within 20 attempts. LSM results were expressed as a median value (kPa) of all valid measurements with associated interquartile range and success rate.

Statistics

All statistical analyses were performed by using R software (R Development Core Team 2008) and IBM SPSS (Armonk, New York, NY), and P value $<$ .05 was considered significant.

Patient characteristics were summarized as means and standard deviation or numbers of cases and percentages, as appropriate.

The interobserver and intraobserver agreement between VCTE values was assessed by using the ICC and respective 95% confidence intervals.

The relationships between LSM and other variables (clinical, histologic, and biological parameters) were evaluated by using the Kendall rank correlation coefficient. Multivariate analyses were performed by using multiple linear regression to investigate the variables independently associated with LSM (LSM was transformed by using the Box–Cox transformation[1] owing to its skewed distribution). A backward selection procedure, which was based on the minimization of the Akaike information criterion, selected independent features significantly associated with LSM.

Diagnostic performance was assessed by AUROC curves by using liver biopsy as reference. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for optimal cutoff values obtained by maximizing the sum of sensitivity and specificity (Youden Index) for the selected diagnosis.

In addition, the Obuchowski measure was assessed,[2] taking into account the distribution of fibrosis stages in the study population. The Obuchowski index is a multinomial version of the AUROC and can be interpreted as the probability that LSM will correctly rank 2 randomly chosen patients sampled with different fibrosis stages F0–F4. Weighting was based on the relative proportion of 5 fibrosis stages in the study population. The Obuchowski measure was assessed by using a penalty function similar to that reported in Lambert et al.[3] This penalty function was proportional to the difference in fibrosis grade units, ie, a penalty of 1/4 when the difference between grades was 1, 2/4 when the difference was 2, $^3/_4$ when the difference was 3, and 1 when the difference was 4.

To further investigate the infuence of patient's BMI on the diagnostic performances of VCTE, the median BMI for each fibrosis stage was used to split the studied population into low and elevated BMI subgroups with identical distribution of patients across the 5 fibrosis stages: BMI below the median BMI for the given fibrosis stage (BMI– group) and BMI above the median BMI for the given fibrosis stage (BMI+ group). AUROCs were then calculated for both populations for prediction of F$\geq$2, F$\geq$3, and F4.

To compare the diagnostic accuracy of VCTE with the biomarkers APRI and FIB-4, differences between AUROCs were determined by the DeLong test, which takes into account the obvious correlations between the measurements.[4]

Results

In phase 1 (March 2005–February 2006), 237 patients were enrolled at 3 sites in the United States; 49 subjects were excluded because of the following: lack of either biopsy or LSM result (n = 15), unreliable LSM result (<8 valid measurements within 20 attempts, n = 32), or inadequate quality of biopsy (n = 2). Thus, 188 patients were eligible for statistical analysis in phase 1. For phase 2 (March 2007–September 2008), 670 patients were enrolled at 6 sites; 110 were excluded because of the following: lack of biopsy or LSM result (n = 8), unreliable LSM result (<8 valid measurements within 20 attempts, n = 77), or inadequate quality of biopsy, n = 25). Thus, 560 patients were eligible for statistical analysis in phase 2.

Discussion

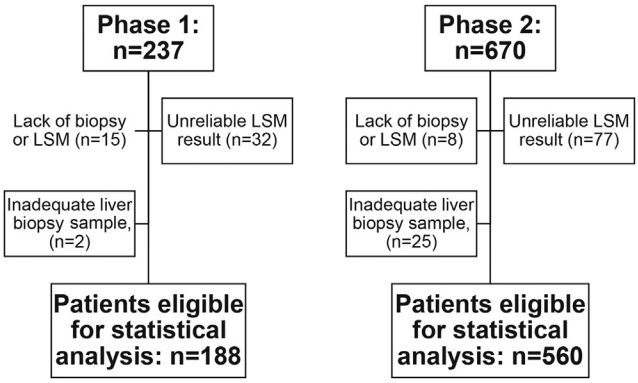
BMI was the only independent factor associated with unreliable LSM measurements (P < .001, data not shown). This finding is consistent with previously published studies and confirms that obesity represents a

limitation of VCTE.[5] However, it must be stressed that the VCTE examinations were performed with the conventional M probe only; the new XL probe has been recently developed to target the overweight population and exhibited a good applicability rate.[6]

As a limitation of the study, it should be pointed out that it was conducted in chronic viral hepatitis patients only; therefore, additional studies are still required to evaluate the usefulness of VCTE in nonalcoholic steatohepatitis patients.

References

1. Sakia R. The Box-Cox transformation technique: a review. The Statistician 1992;41:169–178.
2. Obuchowski NA, Goske MJ, Applegate KE. Assessing physicians' accuracy in diagnosing paediatric patients with acute abdominal pain: measuring accuracy for multiple diseases. Stat Med 2001;20:3261–3278.
3. Lambert J, Halfon P, Penaranda G, et al. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. Clin Chem 2008;54:1372–1378.
4. 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.
5. Castera 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.
6. 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;35:199–208.



Supplementary Figure 1. Flow chart of patients for both study phase 1 and phase 2.