Title

Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease

Journal

Hepatology, 67(1)

ISSN

0270-9139

Authors

Vuppalanchi, Raj
Siddiqui, Mohammad S
Van Natta, Mark L
et al.

Publication Date

2018

DOI

10.1002/hep.29489

Peer reviewed

HHS Public Access

Author manuscript

Hepatology. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Hepatology. 2018 January ; 67(1): 134–144. doi:10.1002/hep.29489.

Performance Characteristics of Vibration-Controlled Transient Elastography for Evaluation of Non-Alcoholic Fatty Liver Disease

Raj Vuppalanchi[1], **Mohammad S. Siddiqui**[2], **Mark L. Van Natta**[3], **Erin Hallinan**[3], **Danielle Brandman**[4], **Kris Kowdley**[5], **Brent A Neuschwander-Tetri**[6], **Rohit Loomba**[7], **Srinivas Dasarathy**[8], **Manal Abdelmalek**[9], **Edward Doo**[10], **James A. Tonascia**[3], **David E. Kleiner, MD**[11], **Arun J. Sanyal**[2,¶], and **Naga Chalasani**[1,¶] **for the NASH Clinical Research Network**

[1]Indiana University School of Medicine, Indianapolis, IN

[2]Virginia Commonwealth University, Richmond, VA

[3]The Johns Hopkins University School of Public Health, Baltimore, MD

[4]University of California at San Francisco, San Francisco, CA

[5]Swedish Medical Center, Seattle, WA

[6]St Louis University, St Louis, MO

[7]University of California San Diego, San Diego, CA

[8]Cleveland Clinic Foundation, Cleveland, OH

[9]Duke University, Durham, NC

[10]Liver Disease Branch, NIDDK, National Institutes of Health, Bethesda, MD

[11]National Cancer Institute, Baltimore, MD

Abstract

Background—Vibration-controlled transient elastography (VCTE) estimates liver stiffness measurement (LSM) and controlled attenuation parameter (CAP) which are noninvasive assessments of hepatic fibrosis and steatosis respectively. However, prior VCTE studies reported high failure rate in patients with non-alcoholic fatty liver disease (NAFLD).

Aim—To examine the performance characteristics of Fibroscan 502 Touch with two probes, medium (M+) and extra-large (XL+), in patients with NAFLD in a multicenter setting.

Methods—A total of 1696 exams were attempted in 992 patients (BMI: 33.6 ± 6.5 kg/m^2) with histologically confirmed NAFLD. Simultaneous assessment of LSM and CAP was performed

¶Correspondence may be addressed to: Naga Chalasani, MD at nchalasa@iu.edu or Arun J Sanyal, MD at arun.sanyal@vcuhealth.org.

Conflicts of Interests: Mr. Van Natta and Drs. Tonascia, Hallman, Dasarathy, Brandman, Kleiner report no conflicts of interests. Drs. Chalasani, Abdelmalek, Sanyal, Kowdley, Neuschwander-Tetri, and Loomba have consulting agreements and/or research grants from various pharmaceutical companies but none have any consulting agreements with Echosens. The Fibroscan machines were provided at no-charge by the Echosens to the NASH CRN adult clinical centers through a clinical trials agreement with the NIDDK. Echosens had no input into study design or data analysis but had the opportunity to review this manuscript ahead of its submission.

using Fibroscan 502 Touch with an automatic probe selection tool. Testing was conducted twice in patients by either a single operator (88%) or two operators (12%). Failure was defined as the inability to obtain a valid examination. An examination was considered unreliable if LSM IQR/median was >30%. Significant disagreement between two readings was defined as greater than >95% limits of agreement between two readings.

Results—A total of 1641 examinations yielded valid results with a failure rate of 3.2% (55/1696). The proportion of unreliable scans for LSM was 2.4%. The proportion of unreliable scans with operator experience in the top quartile (≥ 59 procedures) was significantly lower than lower three quarters combined (1.6% vs.4.7%, p=0.01 by Fisher's Exact test). The significant disagreement between first and second readings for LSM and CAP when obtained back to back was 18% and 11% respectively.

Conclusion—VCTE for estimation of LSM and CAP can be successfully deployed in a multicenter setting with low failure (3.2%) and high reliability (>95%) rates and high reproducibility.

Keywords

Fibroscan; Vibration Controlled Transient Elastography; Continuous Attenuation Parameter; NAFLD; Fibrosis; Steatosis

Introduction

Non-alcoholic fatty liver disease (NALFD) is a common cause of chronic liver disease and is estimated to occur in up to one-third of individuals in the United States.[1] The clinical spectrum of NAFLD ranges from relatively benign non-alcoholic fatty liver (NAFL) to non-alcoholic steatohepatitis (NASH) that can progress to cirrhosis, liver failure or hepatocellular cancer.[2, 3] In NAFL, only macrosteatosis is present on liver biopsy whereas, in NASH, macrosteatosis is present with a variable mix of hepatocyte ballooning and inflammation with or without fibrosis.[4, 5] The distinction of different forms of NAFLD is important in the clinical management due to distinctly different prognosis.[6, 7] Furthermore, fibrosis has emerged as the strongest predictor of long-term outcomes in patients with NAFLD.[8] Liver biopsy has long been regarded as the gold standard for diagnosis and prognostication of patients with NAFLD.[9] However, histological interpretation of liver biopsy is subject to micro-inhomogeneity of liver tissue, sampling error based on length of liver biopsy core, presence of unfragmented core, and observer variability among the pathologists.[10, 11] Moreover, it requires the patient to undergo an invasive procedure that could be associated with discomfort and could rarely be associated with life threatening complications.[12–14]

Liver stiffness measurement (LSM) assessed by vibration controlled transient elastography (VCTE) has been shown to be an easy to perform, non-invasive test to reliably estimate the degree of liver fibrosis in patients with NAFLD.[15–19] VCTE measures the speed of a mechanically induced shear wave using pulse-echo ultrasonic acquisitions in a much larger portion of the tissue, approximately 100 times more than a liver biopsy core. However, prior studies evaluating the performance of VCTE in NAFLD have been limited by medium (M+) size probe with ultrasound probe frequency of 3.5 MHz to measure LSM at a depth of 2.5

and 6.5 centimeters (cm) from the skin.[20, 21] However, using M+ probe alone, the LSM was not reliable in subtanital number of the patients with body mass index (BMI) as the major determinant for diagnostic errors in predicting fibrosis either by overestimating or underestimating the stage of fibrosis.[22, 23] The failure rate reported in these studies have been as high as 25% presumably due to skin to liver capsule distance greater than 2.5 cm.[20, 21, 24] Despite the failure rate, several studies reported correlation between stage of liver fibrosis and LSM value in patients with NAFLD.[20, 25–27] The optimal cut-off values corresponding to various stages of fibrosis however have not been very apparent due to overlapping values of LSM across various stages. In a recent study from NASH CRN reported that LSM of 15.5 kPa had an AUROC of 0.93 for differentiating cirrhosis from non-cirrhosis stage of fibrosis with a specificity of 90%.[28]

The newer version of VCTE available in the USA since early 2013, has several features that not only overcome its prior limitations but also enhances its role as a diagnostic tool in the evaluation of patients with NAFLD. First, it is approved by the regulatory authorities to measure 3.5 MHz ultrasound coefficient of attenuation [CAP: Controlled Attenuation Parameter --the ultrasonic attenuation in the liver tissue depends on the viscosity (fat) of the medium (liver) and the distance of propagation of the ultrasonic signals into the liver].[29, 30] While LSM is measured in kilo pascals (kPa), CAP is measured in decibels per meter (dB/m) and reflects the decrease in the amplitude of ultrasound signal in the liver.[30–32] Therefore, higher CAP is reflective of the higher degree of steatosis. CAP is displayed only when LSM is valid as it is only computed from the ultrasound signals used for acquiring LSM. The shear wave speed with estimation of stiffness and CAP currently allow for simultaneous assessment of both liver fibrosis and steatosis.[33, 34] Several studies have also reported a good correlation between degree of hepatic steatosis or hepatic fat content by magnetic resonance imaging - proton density fat fraction.[35, 36] Second, is the availability of extra-large (XL+) probe with an ultrasound frequency of 2.5 MHz for measurement of shear wave at skin to liver capsule depth of 3.5 and 7.5 cms. Lastly, the availability of an automatic probe selection software tool that determines the choice of the probe based on the skin liver capsule distance. It is therefore possible that the failure rate of current VCTE may be much lower than the previous version in the evaluation of NAFLD.[23, 30, 37] The aim of the current study is to report the success and reliability of the newer, Fibroscan 502 Touch which offers a machine-derived choice between M+ and XL+ probe for simultaneous assessment of LSM and CAP in patients with NAFLD, particularly in a multicenter setting. Furthermore, a thorough understanding of factors that could potentially influence the LSM and CAP values is necessary.

Methods

Study design

Individuals who are ≥ 18 years with histologically proven NAFLD and participating in the prospective NAFLD Database 2 study underwent simultaneous assessment for LSM and CAP using Fibroscan 502 Touch (Echosens™ North America 1050 Winter Street Waltham, MA 02451). The NAFLD Database 2 study was initiated by the NASH Clinical Research Network (NASH CRN) in December 2009 and its inclusion and exclusion criteria are similar

to the previously reported NASH CRN Database study except that all participants required histologically confirmed NAFLD for eligibility.[38] Eligible participants were enrolled into the NASH CRN Database study at eight medical centers across the United States: Case Western Reserve (Cleveland, OH); Duke University (Durham, NC); Indiana University (Indianapolis, IN); Saint Louis University (St. Louis, MO); University of California, San Diego (San Diego, CA); University of California, San Francisco (San Francisco, CA); Virginia Mason Medical Center and Swedish Medical Center (Seattle, WA); and Virginia Commonwealth University (Richmond, VA). The data were stored, monitored, and analyzed at the Data Coordinating Center at the Johns Hopkins Bloomberg School of Public Health.[38] This study (NCT01030484) was approved by the Institutional Review Boards of the respective participating institutions and all participants provided written informed consent prior to enrollment. No employees of Echosens were investigators in this study, or controlled the acquisiiton, analysis, interpretation, or reporting of results.

Study Visit and Procedures—All participants presented for an outpatient study visit after 12 hours of fasting to the respective clinical research centers after overnight fast and were evaluated by the study nurse and the clinical investigator. Participants underwent anthropometric measurements and completed study-specific questionnaires, physical examination, and blood tests. Routine laboratory studies were performed on fresh samples in Clinical Laboratory Improvement Amendments (CLIA) certified laboratories at each clinical site according to standard clinical protocols.[38]

Equipment and Technician expertise—Fibroscan® 502 Touch with two probes - Medium (M+) and Extra-large (XL+) was available at each of the participating medical centers for measuring LSM and CAP. The type of probe required for each participant was based on the automatic probe selection tool embedded within the Fibroscan® 502 Touch operating software. These machines were provided by the Echosens company to adult clinical centers of the NASH CRN through a clinical trials agreement between the Echosens and the NIDDK.

All studies were performed by a dedicated study coordinator at each site using a standardized protocol.[38] All study coordinators were new to the Fibroscan technology and underwent standardized training (didactic followed by hands-on testing in 5 participants under the supervision of a proctor) by Echosens North America and were certified before conducting VCTEs for the current study. Each patient was placed supine with right arm raised behind his/her head and remained still during the procedure. With probe over the liver region, readings were attempted until ten valid measurements were obtained. All studies were started with an M+ probe with XL+ probe as rescue only when prompted by the automatic probe selection tool. In rare instances, where the machine's recommendation fluctuated between M+ and XL+ probes, the study coordinator was instructed to choose the XL+ probe. Two scans were performed during the same visit several minutes apart by the same coordinator (intra-operator assessment) or by a second coordinator (inter-operator assessment) in a subset of participants.

Statistical Analysis

Demographics, anthropometrics, liver tests, serum chemistries, histologic characteristics and Fibroscan metrics were compared between M+ and XL+ probe types using chi-square tests for categorical variables and t-tests for continuous variables. Both means with standard deviations and medians with interquartile were presented for Fibroscan data due to possible non-normality of the distribution. Correlation between 1st and 2nd readings used Spearman correlation coefficient and Fisher's z-transformation.

The unreliability of LSM values was defined as IQR/Median > 30%. Significant disagreement between 1st and 2nd readings was defined as absolute value of the difference > 95% limits of agreement between the 2 readings i.e., using linear regression, the difference between the 2 readings of LSM and CAP was not related to the mean of the 2 readings. This was further visualized using Bland-Altman plots. The B&A plot analysis is a simple way to evaluate a bias between the mean differences, and to estimate an agreement interval, within which 95% of the differences of the second method, compared to the first one, fall. It simply quantifies the bias and a range of agreement, within which 95% of the differences between one measurement and the other are included. It is also possible to say that the bias is significant, when the line of equality is not within the confidence interval of the mean difference. For analyses of LSM and CAP, the mean of the two readings or first reading if second reading was missing was used; observations with missing data for the second reading (n=28 for LSM; n=24 for CAP) were excluded from selected analyses of inter-and intra-observer differences of LSM and CAP. The CAP data for 183 patients were missing due to computer software problem using XL+ probe at the start of FibroScan assessments.

Multiple logistic regression with Akaike's Information Criterion (AIC) was used to assess associations for the binary outcomes of unreliability and significant disagreement from a candidate set including age, gender, ethnicity, race, body mass index, waist circumference, LSM (or CAP), INR, ALT, AST, serum bilirubin, serum alkaline phosphatase, blood platelet count, steatosis score, fibrosis stage, NASH diagnosis, time from biopsy, probe type, and same or different operator. Robust multiple linear regression was used to test for differences by probe type in LSM adjusted for fibrosis stage and BMI, and CAP adjusted for steatosis score and BMI.[39] LSM was modeled both untransformed and log transformed due to non-normality of the distribution. All analyses were conducted using SAS (Version 9.3 of the SAS System for Windows, Cary, NC: SAS Institute Inc., 2002–2004) and Stata (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).

Results

Study Population

Nine hundred and ninety two individuals who underwent evaluation with VCTE as part of the NAFLD Database 2 Study between April 2014 and May 2016 were included in this analysis. Of them, 961 participants had two VCTE examinations back to back, with second VCTE procedure by the same operator in 838 patients and by a different operator in 123 patients. Total number of LSM and CAP available for analysis were 1641 and 1475 respectively.

The selected clinical, laboratory, and histological features of our study cohort are described in Table 1. Overall, the mean age was 51 years with 65% women and the cohort is predominantly Caucasian (84% with 10% Hispanic ethnicity). The mean BMI was 33.6 ± 6.5 kg/m^2 with 31% belonging to normal/overweight and obese categories each and the remaining 36% belonged to extreme obese category (Table 1). The study cohort consisted predominantly of NASH patients with 54% showing definite NASH and 19% exhibiting borderline NASH. The mean duration between liver biopsy and VCTE was 2.6 years. Histograms of LSM (Figure 1, panel a) and CAP (Figure 1, panel b) showed a right-skewed distribution for LSM and a normal distribution for CAP.

VCTE Failure Rate

The failure rate for VCTE for obtaining LSM and CAP was 3.2% (55/1696) based on the total number of scans performed in the current study. The failure rate based on the total number of patients included in this study was slightly higher at 5.5% (55/992). The 3 most frequent reasons for failure: inability to obtain a scan due to excess skin to liver capsule distance (n=19); machine/technician error (n=13); and patient refusal (n=11). Other reasons included invalid readings (n=6), intolerable pain (n=4) and undetectable liver (n=1).

Performance Characteristics

Both median LSM and median CAP values were significantly higher in the higher BMI categories (Table 2). The use of XL+ probe was significantly greater in higher BMI categories, i.e., 27% in normal/overweight, 61% in obese, and 87% in extreme obese patients ($p<0.001$). The mean difference between the 1st and 2nd readings of LSM was 0.0 ± 5.5 kPa and did not significantly differ among different BMI categories (Table 2). The mean difference between the 1st and 2nd CAP readings was 2 ± 36 dB/m and it was also not significantly different among various BMI categories (Table 2). The Bland-Altman plot to describe the agreement between the two LSM readings showed that 95% of the difference between the 1st and 2nd readings occurred within ± 2.9 kPa with no bias over the range of values (4–75 kPa) (Figure 1, panel c). Similarly, the Bland-Altman plot to describe the agreement between the two CAP readings showed that 95% of the difference between the 1st and 2nd readings occurred within ±51 dB/m with no bias over the range of values (100–400 dB/m) (Figure 1, panel d).

Factors associated with Unreliable LSM

Unreliable results due to significant variability i.e., median LSM values with IQR/median >30% were found in 2.4% (39/1651) of VCTE scans. The relationship between operator experience in quartiles and proportion of unreliable scans was first examined (Table 3). The proportion of unreliable scans in the 4th quartile was 1.6% and was significantly lower than first three quarters combined (4.7%, p=0.01 by Fisher's Exact test). Logistic regression model with all variables from Table 1 using AIC criteria showed that BMI category (OR: 2.7, 95% CI: 1.5, 4.8; p=0.001) and Hispanic ethnicity (OR: 3.2, 95% CI: 1.1, 9.3; p=0.03) were associated with higher odds of unreliable LSM (Table 4). On the other hand lower values of international normalized ratio (INR) (OR: 0.94 per %, 95% CI: 0.89, 0.99; p=0.01) and ALT (OR: 0.91 per 10 U/L, 95% CI: 0.78, 1.05; p=0.19) were associated with decreased odds of unreliable LSM values (Table 4). There was a 25% decrease in the odds of

unreliability per increasing quartile of prior exams adjusted for other confounders (OR: 0.75, 95% CI: 0.52, 1.08; p=0.12) (Table 4).

Factors Associated with Significant Disagreement between 1st and 2nd VCTE Readings

Values of LSM and CAP between 1st and 2nd exams were compared, and significant disagreement was defined as absolute value of the difference > 95% limits of agreement between the 2 readings. Using linear regression, the difference between the 2 readings of LSM and CAP was not related to the mean of the 2 readings (Change in LS difference / 1 unit increase in LS mean = 0.002 kPa (SE=0.016); Change in CAP difference / 1 unit increase in CAP mean = −0.001 dB/m (SE=0.024). This can also be visually seen in the Bland-Altman plots in Figures 1c and 1d. The rate of significant disagreement for LSM was 18% (169 out of 964 paired LSM measurements) while the rate of significant disagreement for CAP was 11% (85 out of 785 paired CAP measurements). The factors associated with significant disagreement between 1st and 2nd measurements of LSM and CAP are shown in Table 5. While the significant disagreement between two LSM values increased with increasing LSM score (OR: 1.06, 95% CI: 1.04, 1.09; p<0.001), higher fibrosis stage (OR: 1.22, 95% CI: 1.02, 1.46; p=0.03), and AST (OR: 1.07, 95% CI: 1.01, 1.13; p=0.03), factors such as male gender (OR:0.52, 95% CI: 0.32, 0.86; p=0.01), White race (OR: 0.50, 95% CI: 0.26, 0.96, p=0.04) and platelet count (OR: 0.96, 95% CI: 0.93, 1.00; p=0.04) were associated with lower significant disagreement. Factors associated with less disagreement between 1st and 2nd CAP scores were higher CAP (per every 10 dB/m) values (OR: 0.91, 95% CI: 0.86, 0.96; p=0.001) and steatosis grade (OR: 0.67, 95% CI: 0.49, 0.92; p=0.01). Waist circumference was associated with higher disagreement in both CAP (OR: 1.03, 95% CI: 1.01, 1.05, p=0.006) and LSM (OR: 1.02, 95% CI: 1.00, 1.04; p=0.09) (Table 5).

The correlation between 1st and 2nd measurements obtained by the same coordinator (intra-observer) and by a different coordinator (inter-observer) was assessed in 838 and 123 participants, respectively. The correlation between 1st and 2nd readings was high for both inter- (r=0.84) and intra-operator (r=0.90) LSM values. This correlation was significantly higher when M+ probe was used as compared to XL+ probe (inter-operator r=0.90 vs. 0.77, p=0.02; intra-operator r=0.94 vs. 0.87; p<0.001). The correlation between the 1st and 2nd CAP readings was high for both inter- (r=0.70) and intra- (r=0.82) operator values. The inter-operator correlation did not differ by M+ vs. XL+ probe type (r=0.64 vs 0.68; p=0.71) but the intra-operator correlation was significantly higher using M+ vs. XL+ probe (r=0.85 vs. 0.75, p=0.0003).

Probe Size and Performance Characteristics

The mean and median LSM values with the XL+ probe were significantly higher than with the M+ probe (mean ± SD 12.5±13.1 vs. 9.7±8.8 kPa, p<0.0001; median[IQR] = 7.9[5.7, 13.6] vs. 6.8[4.9, 10.8] kPa, p<0.0001). However, when adjusted for BMI and fibrosis stage, mean LSM using M+ probe was significantly higher by 0.8 kPa (95% CI: 0.3,1.4; p=0.003) compared to XL+ probe and the ratio of LSM using M+ / XL+ was significantly higher by 6% (95% CI=3%, 9%; p<0.001) (back-transforming log LSM) (Table 6). The mean CAP score in XL+ probe group was significantly higher than the M+ probe group (326±47 vs. 288±52 dB/m, p<0.0001) (Table 2). When adjusted for BMI and severity of hepatic steatosis

by histology, CAP scores obtained using XL+ probe were higher by 16 dB/m (95% CI: 8, 24; p <0.001) (Table 6).

Discussion

The VCTE examination offers an opportunity to noninvasively assess the extent of liver fibrosis and hepatic steatosis in patients with NAFLD in an outpatient clinic setting. Moreover, the change in LSM and CAP values over time may provide an opportunity to monitor disease progression without a liver biopsy. Soon after its approval in 2013, there has been interest in incorporating VCTE into ambulatory gastroenterology and hepatology practices throughout North America. However, the performance of VCTE via Fibroscan®502 Touch which offers two probes and an automated probe selection by the device is not well understood.

The main findings from this study are (a) failure rate for obtaining valid LSM and CAP using currently available VCTE device and software technology is quite minimal; (b) unreliable liver stiffness measurements are rare but are associated with obesity and operator inexperience; and (c) reproducibility of liver stiffness and CAP measurements is high whether the test is repeated by the same, or by another operator.

In a recent study conducted in the United States by Tapper et al., VCTE was performed in 164 patients with NAFLD utilizing an M+ probe and it reported 73% success rate for obtaining reliable LSM which was defined as 10 valid measurements and IQR ≤ 30% of the median. If we apply the same definition as Tapper et al., the success rate for VCTE in our experience was high at 94%. We believe the improved performance of the VCTE in our study is due to the availability of the XL+ probe which was utilized in 60% of the participants. In the study by Tapper et al. extreme obesity was strongly associated with uninterpretable VCTE results when obtained with an M+ probe (OR: 8.76; 95% CI 3.82–21.0). In our study, nearly 90% of the extremely obese participants had their VCTE performed with an XL+ probe and thus overcoming the limitation of the M+ probe in obtaining valid reading in those with extreme obesity. Finally, Hispanic ethnicity was associated with unreliable LSM in the current cohort. We suspect that this association may be related to greater prevalence of abdominal adiposity, shorter height and possibly smaller intercostal space.

The divergent relationship between the LSM and CAP values with the size of the probe is interesting. Although caution should be exercised in interpreting these findings since these values are not obtained using both probes in the same patient, the current study revealed that LSM values with XL+ probe are lower by 0.8 kPa while CAP values are increased by approximately 16 dB/m. Although the lower LSM values by the XL+ probe, when compared to M+ probe in the same individual, has been described before,[40] a higher CAP values with the XL+ probe for similar degree for steatosis is a novel finding. These results are important when interpreting the changes in the values of LSM and CAP during the course a longitudinal study when the size of the probe may inadvertently change. Therefore, consideration should be given for ± 1 kPa and ± 16 kPa when establishing a threshold cut-off for the stage of fibrosis and degree of steatosis respectively.

An important shortcoming of this study is that it did not examine the relationship between LSM and CAP and liver histology because the mean duration between liver biopsy and VCTE was 2.6 years. We believe that it is best to investigate the relationship between VCTE and liver histology obtained at a much closer interval. In an earlier study from the NASH CRN, the relationship between serum keratin 18 fragments and liver histology was much stronger when they were measured in close proximity to the liver biopsy.[41]

In summary, VCTE for estimation of LSM and CAP can be successfully deployed in a multi-center setting with a very low failure rate (<2.5%) and high reliability rate (>95%). Although the reasons for failure and unreliable scans appear to be related to factors associated with the patient, VCTE technology, or the operator, the current study suggests that with sufficient operator experience, it is possible to minimize failure and unreliability rates. Our study demonstrates high performance characteristics for currently available VCTE, supporting its use in clinical practice and in multicenter research setting.

Acknowledgments

Source of funding (adult and pediatric)

The Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (grants U01DK061718, U01DK061728, U01DK061731, U01DK061732, U01DK061734, U01DK061737, U01DK061738, U01DK061730, U01DK061713). Additional support is received from the National Center for Advancing Translational Sciences (NCATS) (grants UL1TR000439, UL1TR000077, UL1TR000436, UL1TR000150, UL1TR000424, UL1TR000006, UL1TR000448, UL1TR000040, UL1TR000100, UL1TR000004, UL1TR000423, UL1TR000058, UL1TR000454). This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute.

Abbreviations

NAFLD	Nonalcoholic Fatty Liver Disease
NASH	Nonalcoholic Steatohepatitis
NASH CRN	NASH Clinical Research Network
VCTE	Vibration Controlled Transient Elastography
LSM	Liver Stiffness Measurement
CAP	Controlled Attenuation Parameter

References

1. Williams CD, SJ, Asike M, Torres D, Shaw J, Contreras M, Landt C, Harrison SA. Prevalence of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis Among a Largely Middle-Aged Population Utilizing Ultrasound and Liver Biopsy: A Prospective Study. Gastroenterology. 2011:124–131.
2. Wong RJ, CR, Ahmed A. Nonalcoholic Steatohepatitis Is the Most Rapidly Growing Indication for Liver Transplantation in Patients With Hepatocellular Carcinoma in the U.S. Hepatology. 2014:2188–2195. [PubMed: 24375711]
3. Charlton MBJ, Pedersen RA, Watt KD, Heimback JK, Dierkhisin RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. Gastroenterology. 2011:1249–1253. [PubMed: 21726509]

4. Brunt EM, Tiniakos DG. Histopathology of nonalcoholic fatty liver disease. World J Gastroenterol. 2010; 16:5286–96. [PubMed: 21072891]
5. Brunt EM. Pathology of nonalcoholic fatty liver disease. Nature reviews Gastroenterology & hepatology. 2010; 7:195–203. [PubMed: 20195271]
6. Dam-Larsen S, Becker U, Franzmann MB, et al. Final results of a long-term, clinical follow-up in fatty liver patients. Scand J Gastroenterol. 2009; 44:1236–43. [PubMed: 19670076]
7. Adams LA, Sanderson S, Lindor KD, et al. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. J Hepatol. 2005; 42:132–8. [PubMed: 15629518]
8. Angulo PKD, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharcenwitthya P, Mills PR, Keach JC, Lafferty HD, Stahler A, Haflidadottir S, Bendtsen F. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2015; 149:389–397. [PubMed: 25935633]
9. Chalasani NYZ, Lavine J, Diehl AM, Brunt E, Cusi K, Charlton M, Sanyal A. The Diagnosis and Management of Non-Alcoholic Fatty Liver Disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology. 2012; 55:2005–23. [PubMed: 22488764]
10. The METAVIR Study Group. Intraobserver and Interobserver Variations in Liver Biopsy Interpretation in Patients with Chronic Hepatitis C. Hepatology. 1994:15–20. [PubMed: 8020885]
11. Ratziu VCF, Heurtier A, Bombert S, Giral P, Bruckert E, Grimaldi A, Capron F, Poynard T. Sampling Variability of Liver Biopsy in Nonalcoholic Fatty Liver Disease. Gastroenterology. 2005:1898–1906. [PubMed: 15940625]
12. Szymczak ASK, Inglot M, Gladysz A. Safety and Effectiveness of Blind Percutaneous Liver Biopsy: Analysis of 1412 Procedures. Hepat Mon. 2012; 12:32–37. [PubMed: 22451841]
13. Behrens GFH, Giusto D, Patel J, Van-Thiel DH. Transjugular Liver Biopsy: Comparison of Sample Adequacy with the Use of Two Automated Needle Systems. J Vasc Interv Radiol. 2011:341–345. [PubMed: 21194967]
14. Kalambokis GMP, Vibhakorn S, Marelli L, Cholongitas E, Senzolo M, Patch D, Burroughs A. Transjugular liver biopsy – Indications, adequacy, quality of specimens, and complications – A systematic reviewTransjugular liver biopsy. Journal of Hepatology. 2007:284–294. [PubMed: 17561303]
15. Castera LWM, Pambrun E, Paradis V, Perez P, Loko MA, Asselineau J, Dabis F, Degos F, Salmon D. Comparison of transient elastography (FibroScan), FibroTest, APRI and two algorithms combining these non-invasive tests for liver fibrosis staging in HIV/HCV coinfected patients: ANRS CO13 HEPAVIH and FIBROSTIC collaboration. HIV Medicine. 2014:30–39. [PubMed: 24007567]
16. Obara NUY, Fukushima K, Nakagome Y, Kakazu E, Kimura O, Wakut Y, Kido O, Ninomiya M, Kogure T, Inoue J, Kondo Y, Shiina M, Iwasaki T, Yamamoto T, Shimosegawa T. Transient elastography for measurement of liver stiffness measurement can detect early significant hepatic fibrosis in Japanese patients with viral and nonviral liver diseases. Journal Gastroenterol. 2008; 43:720–728.
17. Corpechot CGF, El-Naggar A, Kemgang A, Wendum D, Poupon R, Carrat F, Chazouillères O. Baseline Values and Changes in Liver Stiffness Measured by Transient Elastography Are Associated With Severity of Fibrosis and Outcomes of Patients With Primary Sclerosing Cholangitis. Gastroenterology. 2014:970–979. [PubMed: 24389304]
18. Yoneda M, Mawatari H, Fujita K, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). Dig Liver Dis. 2008; 40:371–8. [PubMed: 18083083]
19. Shin SH, Kim SU, Park JY, et al. Liver stiffness-based model for prediction of hepatocellular carcinoma in chronic hepatitis B virus infection: comparison with histological fibrosis. Liver International. 2015; 35:1054–1062. [PubMed: 24930484]
20. Tapper EB, Challies T, Nasser I, et al. The Performance of Vibration Controlled Transient Elastography in a US Cohort of Patients With Nonalcoholic Fatty Liver Disease. Am J Gastroenterol. 2016; 111:677–84. [PubMed: 26977758]

21. Castera L, Foucher J, Bernard PH, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. Hepatology. 2010; 51:828–35. [PubMed: 20063276]
22. Petta S, Di Marco V, Camma C, et al. Reliability of liver stiffness measurement in non-alcoholic fatty liver disease: the effects of body mass index. Aliment Pharmacol Ther. 2011; 33:1350–60. [PubMed: 21517924]
23. 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; 107:1862–71. [PubMed: 23032979]
24. Pang JX, Pradhan F, Zimmer S, et al. The feasibility and reliability of transient elastography using Fibroscan(R): a practice audit of 2335 examinations. Can J Gastroenterol Hepatol. 2014; 28:143–9. [PubMed: 24619636]
25. Naveau S, Lamouri K, Pourcher G, et al. The diagnostic accuracy of transient elastography for the diagnosis of liver fibrosis in bariatric surgery candidates with suspected NAFLD. Obes Surg. 2014; 24:1693–701. [PubMed: 24841950]
26. Lee HW, Park SY, Kim SU, et al. Discrimination of Nonalcoholic Steatohepatitis Using Transient Elastography in Patients with Nonalcoholic Fatty Liver Disease. PLoS One. 2016; 11:e0157358. [PubMed: 27284700]
27. Wong VW, Vergniol J, Wong GL, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology. 2010; 51:454–62. [PubMed: 20101745]
28. Siddiqui, Mohammad SRV., VanNatta, Mark, Brandman, Danielle, Tonascia, James T., Chalasani, Naga, Middleton, Michael S., Sanyal, Arun J. NASH-CRN. The diagnostic accuracy of vibration controlled transient elastography for cirrhosis in subjects with nonalcoholic fatty liver disease. Journal of Hepatology (Abstract EASL 2017). 2017
29. Sasso M, Beaugrand M, de Ledinghen V, et al. Controlled attenuation parameter (CAP): a novel VCTE guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. Ultrasound Med Biol. 2010; 36:1825–35. [PubMed: 20870345]
30. Sasso M, Audiere S, Kemgang A, et al. Liver Steatosis Assessed by Controlled Attenuation Parameter (CAP) Measured with the XL Probe of the FibroScan: A Pilot Study Assessing Diagnostic Accuracy. Ultrasound Med Biol. 2016; 42:92–103. [PubMed: 26386476]
31. Mueller S, Sandrin L. Liver stiffness: a novel parameter for the diagnosis of liver disease. Hepat Med. 2010; 2:49–67. [PubMed: 24367208]
32. Sasso M, Tengher-Barna I, Ziol M, et al. Novel controlled attenuation parameter for noninvasive assessment of steatosis using Fibroscan((R)): validation in chronic hepatitis C. J Viral Hepat. 2012; 19:244–53. [PubMed: 22404722]
33. Kwok R, Choi KC, Wong GL, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut. 2016; 65:1359–68. [PubMed: 25873639]
34. Tapper EB, Castera L, Afdhal NH. FibroScan (vibration-controlled transient elastography): where does it stand in the United States practice. Clin Gastroenterol Hepatol. 2015; 13:27–36. [PubMed: 24909907]
35. 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. e7. [PubMed: 26677985]
36. Karlas T, Petroff D, Garnov N, et al. Non-invasive assessment of hepatic steatosis in patients with NAFLD using controlled attenuation parameter and 1H-MR spectroscopy. PLoS One. 2014; 9:e91987. [PubMed: 24637477]
37. 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. [PubMed: 21898479]
38. Neuschwander-Tetri BA, Clark JM, Bass NM, et al. Clinical, laboratory and histological associations in adults with nonalcoholic fatty liver disease. Hepatology. 2010; 52:913–24. [PubMed: 20648476]

39. Hamilton, LC. Stata Technical Bulletin. Vol. 2. College Station, TX: Stata Press; 1991a. srd1: How robust is robust regression?; p. 21-26.Reprinted in Stata Technical Bulletin Reprints, vol. 1, pp. 169–175

40. Sirli R, Sporea I, Deleanu A, et al. Comparison between the M and XL probes for liver fibrosis assessment by transient elastography. Med Ultrason. 2014; 16:119–22. [PubMed: 24791843]

41. Vuppalanchi R, Jain AK, Deppe R, et al. Relationship Between Changes in Serum Levels of Keratin 18 and Changes in Liver Histology in Children and Adults With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol. 2014

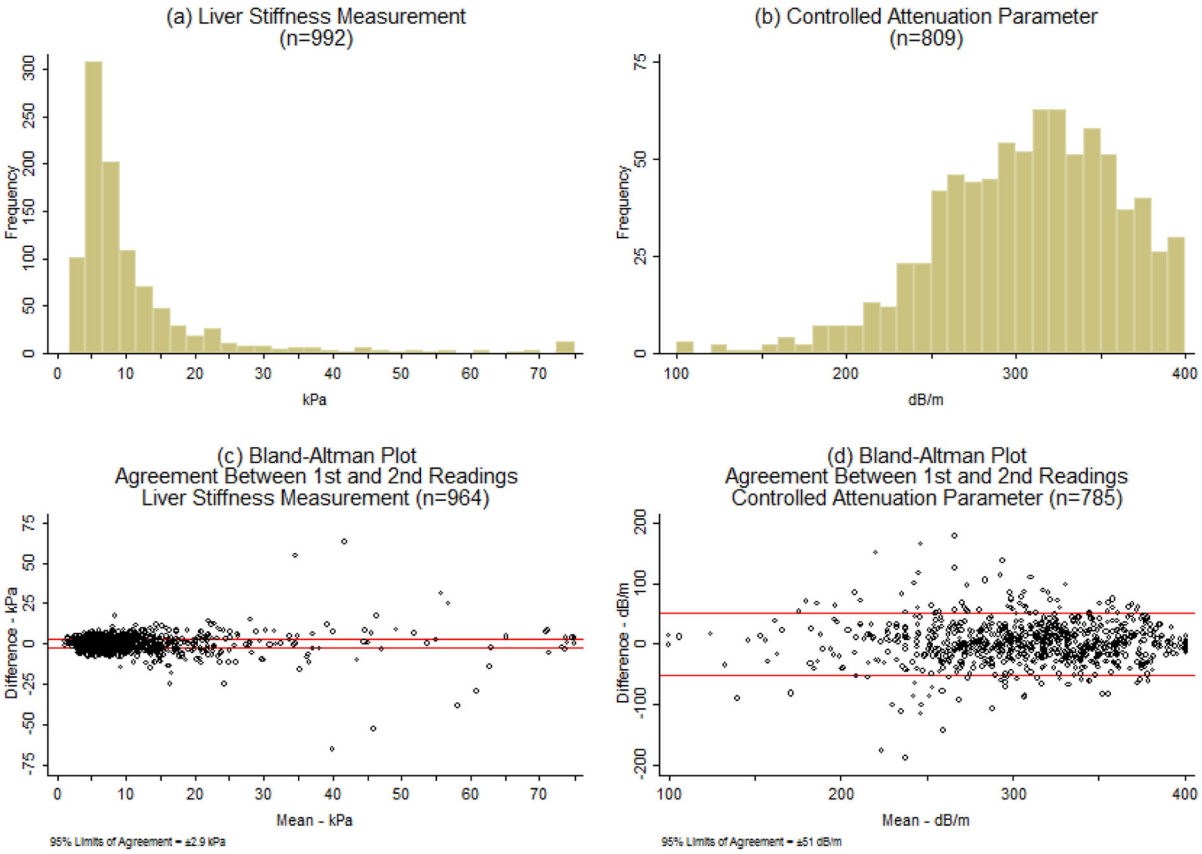

Figure 1.

Histogram of (a) Liver Stiffness Measurement (LSM); (b) Controlled Attenuation Parameter (CAP); (c) Bland-Altman plot of the difference in the 1st and 2nd reading of LSM values against the mean of the two measurements. The limits of agreement are indicated by the red lines—that is, the interval of 1.96 * standard deviation of the measurement differences. 95% of the difference between the 1st and 2nd readings occurred within ± 2.9 kPa with no bias over the range of values (4–75 kPa); (d) Bland-Altman plot of the difference in the 1st and 2nd reading of CAP values against the mean of the two measurements. 95% of the differences between the 1st and 2nd CAP readings occurred within ± 51 dB/m with no bias over the range of values (100–400 dB/m).

Hepatology. Author manuscript; available in PMC 2019 January 01.

Table 1

Selected demographic, histologic, and VTCE features by BMI category at the closest visit to VCTE evaluation. All values are mentioned in means (SD) unless otherwise specified.

	BMI category				
	Normal/Overweight* <30 kg/m^2 (n=311)	**Obese 30–34.9 kg/m^2 (n=311)**	**Extreme Obese 35 kg/m^2 (n=358)**	**Total† (n=992)**	**P-value**
Demographics					
Age, Years	53.5 (12.2)	51.5 (11.9)	49.6 (12.1)	51.4 (12.2)	
Gender - % Male	40.5	35.3	30.6	35.4	0.03
Race					<0.001
% White	81.6	82.1	87.1	84.1	
% Black	1.3	4.6	3.9	3.0	
% Asian	11.5	5.8	1.1	5.8	
% Other / refused	5.6	7.5	7.9	7.0	
Ethnicity - % Hispanic	12.1	10.7	7.9	10.2	0.18
Anthropometrics					
Waist circumference – cm	94 (8)	105 (8)	120 (11)	107 (14)	<0.001
Body mass index – kg/m^2	27.0 (2.3)	32.3 (1.4)	40.6 (4.7)	33.6 (6.5)	<0.001
Liver tests					
AST – U/L	39 (27)	42 (32)	45 (32)	42 (31)	
ALT – U/L	47 (36)	52 (41)	55 (42)	52 (40)	
Alkaline phosphatase – U/L	74 (29)	80 (30)	84 (33)	79 (31)	
Total bilirubin – mg/dL	0.8 (0.4)	0.7 (0.6)	0.7 (0.6)	0.7 (0.6)	
Serum chemistries					
Platelet count – 1K cells / uL	215 (62)	227 (73)	224 (74)	222 (70)	
International normalized ratio	1.05 (0.16)	1.04 (0.16)	1.07 (0.23)	1.05 (0.18)	
Histology at closest visit					
Sample size	288	288	342	893	
Years from biopsy	2.7 (2.8)	2.3 (2.6)	2.3 (2.6)	2.6 (2.6)	
NASH diagnosis					<0.001
No NAFLD - %	3.5	8.0	2.0	4.2	
NAFLD, no NASH - %	32.6	18.3	21.0	23.3	

	BMI category				
	Normal/Overweight* <30 kg/m^2 (n=311)	**Obese 30–34.9 kg/m^2 (n=311)**	**Extreme Obese 35 kg/m^2 (n=358)**	**Total† (n=992)**	**P-value**
Borderline - %	19.4	21.4	17.2	18.9	
Definite - %	44.4	52.2	59.8	53.6	
Fibrosis					0.01
None - %	32.3	26.0	21.6	25.9	
Stage 1 – mild - %	27.4	27.4	23.7	25.9	
Stage 2 – moderate - %	16.0	12.8	18.4	16.4	
Stage 3 – bridging - %	16.0	24.0	26.3	22.5	
Stage 4 – cirrhosis - %	8.3	9.7	9.9	9.4	
Mean stage	1.4 (1.3)	1.6 (1.3)	1.8 (1.3)		
Steatosis					0.27
Grade 0 - %	9.3	10.0	5.2	7.9	
Grade 1 - %	42.2	38.1	44.9	42.5	
Grade 2 - %	28.4	31.5	29.4	29.5	
Grade 3 - %	20.1	20.4	20.4	20.1	
Mean grade	1.6 (0.9)	1.6 (0.9)	1.7 (0.9)		

*55 normal patients + 256 overweight patients;

†12 patients with missing BMI data are included in the total

Hepatology. Author manuscript; available in PMC 2019 January 01.

Table 2

Performance characteristics of the VCTE in various categories of BMI.

	BMI category				
	Normal/Overweight* <30 kg/m² (n=311)	**Obese 30–34.9 kg/m² (n=311)**	**Extreme Obese 35 kg/m² (n=358)**	**Total† (n=992)**	**P-value**
Probe type - % XL probe	26.7%	60.8%	87.4%	59.7%	<0.001
Liver stiffness measurement					
Average 1st and 2nd reading					
Mean (SD) - kPa	9.0 (9.6)	11.0 (11.8)	13.8 (12.8)	11.4 (11.7)	<0.001
Median [IQR] - kPa	6.0 [4.4, 9.3]	7.1 [5.2, 11.8]	9.3 [6.6, 15.0]	7.5 [5.3, 12.2]	
Difference 1st and 2nd reading					
Mean (SD) - kPa	−0.2 (4.8)	0.3 (6.3)	0.0 (5.2)	0.0 (5.5)	0.96
Median [IQR] – kPa	0.0 [−0.7, 0.7]	0.0 [−0.7, 1.1]	0.1 [−1.3, 1.4]	0.0 [−0.9, 1.1]	
95% limits of agreement† - kPa					
Inter-reader (n=123)	±3.0	±3.3	±3.5	±3.6	
Intra-reader (n=838)	±2.0	±2.5	±4.1	±2.8	
Total (n=964)	±2.0	±2.6	±3.9	±2.9	
Controlled Attenuation Parameter (CAP)					
Average 1st and 2nd reading					
Mean (SD) – dB/m	284 (56)	303 (47)	336 (42)	307 (53)	<0.001
Median [IQR] – dB/m	284 [254, 321]	306 [270, 337]	340 [311, 366]	312 [272, 347]	
Difference 1st and 2nd reading					
Mean (SD) – dB/m	3 (32)	−2 (38)	5 (36)	2 (36)	0.07
Median [IQR] – dB/m	3 [−14, 18]	−3 [−21, 16]	0 [−16, 24]	0 [−16, 19]	
95% limits of agreement† - dB/m					
Inter-reader (n=114)	±52	±77	±70	±64	
Intra-reader (n=669)	±45	±49	±57	±49	
Total (n=785)	±46	±54	±58	±51	

*55 normal participants + 256 overweight participants;

†12 patients with missing BMI data are included in the total

Table 3

Proportion of operator's experience as quartiles and Unreliable VCTE scans

			Unreliable exams	
Quartile	**No. prior exams**	**Total exams**	**No.**	**%**
1	1 to 11	244	11	4.9
2	12 to 27	250	10	4.1
3	28 to 58	245	14	5.0
4	59 to 180	253	4	1.6

Table 4

Covariates of unreliable liver stiffness measurement *

	Odds of unreliable liver stiffness measurement		
Covariate	**Odds Ratio**	**95% CI**	**P value**
Body Mass Index (per category †)	2.7	1.5, 4.8	0.001
International normalized ratio (%)	0.94	0.89, 0.99	0.01
Ethnicity (Hispanic vs non-Hispanic)	3.2	1.1, 9.3	0.03
Age (per yr)	1.03	1.00, 1.07	0.08
Prior operator-specific readings (per quartile)	0.75	0.52, 1.08	0.12
ALT (per 10 U/L)	0.91	0.78, 1.05	0.19

* Unreliable liver stiffness defined as IQR/median > 30%. Overall rate of unreliable liver stiffness was 2.4% (39/1651).

Covariates were selected using AIC criteria from a multiple logistic model regressing unreliable liver stiffness measurement on candidate set of age, gender, ethnicity, race, BMI, waist circumference, mean and difference between 1st and 2nd readings of liver stiffness, INR, ALT, AST, bilirubin, alkaline phosphatase, platelet count, steatosis score, fibrosis stage, NASH diagnosis, time from biopsy, and same or different operator, prior operator-specific readings in NASHCRN patients

† BMI categorized as 1=normal/overweight; 2=obese; 3=very obese

Table 5

Factors associated with significant disagreement* between 1st and 2nd readings

	Odds of significant disagreement*		
	OR	**95% CI**	**P-value**
LSM†			
LSM (kPa)	1.06	1.04, 1.09	<0.001
Gender (male vs. female)	0.52	0.32, 0.86	0.01
Fibrosis (stage)	1.22	1.02, 1.46	0.03
AST (10 U/L)	1.07	1.01, 1.13	0.03
Race (white vs. non-white)	0.50	0.26, 0.96	0.04
Platelet count (10K cells/uL)	0.96	0.93, 1.00	0.04
Alkaline phosphatase (10 U/L)	0.99	0.98, 1.00	0.06
Operator (same vs different)	0.59	0.33, 1.06	0.08
Steatosis (score)	0.81	0.64, 1.02	0.08
Waist circumference (cm)	1.02	1.00, 1.04	0.09
Ethnicity (Hispanic vs. not Hispanic)	0.53	0.24, 1.21	0.13
Body mass index (/ category)	1.15	0.78, 1.70	0.47
CAP‡			
CAP (per 10 dB/m)	0.91	0.86, 0.96	0.001
Waist circumference (cm)	1.03	1.01, 1.05	0.006
Steatosis (grade)	0.67	0.49, 0.92	0.01
International normalized ratio (%)	1.01	1.00, 1.02	0.04
Age (years)	0.98	0.96, 1.00	0.05
Race (white vs. non-white)	0.50	0.25, 1.00	0.05

*Defined as absolute value of the difference > 95% limits of agreement between the 2 readings; Predictors were selected using AIC criteria from a multiple logistic model regressing significant disagreement on candidate set of age, gender, ethnicity, race, BMI category, waist circumference, INR, ALT, AST, bilirubin, alkaline phosphatase, platelet count, steatosis score, fibrosis stage, NASH diagnosis, time from biopsy, and operator

†rate of significant disagreement is 18% (169/964)

‡rate of significant disagreement is 11% (85/785)

Table 6

Relationship between probe type and liver stiffness and CAP measurements, adjusted for the body mass index

	XL+ probe		M+ probe		BMI-Adjusted Difference (XL – M)		
	N	BMI-Adjusted Mean	N	BMI-Adjusted Mean	Mean	95% CI	P-value
Liver stiffness (kPa)							
Fibrosis stage							
0	132	5.4	112	6.5	−1.0	−1.6, −0.5	<0.001
1	136	6.4	106	7.2	−0.9	−1.7, −0.1	0.03
2	90	7.2	57	7.8	−0.5	−1.7, 0.6	0.35
3	137	11.3	70	13.7	−2.4	−4.6, −0.2	0.04
4	57	21.3	30	26.2	−4.8	−12.6, 2.9	0.22
Total	552	8.0	375	8.8	−0.8	−1.4, −0.3	0.003
CAP (dB/m)							
Steatosis score							
0	20	268	41	289	−21	−50, 9	0.17
1	160	313	155	300	13	0, 25	0.05
2	117	321	106	306	15	1, 29	0.04
3	80	342	73	310	32	18, 46	<0.001
Total	377	319	375	303	16	8, 24	<0.001