**ORIGINAL ARTICLE**

# How Good Are Controlled Attenuation Parameter Scores from Fibroscan to Assess Steatosis, NASH, and Fibrosis?

Joshua Trowell[1] · Joseph Alukal[1] · Talan Zhang[1] · Li Liu[2] · Anurag Maheshwari[1,3] · Hwan Y. Yoo[1] · Paul J. Thuluvath[1,3]

## Abstract

**Background and Aims** The objective of our study was to determine the concordance rates of steatosis staging by controlled attenuation parameter (CAP) scores from transient elastography (TE) in comparison with liver histology in patients with chronic liver disease and to determine the optimal CAP cutoffs to predict the severity of steatosis and identify those with nonalcoholic steatohepatitis (NASH).

**Methods** Patients ($n = 217$) who had both CAP scores and liver biopsy within a period of 90 days were retrospectively studied. Histology was graded in a blinded fashion by a single pathologist; steatosis was graded on a scale from 0 to 3. Nonalcoholic fatty liver disease activity scores (NAS) scores were calculated for all patients. Optimal CAP cut-points were selected by maximum Youden's index.

**Results** Area under receiver operating characteristic curve (AUROC) for CAP (using cutoff value $\geq 278$ dB/m) in differentiating steatosis 1–3 from 0 was 0.82 (95% CI 0.75–0.89), and 0.79 (95% CI 0.70–0.88) in differentiating steatosis 0–1 from 2 to 3 using CAP cutoff value $\geq 301$ dB/m. With CAP cutoff value $\geq 301$ dB/m, CAP identified NAS 3 or above with AUROC of 0.82 (95% CI 0.74–0.89). The AUROC for TE in differentiating fibrosis (cutoff 11.9 kPa) 3–4 from 0 to 2 was 0.85 (95% CI 0.77–0.92), and 0.84 (95% CI 0.74–0.93) in differentiating (cutoff 14.4 kPa) 4 from 0 to 3.

**Conclusions** Transient elastography is a good modality to accurately diagnose steatosis and NASH and can also differentiate advanced liver fibrosis from early stages.

**Keywords** CAP scores · Liver histology · Steatosis · Fibrosis

## Abbreviations

CAP    Controlled attenuation parameter
NASH   Nonalcoholic steatohepatitis

✉ Paul J. Thuluvath
 thuluvath@gmail.com

[1] Institute of Digestive Health and Liver Diseases, Mercy Medical Center, Baltimore, MD 21202, USA

[2] Department of Pathology, Mercy Medical Center, Baltimore, MD 21202, USA

[3] University of Maryland School of Medicine, Baltimore, MD 21202, USA

## Introduction

Nonalcoholic fatty liver disease (NAFLD) has become a major public health concern since it is associated with an increase in cardiovascular, liver-related, and all-cause mortality [1, 2]. The current global prevalence of NAFLD, estimated at 25%, is projected to increase to 33.5% by 2030 unless there are universal life style modifications in the next decade [1]. The prevalence of nonalcoholic steatohepatitis (NASH) is < 10% in those with NAFLD, but with the growing epidemic of obesity and diabetes this number is expected to increase along with the complications stemming from NASH such as cirrhosis, liver failure and hepatocellular carcinoma (HCC) [2].

Liver biopsy remains the gold standard for the assessment of the severity of steatosis, inflammation and fibrosis in patients with NAFLD or other chronic liver diseases, but there is a growing need to develop a diagnostic tool that is noninvasive, cost-effective, and easily accessible to risk-stratify NAFLD [3]. Vibration Controlled Transient

Elastography (VCTE) (Fibroscan®, Echosens, Paris, France) has been used to assess the presence and severity of liver steatosis using controlled attenuation parameter (CAP) scores and severity of fibrosis using liver stiffness measurement (LSM) [4]. The reported accuracy of LSM to measure fibrosis is around 80%, but there are only few studies on the accuracy and optimal cutoff for CAP scores to diagnosis steatosis and its severity [3]. Moreover, many of these studies that had assessed the optimal cutoff were done in cohorts of patients who had exclusively NAFLD skewing the sensitivity, specificity and AUROC [4–10]. The objective of our study was to determine the accuracy of VCTE to assess steatosis and fibrosis in comparison with liver histology (gold standard) in an unselected patient population with NAFLD and non-NAFLD liver diseases.

## Patients and Methods

This was a single-center retrospective cohort study that included adult patients (age > 21 years) who underwent both liver biopsy and VCTE within a maximum interval of 90 days at our center. At our center, between August 2015 and February 2018, 3096 patients underwent TE and 724 had liver biopsy. Patients ($n = 217$) who had both VCTE and liver biopsy within 90 days of each other were selected. All patients were seen by one of three senior hepatologists. NAFLD and non-NAFLD were defined by history, risk factors, and laboratory tests. Alcohol and other causes of chronic liver diseases were excluded from the NAFLD group. The study was conducted after approval from the institutional review board, and a waiver of informed consent was applied since patient information was collected in a retrospective manner.

We collected demographic and clinical information on all patients using electronic medical records. The following information was collected: age, sex, race, body mass index (BMI), presence of type 2 DM or insulin resistance, liver biochemistry, platelet counts, APRI (AST to Platelet Ratio Index), the type of VCTE probe used, and liver histology including NAFLD activity scores (NAS). Based on the cause of liver disease, patients were classified as having NAFLD or non-NAFLD group.

The primary objective of our study was to evaluate diagnostic performance of CAP by TE by comparing it with liver histology for the detection and staging of steatosis in patients with chronic liver diseases. Our secondary outcome was to determine whether CAP scores from VCTE correlated with NAS scores. We also wanted to determine: (a) cut-points for CAP to identify patients into NAS 3 or above and NAS < 3; (b) cut-points for CAP to differentiate the diagnosis of steatosis 1–3 from 0, and steatosis 0–1 from 2 to 3; and (c) cut-points for TE to differentiate advanced fibrosis (F3–4)

from non-advanced fibrosis (F0–F2), and cirrhosis (F4) from non-cirrhosis (F0–3).

## Liver Histology Assessment

All liver biopsies were re-read and graded by a single pathologist who was blinded to clinical and VCTE data. Histopathological findings were reported according to the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) scoring system. Although NASH CRN scoring was designed for patients with NASH, we used it for both NAFLD and non-NAFLD group for the purpose of this study. NAS was assessed as previously reported on a scale of 0–8 [sum of scores for hepatic steatosis (0–3), lobular inflammation (0–3) and hepatocyte ballooning (0–2)]. Hepatic steatosis was graded from 0 to 3: S0 = steatosis < 5%, S1 = steatosis 5–33%, S2 = steatosis 33–66% and S3 = steatosis > 66%. Steatosis was considered significant at a grade of ≥ S1. Fibrosis was graded (METAVIR scoring system) as follows: F0 = no fibrosis, F1 = mild fibrosis, F2 = moderate fibrosis, F3 = severe fibrosis, and F4 = cirrhosis.

## Vibration Controlled Transient Elastography (TE)

VCTE was performed using a Fibroscan® 502 Touch machine with the option to select an M or XL size probe. The type of probe was automatically selected by the software included in the machine. VCTE was performed by trained operators who were blinded to the results of the liver biopsy and to the clinical data. All patients underwent the procedure after a 3 h fast. Patients who were included in the analysis had 10 or more valid measurements with a success rate (valid measurements/total measurement) ≥ 60% and IQR/median ≤ 30% for LSM.

## Statistical Analysis

Descriptive statistics for characteristics of patients were presented as means and standard deviations (SDs) for continuous variables, and frequencies for categorical variables. The patients' characteristics difference between NAFLD and non-NAFLD was assessed by using Chi square test for categorical variables, and T test for continuous variables; normality was checked for all the continuous variables, nonparametric Wilcoxon test was used when data were not normally distributed, and a variable with $p$ value ≤ 0.05 indicates a significant difference between two groups. Spearman correlation test was used to assess the correlation between CAP and NAS score, CAP and Steatosis grade, and TE and fibrosis stage.

In order to find optimal cutoffs, data were randomly divided into a training dataset and a validation dataset using

a 2:1 ratio. The training dataset was used to build a model, while the validation dataset was used to assess the model's predictive ability. Receiver operating characteristic (ROC) analysis was performed to determine optimal cutoffs by maximum Youden's index. Cut-points were derived from the training data, and the performances of cut-points were further tested on the validation data.

As an exploratory analysis, we further assessed factors which had influences on misclassification in steatosis grade, NAS 3 or above and fibrosis stage separately based on the cutoffs. Misclassification errors were defined as either false positive or false negative in differentiating disparate steatosis grade, NAS 3 or above or fibrosis stage, separately. Logistic regression was performed to evaluate each factor's influence on misclassification in steatosis, NAS and fibrosis stage separately. For example, a patient was misclassified if the patient was falsely classified as NAS 3 or above based on Fibroscan, while NAS was below 3 on liver histology, or the patient was falsely classified as below NAS 3 by Fibroscan, while NAS was 3 or above by histology. We started with univariate analysis, followed by multivariate analysis using a forward model selection approach. The final model was selected by balancing goodness of fit (e.g., Bayesian information criteria). Any variable with univariate effect ($p$ value $\leq 0.05$) was considered a candidate for initial multivariate modeling. The final model retained variables with a $p$ value 0.05 or less. Estimations of adjusted odds ratios and 95% CIs were reported. Data were analyzed using SAS 9.4 (SAS Institute, Cary, North Carolina, USA).

## Results

Of the entire cohort, 92 patients had NAFLD and 125 had non-NAFLD. In the non-NAFLD group, 62 (50%) had hepatitis C and 24% had abnormal liver enzymes without any definitive liver pathology. Hepatitis B, alcoholic liver disease, and cryptogenic cirrhosis accounted for the rest. The clinical characteristics of the study cohort are shown in Table 1. Patients were mostly middle aged ($55.6 \pm 11.8$ years), 54.4% were female, and 51.6% were obese (BMI $\geq 30$). Type 2 diabetes was found in 40 (43%) NAFLD patients versus 27 (22%) patients with non-NAFLD. Blacks comprised 30% of the entire cohort, but only 13% of NAFLD group.

Out of the entire cohort of patients ($n = 217$), 68 (32%) had a NAS score of 3 or more and all of them, except one, were found in the NAFLD group (Table 1). By liver histology, 12 (6%) patients had no fibrosis and 38 (18%) patients had cirrhosis (Table 1). XL probe was used in 67 (31%) patients, but it was used more often in NAFLD patients (45%). The mean IQR for VCTE was 18%.

Training and validation patients were randomly split into groups of 151 and 66 patients, respectively. Patient characteristics were similar in both training and validation datasets. Spearman correlation showed that there was a significant correlation between CAP and steatosis grade (Spearman correlation: 0.57, 95% CI 0.47, 0.65, $p < 0.0001$), CAP and NAS score (Spearman correlation: 0.61, 95% CI 0.52, 0.69, $p < 0.0001$), and TE and fibrosis stage (Spearman correlation: 0.52, 95% CI 0.42, 0.61, $p < 0.0001$). The analysis was done in NAFLD group, and those results are shown in the supplementary table.

### Diagnostic Accuracy of CAP for the Estimation of Steatosis Grade

The sensitivity and specificity to diagnose the severity of steatosis varied based on the cutoff values used for CAP scores. Using the maximum value of Youden's index, the AUROC for steatosis grade $\geq 1$ with CAP cutoff value of 278 dB/m was 0.82 (95% CI 0.75–0.89) and for steatosis grade $\geq 2$ with CAP cutoff value of 301 dB/m was 0.79 (95% CI 0.70–0.88) based on training data (Table 2, Fig. 1). Applying CAP on the validation data set gave higher AUROC: for steatosis grade $\geq 1$ with CAP cutoff value of 278 dB/m, it was 0.84, and for steatosis grade $\geq 2$ with CAP cutoff value of 301 dB/m, it was 0.82 (Table 2, Fig. 1). Sensitivity and specificity for different cutoff values are shown in Table 2.

### Diagnostic Accuracy of CAP for the Estimation of NAS

AUROC for CAP in differentiating NAS 3 or above from NAS $< 3$ using CAP cutoff value of 301 dB/m was 0.82 (95% CI 0.74–0.89) based on training data. Validation data depicted an AUROC of 0.80 (Table 2; Fig. 2).

### Diagnostic Accuracy of TE for the Estimation of Fibrosis

Liver stiffness increased with increasing fibrosis. Based on the training data, the AUROC for TE with cutoff 11.9 kPa in differentiating fibrosis 3–4 from 0 to 2 was 0.85 (95% CI 0.77–0.92), and 0.84 (95% CI 0.74–0.93) for TE with cutoff 14.4 kPa in differentiating fibrosis 4 from 0 to 3. AUROC based on validation data for TE in differentiating fibrosis 3–4 from 0 to 2 and fibrosis 4 from 0 to 3 was 0.78 and 0.86 with TE cutoffs 11.9 kPa and 14.4 kPa, respectively (Table 2; Fig. 3).

**Table 1** Patient characteristics

| Characteristics | Total patients ($n = 217$) | NAFLD ($n = 92$) | Not NAFLD ($n = 125$) | $p$ value |
|---|---|---|---|---|
| Age, mean (years) | $55.6 \pm 11.8$ | $55.3 \pm 12.6$ | $55.8 \pm 11.2$ | 0.91 |
| Female, $n$ (%) | 118 (54%) | 52 (57%) | 66 (53%) | 0.59 |
| *Race* | | | | |
| White, $n$ (%) | 143 (66%) | 72 (78%) | 71 (57%) | <0.001 |
| Black, $n$ (%) | 65 (30%) | 12 (13%) | 53 (42%) | |
| Other, $n$ (%) | 9 (3%) | 8 (8%) | 1 (1%) | |
| *Ethnicity* | | | | |
| Non-Hispanic, $n$ (%) | 164 (76%) | 70 (76%) | 94 (75%) | 0.11 |
| Type 2 diabetes, $n$ (%) | 67 (31%) | 40 (43%) | 27 (22%) | <0.001 |
| BMI, Mean ($\geq 30$ kg/m$^2$) ($\pm$ SD) | $30.9 \pm 6.22$ | $33.0 \pm 5.72$ | $29.3 \pm 6.15$ | <0.001 |
| *Laboratory results* | | | | |
| ALT, mean ($\pm$ SD) | 48.9 (72.3) | 56.2 (48.4) | 56.2 (48.4) | 0.18 |
| Alkaline phosphatase, mean ($\pm$ SD) | 104 (87.0) | 95.8 (41.1) | 108 (108) | 0.23 |
| Bilirubin, total, mean ($\pm$ SD) | 0.59 (0.63) | 0.58 (0.42) | 0.59 (0.75) | 0.91 |
| INR, Mean ($\pm$ SD) | 1.07 (0.25) | 1.06 (0.20) | 1.08 (0.28) | 0.60 |
| Albumin, mean ($\pm$ SD) | 4.29 (0.43) | 4.42 (0.37) | 4.20 (0.45) | <0.001 |
| Creatinine, mean ($\pm$ SD) | 0.94 (0.54) | 0.89 (0.22) | 0.96 (0.68) | 0.29 |
| Platelet count, mean (1000 cells/UL) ($\pm$ SD) | 237 (69.1) | 243 (78.1) | 233 (64.5) | 0.30 |
| APRI, mean ($\pm$ SD) | 0.75 (3.55) | 0.55 (0.58) | 43.69 (85.04) | 0.44 |
| *Histology* | | | | |
| Ballooning grade, $n$ (%) | | | | <0.001 |
| Grade 0 | 139 (64%) | 19 (21%) | 120 (96%) | |
| Grade 1 | 51 (24%) | 48 (52%) | 3 (2%) | |
| Grade 2 | 27 (12%) | 25 (27%) | 2 (2%) | |
| Lobular Inflammation, $n$ (%) | | | | <0.001 |
| Grade 0 | 146 (67%) | 23 (25%) | 123 (98%) | |
| Grade 1 | 52 (24%) | 50 (54%) | 2 (2%) | |
| Grade 2 | 15 (7%) | 15 (16%) | 0 (0%) | |
| Grade 3 | 4 (2%) | 4 (4%) | 0 (0%) | |
| Steatosis, $n$ (%) | | | | <0.001 |
| Grade 0 | 123 (57%) | 11 (12%) | 112 (90%) | |
| Grade 1 | 58 (27%) | 45 (49%) | 13 (10%) | |
| Grade 2–3 | 36 (16%) | 36 (39%.0) | 0 (0.0%) | |
| NAFLD activity score (NAS), $n$ (%) | | | | <0.001 |
| 0 | 119 (55%) | 8 (9%) | 111 (89%) | |
| 1 | 15 (7%) | 6 (7%) | 9 (7%) | |
| 2 | 15 (7%) | 11 (12%) | 4 (3%) | |
| $3 \geq$ | 68 (32.0%) | 67 (73%) | 1 (1%) | |
| Fibrosis, $n$ (%) | | | | 0.093 |
| Stage 0 | 12 (6%) | 2 (2%) | 10 (8%) | |
| Stage 1 | 85 (39%) | 32 (35%) | 53 (42%) | |
| Stage 2 | 53 (24%) | 24 (26%) | 29 (23%) | |
| Stage 3 | 29 (13%) | 12 (13%) | 17 (14%) | |
| Stage 4 | 38 (18%) | 22 (24%) | 16 (13%) | |
| *Transient elastography (TE)* | | | | |
| Probe size, $n$ (%) | | | | <0.001 |
| M | 150 (69.1) | 51 (55.4) | 99 (79.2) | |
| XL | 67 (30.9) | 41 (44.6) | 26 (20.8) | |
| CAP Score, mean ($\pm$ SD) | $275 \pm 66.1$ | $318 \pm 47.9$ | $244 \pm 60.0$ | <0.001 |
| Fibrosis Score, mean ($\pm$ SD) | $12.8 \pm 10.9$ | $14.4 \pm 10.9$ | $11.5 \pm 10.9$ | <0.001 |

**Table 1** (continued)

| Characteristics | Total patients ($n = 217$) | NAFLD ($n = 92$) | Not NAFLD ($n = 125$) | $p$ value |
|---|---|---|---|---|
| Fibrosis IQR, mean ($\pm$ SD) | 0.86 $\pm$ 0.18 | 0.86 $\pm$ 0.18 | 0.86 $\pm$ 0.18 | 0.74 |
| *Interval between biopsy and TE (days)* | | | | |
| Mean, mean ($\pm$ SD) | 25.2 $\pm$ 1.5 | 25.1 $\pm$ 2.3 | 25.7 $\pm$ 1.9 | |

## Factors That Had Influences on Misclassification

### Steatosis Grade

Age, gender, BMI, T2DM, total bilirubin, glucose, and hemoglobin had individual influences on misclassifying steatosis (Table 3). In multivariate model, male, higher BMI and patients with T2DM had higher odds to be misclassified (Table 3).

### NAS 3 or Above

Total bilirubin and platelets had individual influence on misclassifying NAS (Table 3). In multivariate model, higher platelets and patients with T2DM had higher odds to be misclassified (Table 3).

### Fibrosis Stage

Only age had individual influences on misclassifying fibrosis stage (Table 3). In multivariate model, older patients, higher BMI, and patients with higher ALT had higher odds to be misclassified (Table 3).

## Discussion

In an unselected patient population with NAFLD and non-NAFLD, we compared the diagnostic accuracy of CAP scores to estimate the severity of steatosis using liver biopsy as the "gold standard". Our study demonstrated that CAP scores can be reliably used to diagnose steatosis of grade $\geq 1$ at a cutoff of 278 dB/m (AUROC of 0.81) and grade $\geq 2$ at a cutoff of 301 dB/m (AUROC of 0.79) and corroborate with results from other similar studies [7, 9, 12]. The AUROC to diagnose NASH (NAS score $\geq 3$) using a cutoff CAP score of 301 dB/m was 0.82 again, indicating the validity of this noninvasive diagnostic modality. The study also showed that liver stiffness measurement (LSM) scores can be used to accurately differentiate advanced fibrosis (F3 and F4) from early stage (F0–F2) as well as cirrhosis (F4) from any stage.

To date, liver biopsy remains the best available test to diagnose and stage NAFLD. Even though there are studies on the accuracy of TE to diagnose steatosis and fibrosis, majority of these studies were done exclusively in patients with NAFLD. Our study is different since we explored the utility of TE in NAFLD and non-NAFLD patients. The results from our study our encouraging and implies the wider applicability of TE to assess steatosis in patients with any form of chronic liver diseases. Different studies have reported cutoff values of varying ranges to grade steatosis. A recent prospective study analyzed 393 patients with NAFLD to evaluate the accuracy of VCTE and found that CAP scores were able to detect $> 5\%$ steatosis with an AUROC of 0.76, but was neither accurate to differentiate between higher steatosis grades nor to diagnose NASH since the AUROC was suboptimal [7]. Another prospective study from the UK studied 404 patients who had a liver biopsy within two weeks of Fibroscan and demonstrated that a CAP score cutoff of 302 dB/m could accurately diagnose steatosis ($> S1$) with an AUROC of 0.87 (95% CI 0.82–0.92) [11]. In a meta-analysis of 24 studies on CAP accuracy, it was determined that a cutoff of 214 dB/m diagnosed steatosis grade $\geq 1$, a cutoff of 255 dB/m diagnosed steatosis grade $\geq 2$, and a cutoff of 281 determined steatosis grade 3 [12]. Another recent individual patient data meta-analysis reports a CAP cutoff of 248 dB/m (237–261) to identify steatosis $> S0$ (AUROC of 0.82) and 268 dB/m (257–284) for $> S1$ (AUROC of 0.86) [13]. Using maximum value of Youden's index, we found that higher CAP cutoff values were necessary to improve the accuracy of staging steatosis in our population. In our study, using the maximum value of Youden's index, the AUROC for steatosis grade $\geq 1$ with CAP cutoff value of 278 dB/m was 0.82 (95% CI 0.75–0.89) and for steatosis grade $\geq 2$ with CAP cutoff value of 301 dB/m was 0.79 (95% CI 0.70–0.88) in the training group with marginally higher AUROC (0.84 and 0.82, respectively) in the validation group. We do not have an explanation for this, but it could be related to metabolic factors or body habitus of our cohort, but it is important to note that our cutoff values are similar to a study published from the UK [11].

It is interesting to point out that in our group of patients, a cutoff CAP score of 301 dB/m was able to diagnose NASH (NAS score $\geq 3$) with an AUROC of 0.82. In the aforementioned studies neither CAP nor LSM was able to capture this diagnosis [7, 11]. In the US cohort, the AUROC to diagnose NASH using CAP scores was 0.58 and in the UK cohort the AUROC using CAP scores was 0.71 [7, 11]. The discordance between these studies merits further investigation. A noninvasive diagnostic technique that can distinguish NASH

**Table 2** Diagnostic accuracy of CAP and liver stiffness measurement in differentiating steatosis grade, NAS, and fibrosis stage

| | Cutoffs (dB/m; kPa) | Training | | | | | Validation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | AUROC (95% CI) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | AUROC (95% CI) |
| *Diagnostic accuracy of CAP (dB/m) scores in differentiating Steatosis grade (0, 1, 2–3)* | | | | | | | | | | | |
| Steatosis 1–3 versus 0 | 278 | 81.54 | 72.09 | 68.83 | 83.78 | 0.82 (0.75, 0.89) | 75.86 | 81.08 | 75.86 | 81.08 | 0.84 (0.75, 0.94) |
| Steatosis 2–3 versus 0–1 | 301 | 80.00 | 69.84 | 34.48 | 94.62 | 0.79 (0.70,0.88) | 81.82 | 72.73 | 37.50 | 95.24 | 0.82 (0.72, 0.93) |
| *Diagnostic accuracy of CAP (dB/m) scores in differentiating NAS (≥3 and <3)* | | | | | | | | | | | |
| NAS≥3 versus <3 | 301 | 77.27 | 77.57 | 58.62 | 89.25 | 0.82 (0.74, 0.89) | 82.35 | 79.59 | 58.33 | 92.86 | 0.80 (0.68, 0.92) |
| *Diagnostic accuracy of liver stiffness measurement (kPa) in differentiating fibrosis stage (F0–F2, F3 and F4)* | | | | | | | | | | | |
| F3–4 versus F0–2 | 11.90 | 75.00 | 81.55 | 65.45 | 87.50 | 0.85 (0.77, 0.92) | 73.68 | 74.47 | 53.85 | 87.50 | 0.78 (0.66, 0.90) |
| F4 versus F0–3 | 14.40 | 80.00 | 83.47 | 54.55 | 94.39 | 0.84 (0.74, 0.93) | 75.00 | 79.31 | 33.33 | 95.83 | 0.86 (0.75, 0.97) |

*PPV* positive predictive value, *NPV* negative predictive value

from the vast majority of patients with simple steatosis (fatty liver) with fair accuracy could potentially have significant clinical and research implications, especially when it comes to enrolling patients for clinical trials to reduce failure rates.

Since there is evidence that links advanced fibrosis with all-cause mortality, there is a critical need to identify this subset of NASH population early in the disease course so that therapeutic intervention can be applied in a timely fashion. Although histology is considered as gold standard for staging fibrosis, biopsy is associated with complications and subject to sampling error as well as inter and intra-observer variability. Reports from previous studies have indicated that VCTE can accurately demarcate advanced fibrosis from early stages, suggesting that a biopsy may be unnecessary in patients with early fibrosis [11]. The optimal LSM cutoff score in our study using Youden's index to distinguish advanced fibrosis (F3 and F4) from early stage was 11.9 kPa (sensitivity of 75% and specificity of 81.6%). Our cutoff values are higher than previously reported in the US (8.6 kPa) and UK (9.7 kPa) cohort studies [7, 11]. However, the LSM cutoff score to diagnose cirrhosis (F4) in our study of 14.4 kPa (sensitivity of 80% and specificity of 83.47%) is similar to two previous studies (13.1 kPa in one study and 13.6 kPa in another study [7, 11].

For wider application of VCTE, the cutoff values need to be defined using larger unselected cohorts. There have been many studies that had explored the utility of LSM by VCTE for assessing liver fibrosis, and these reports had suggested varying cutoff values (stage 0 < 5.5 kPa, 5.6–6.5 kPa for F1, 6.6–7.8 kPa for F2, 7.1–10.4 kPa for F3, and 10.3–22.3 kPa for F4 [3]. Although LSM has been reported to have excellent AUROC for advanced fibrosis, it is not an ideal tool to differentiate F2 and F3 [14]. Previous studies have suggested advanced steatosis or inflammation could increase the likelihood of over staging fibrosis in the presence of NASH and may be a contributing factor for failure of VCTE in a subset of patients [15–17]. Reports have also indicated that the cutoff value for CAP and LSM scores may vary depending on the type of probe used (M vs XL); however, this could be related to the lack of availability of the XL probe in some of these studies involving extremely obese patients [9, 10]. Even though our study did not explore the failure rates or differentiate CAP and LSM scores based on the type of probe, it is worth pointing out that 45% of our NAFLD patients were assessed using the XL probe as prompted by Fibroscan and this may explain the higher AUROC value in our cohort for CAP and LSM scores. According to data from a recent UK cohort, neither probe type nor steatosis had any association with LSM and the only histological parameter that could influence LSM was the degree of fibrosis [11]. We looked into factors that could contribute to misclassification/failure of Fibroscan in grading steatosis and fibrosis. Multivariate analysis revealed that patients with type 2 DM
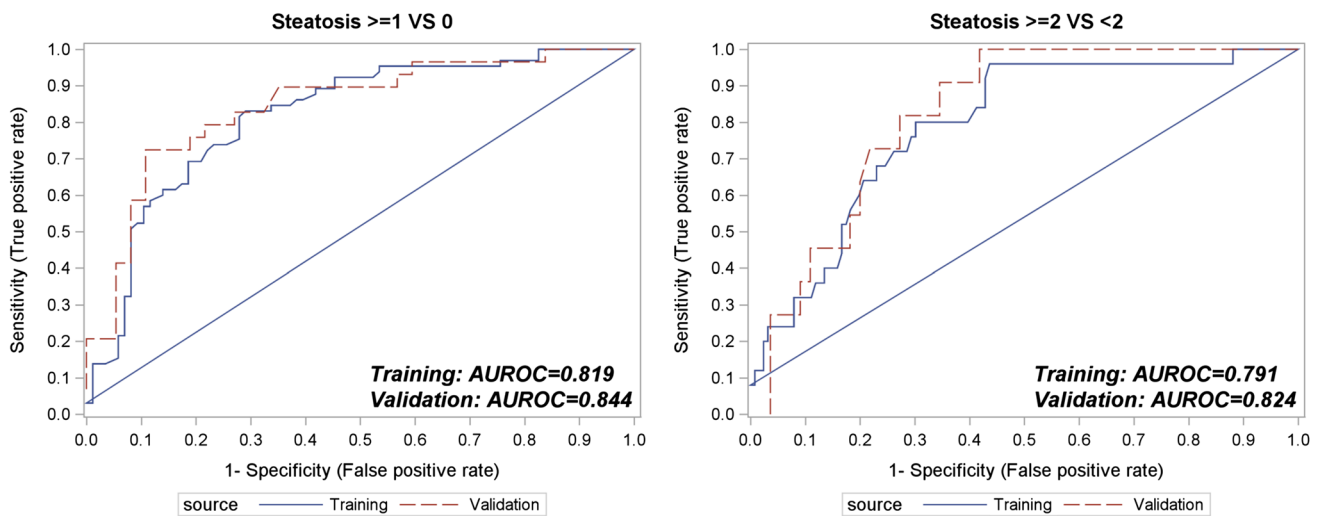
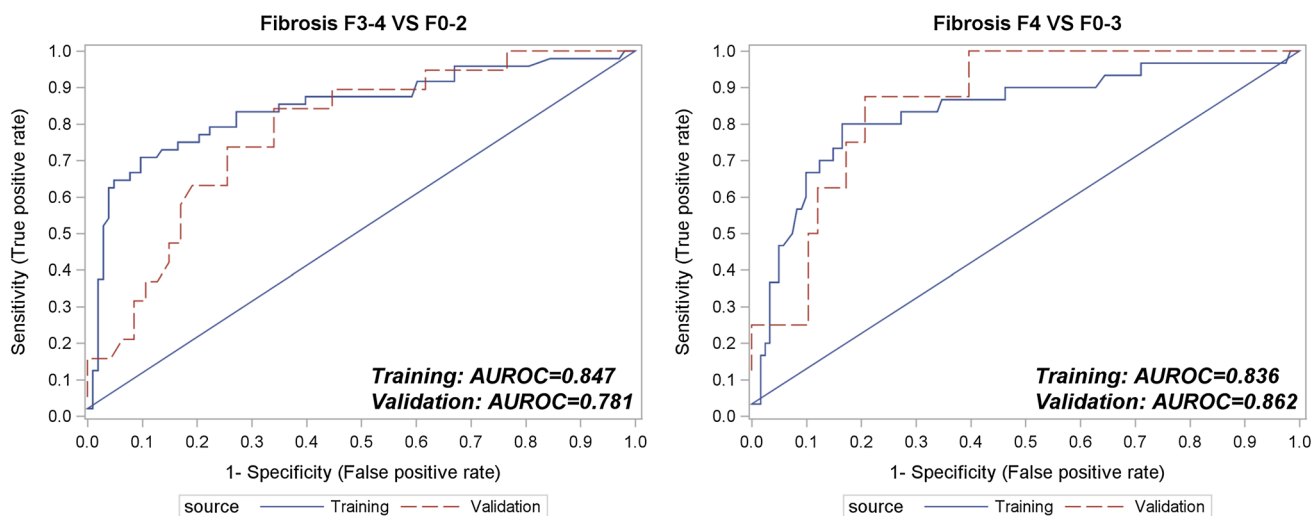Fig. 1 ROC curve of CAP in predicting steatosis ≥ 1 or ≥ 2. AUROCs are shown in training and validation datasets separately
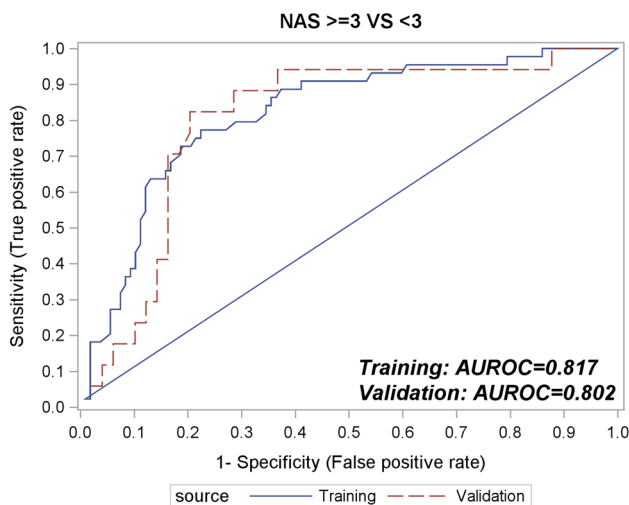


Fig. 2 ROC curve of CAP in predicting NAS ≥ 3. AUROCs are shown in training and validation datasets separately

and higher BMI had a higher chance of being misclassified into a higher steatosis grade, while variables associated with LSM failure included older age, high BMI, and higher ALT. Previous studies had also reported that factors such as increased BMI, necroinflammation (high ALT), cholestasis (higher alkaline phosphatase), and presence of right heart failure may affect the ability of Fibroscan to assess fibrosis accurately; however, data are limited on confounders for grading steatosis [18].

The use of VCTE in clinical practice has many benefits including relatively low cost, application at the bedside in outpatient setting,and no procedure-related complication, and yet it cannot entirely replace liver histology because of ~ 20% false positive and negative results. Another

noninvasive imaging method that is gaining popularity to diagnose and quantify steatosis and fibrosis is magnetic resonance imaging (MRI). A prospective cross-sectional study involving 104 patients compared the efficacy of magnetic resonance elastography (MRE) and MRI-proton density fat fraction (MRI-PDFF) versus Fibroscan (M and XL probe) with liver biopsy as reference, and showed that MRI-PDFF was superior to CAP score in diagnosing any steatosis (grade 1–3 vs 0 with AUROC of 0.99), and MRE was more accurate than TE in diagnosing fibrosis of any stage (stage F1–F4 vs F0) [5]. Despite the higher accuracy of MRE, the high cost and limited availability of this imaging technique make it less attractive [2–6]. It is very unlikely that MRE will replace Fibroscan based on the current availability and costs for longitudinal assessment of steatosis and fibrosis [9].

Our study has few clinical implications and limitations. The sensitivity and specificity for diagnosing steatosis using VCTE is good, and more importantly, CAP scores showed a good correlation with NASH severity (NAS ≥ 3) when appropriate cutoff values were utilized. Therefore, VCTE could be valuable for screening patients for NAFLD trials to reduce screen failure rates, and possibly for treatment decisions in the future. The major limitation of our study is the retrospective nature of data collection despite the "blinded" analysis of histology. Another limitation is the relatively small sample size, but unlike some other previous studies, we did not use an exclusive population of NAFLD which improved the validity of our observations and reduced the bias. Despite these limitations, our study is robust since we compared the validity of Fibroscan against liver biopsy which is considered the gold standard and confirmed our observations both in the training and validation cohorts. Also, the availability of M and XL probe in our study reduced some of the shortcomings encountered in

**Fig. 3** ROC curve of TE in predicting Fibrosis F3–4 or cirrhosis F4. AUROCs are shown in training and validation datasets separately

**Table 3** Patient and biochemical characteristics that potentially caused misclassification on univariate and multivariate analysis

| Variable | Steatosis grade | NAS $\geq 3$ | Fibrosis stage |
|---|---|---|---|
| *Univariate analysis* | | | |
| Age | 0.01 | 0.19 | 0.03 |
| Gender | 0.02 | 0.40 | 0.12 |
| Ethnicity | 0.93 | 0.62 | 1.00 |
| Hispanic | 0.91 | 0.26 | 0.86 |
| BMI | 0.003 | 0.44 | 0.13 |
| T2DM | 0.007 | 0.11 | 0.14 |
| Diagnosis | 0.61 | 0.99 | 0.93 |
| AST | 0.34 | 0.52 | 0.34 |
| ALT | 0.19 | 0.13 | 0.08 |
| APRI | 0.42 | 0.28 | 0.32 |
| ALP | 0.50 | 0.12 | 0.21 |
| Total bilirubin | 0.04 | 0.03 | 0.25 |
| Direct bilirubin | 0.28 | 0.12 | 0.52 |
| Albumin | 0.97 | 0.75 | 0.27 |
| Glucose | 0.02 | 0.73 | 0.16 |
| Hemoglobin | 0.03 | 0.62 | 0.78 |
| Triglycerides | 0.50 | 0.91 | 0.95 |
| Total cholesterol | 0.77 | 0.24 | 0.75 |
| HDL cholesterol | 0.34 | 0.73 | 0.34 |
| LDL cholesterol | 0.49 | 0.28 | 1.00 |
| Platelets | 0.64 | 0.02 | 1.00 |
| INR | 0.20 | 0.19 | 0.33 |
| Creatinine | 0.32 | 0.19 | 0.23 |
| *Multivariate analysis OR (95% CIs); p values* | | | |
| Age | | | 1.05 (1.01, 1.10); 0.01 |
| Female (ref = M) | 2.07 (1.16, 3.67); 0.01 | | |
| BMI | 1.07 (1.02, 1.13); 0.01 | | 1.05 (1.00, 1.10); 0.08 |
| T2DM (ref = no) | 1.90 (1.03, 3.51); 0.04 | 2.05 (1.00, 4.20); 0.05 | |
| ALT | | | 1.01 (1.00, 1.01); 0.03 |
| Platelets | | 1.01 (1.00, 1.01); 0.02 | |

prior studies. Future studies should explore the validity of pre-determined cutoff values in a prospective manner in a cohort of patients with and without NAFLD. Larger studies should also establish optimal cutoff values for both CAP and LSM scores using either M or XL probes to determine the severity of steatosis, NASH, and fibrosis.

## Compliance with Ethical Standards

## References

1. Loomba R, Sanyal AJ. The global NAFLD epidemic. *Nat Rev Gastroenterol Hepatol*. 2013;10:686–690.
2. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. *JAMA*. 2015;313:2263–2273.
3. Cheah MCC, McCullough AJ, Goh GB. Current modalities of fibrosis assessment in non-alcoholic fatty liver disease. *J Clin Transl Hepatol*. 2017;5:261–271.
4. Imajo K, Kessoku T, Honda T, 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.
5. 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.
6. Caussy C, Alquiraish MH, Nguyen P, et al. Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. *Hepatology*. 2018;67:1348–1359.
7. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al. Vibration-controlled transient elastography to assess fibrosis and steatosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol*. 2019;17:156–163.
8. Petta S, Wong VW, Cammà C, et al. Improved noninvasive prediction of liver fibrosis by liver stiffness measurement in patients with nonalcoholic fatty liver disease accounting for controlled attenuation parameter values. *Hepatology*. 2017;65:1145–1155.
9. Vuppalanchi R, Siddiqui MS, Van Natta ML, et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. *Hepatology*. 2018;679:134–144.
10. 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;111:677–684.
11. Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2019;156:1717–1730.
12. Mikolaseviv I, Orlic L, Franjic N, et al. Transient elastography (Fibroscan) with controlled attenuation parameter in the assessment of liver steatosis and fibrosis in patients with nonalcoholic fatty liver disease—where do we stand? *World J Gastroenterol*. 2016;22:7236–7251.
13. Karlas T, Petroff D, Sasso M, Fan JG, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *J Hepatol*. 2017;66:1022–1030.
14. Kwok R, Tse YK, Wong GL, et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease—the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment Pharmacol Ther*. 2014;39:254–269.
15. Arena U, Vizzutti F, Corti G, et al. Acute viral hepatitis increases liver stiffness values measured by transient elastography. *Hepatology*. 2008;47:380–384.
16. Sagir A, Erhardt A, Schmitt M, Häussinger D. Transient elastography is unreliable for detection of cirrhosis in patients with acute liver damage. *Hepatology*. 2008;47:592–595.
17. Millonig G, Reimann FM, Friedrich S, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *Hepatology*. 2008;48:1718–1723.
18. Tapper EB, Castera L, Afdhal NH. Fibroscan (vibration-controlled transient elastograpy): where does it stand in the United States practice. *Clin Gastroenterol Hepatol*. 2015;13:27–36.