







Accuracy of Transient Elastography in Assessing Fibrosis at Diagnosis in Naïve Patients With Primary Biliary Cholangitis: A Dual Cut-Off Approach

Laura Cristoferi ^{1,3*}, Vincenza Calvaruso,^{4*} Diletta Overi ⁵, Mauro Viganò,⁶ Cristina Rigamonti,⁷ Elisabetta Degasperi ⁸, Vincenzo Cardinale,⁹ Sara Labanca,¹⁰ Nicola Zucchini,¹¹ Anna Fichera,⁴ Vito Di Marco,⁴ Monica Leutner,¹² Rosanna Venerè,¹³ Antonino Picciotto,¹⁰ Martina Lucà,^{1,2} Giacomo Mulinacci,^{1,2} Andrea Palermo ^{1,2}, Alessio Gerussi ^{1,2}, Daphne D'Amato,^{1,2} Sarah Elisabeth O'Donnell,^{1,2} Federica Cerini,⁶ Carla De Benedittis,⁷ Federica Malinverno,^{1,2} Vincenzo Ronca,^{1,2} Clara Mancuso,^{1,2} Nora Cazzagon,¹⁴ Antonio Ciaccio,^{1,2} Donatella Barisani,¹ Marco Marzioni,¹⁵ Annarosa Floreani,^{16,17} Domenico Alvaro,⁹ Eugenio Gaudio,⁵ Pietro Invernizzi,^{1,2} Guido Carpino ^{18**}, Alessandra Nardi,^{19**} and Marco Carbone^{1,2**}, on the behalf of the Italian PBC Registry

BACKGROUND AND AIMS: Liver fibrosis holds a relevant prognostic meaning in primary biliary cholangitis (PBC). Noninvasive fibrosis evaluation using vibration-controlled transient elastography (VCTE) is routinely performed. However, there is limited evidence on its accuracy at diagnosis in PBC. We aimed to estimate the diagnostic accuracy of VCTE in assessing advanced fibrosis (AF) at disease presentation in PBC.

APPROACH AND RESULTS: We collected data from 167 consecutive treatment-naïve PBC patients who underwent liver biopsy (LB) at diagnosis at six Italian centers. VCTE examinations were completed within 12 weeks of LB. Biopsies were scored by two blinded expert pathologists, according to the Ludwig system. Diagnostic accuracy was estimated using the area under the receiver operating characteristic curves (AUROCs) for AF (Ludwig stage \geq III). Effects of biochemical and clinical parameters on liver stiffness measurement (LSM) were appraised. The derivation cohort consisted of 126 patients with valid LSM and LB; VCTE identified patients with AF with an AUROC of 0.89. LSM cutoffs \leq 6.5 and $>$ 11.0 kPa enabled to exclude and confirm, respectively,

AF (negative predictive value [NPV] = 0.94; positive predictive value [PPV] = 0.89; error rate = 5.6%). These values were externally validated in an independent cohort of 91 PBC patients (NPV = 0.93; PPV = 0.89; error rate = 8.6%). Multivariable analysis found that the only parameter affecting LSM was fibrosis stage. No association was found with BMI and liver biochemistry.

CONCLUSIONS: In a multicenter study of treatment-naïve PBC patients, we identified two cutoffs (LSM \leq 6.5 and $>$ 11.0 kPa) able to discriminate at diagnosis the absence or presence, respectively, of AF in PBC patients, with external validation. In patients with LSM between these two cutoffs, VCTE is not reliable and liver biopsy should be evaluated for accurate disease staging. BMI and liver biochemistry did not affect LSMs. (HEPATOLOGY 2021;74:1496-1508).

P primary biliary cholangitis (PBC) is an autoimmune liver disease characterized by destructive cholangitis affecting the small intrahepatic

Abbreviations: AF, advanced fibrosis; AIH, autoimmune hepatitis; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AMA, antimitochondrial antibody; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; AUROCs, area under the receiver operating characteristic curves; BMI, body mass index; CPGs, Clinical Practice Guidelines; EASL, European Association for the Study of the Liver; FiB-4, Fibrosis-4 score; IQR/M, interquartile range/median; kPa, kilopascal; LB, liver biopsy; LF, liver fibrosis; LFTs, liver function tests; LR⁺, positive likelihood ratio; LR⁻, negative likelihood ratio; LSM, liver stiffness measurement; NPV, negative predictive value; PBC, primary biliary cholangitis; PPV, positive predictive value; ROC, receiving operator characteristic; UDCA, ursodeoxycholic acid; URS, UDCA response score; VCTE, vibration-controlled transient elastography; ULN, upper limit of normal.

Received November 3, 2020; accepted March 2, 2021.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31810/supinfo.

*These authors contributed equally and are joint first authors.

**These authors contributed equally and are joint last authors.

bile ducts, leading to chronic cholestasis and fibrosis. Many patients eventually develop end-stage liver disease with attendant need for liver transplantation.⁽¹⁾ PBC pathophysiology is characterized by three main processes: inflammation, cholestasis, and fibrosis. Surrogate markers of inflammation and cholestasis (i.e., alkaline phosphatase [ALP], bilirubin, and transaminases) are used to assess the response to therapy with ursodeoxycholic acid (UDCA); as such, they have been included in binary criteria and continuous prognostic scores to quantify treatment benefit (e.g., Paris criteria, GLOBE, and UK PBC score) after UDCA, which represent the basis for risk stratification in PBC.⁽²⁻⁴⁾ In addition, our group has recently proposed the UDCA response score (URS), which, including

biochemical and clinical parameters, enables an accurate prediction of UDCA response at diagnosis.⁽⁵⁾

Fibrosis staging at baseline is currently overlooked in PBC and not integrated into a paradigm of management such as biochemical response. Recently, the GLOBAL PBC and the UK-PBC study groups showed that histological fibrosis grants prognostic value beyond biochemical response at 1 year; this highlighted the need to incorporate liver fibrosis (LF) stage (or its surrogate markers) into paradigms of risk stratification of PBC at diagnosis.^(6,7)

Liver biopsy (LB) is currently the gold standard for LF staging in chronic liver disease. However, because of its invasiveness and potential sampling error, it is not recommended for staging purposes at diagnosis

This research was partially supported by the Italian Ministry of University and Research (MIUR)–Department of Excellence project PREMIA (PREcision Medicine Approach: bringing biomarker research to clinic). The study was partially supported by the Grants titled “Biocompatible Nano-assemblies to Increase the Safety and the Efficacy of Steroid Treatment Against Liver Inflammation” (grant/award no.: GR-2018-12367794) and “the Role of Auto-reactive Hepatic Natural Killer Cells in the Pathogenesis of Primary Biliary Cholangitis” (grant/award no.: PE-2016-02363915).

© 2021 The Authors. HEPATOLOGY published by Wiley Periodicals LLC on behalf of American Association for the Study of Liver Diseases. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.31810

Potential conflict of interest: Dr. Calvaruso received grants from Intercept, AbbVie, and Gilead. Dr. Floreani consults for Intercept. Dr. Invernizzi received grants from Intercept and Gilead. Dr. Carbone consults for and advises for Intercept.

ARTICLE INFORMATION:

From the ¹Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy; ²European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy; ³Bicocca Bioinformatics Biostatistics and Bioimaging Centre-B4, School of Medicine and Surgery, University of Milan-Bicocca, Monza, Italy; ⁴Section of Gastroenterology and Hepatology, PROMISE, University of Palermo, Palermo, Italy; ⁵Department of Anatomical, Histological, Forensic Medicine and Orthopedics Sciences, Sapienza University of Rome, Rome, Italy; ⁶Division of Hepatology, Ospedale San Giuseppe, University of Milan, Milan, Italy; ⁷Department of Translational Medicine, Università degli Studi del Piemonte Orientale “A. Avogadro”, Novara, Italy; ⁸CRC “A. M. e A. Migliavacca” Center for Liver Diseases, Division of Gastroenterology and Hepatology, Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy; ⁹Department of Medico-Surgical Sciences and Biotechnologies, Polo Pontino “Sapienza” University of Rome, Latina, Italy; ¹⁰Department of Internal Medicine, University of Genoa, Genoa, Italy; ¹¹Pathology Department, San Gerardo Hospital, Monza, Italy; ¹²Histopathology Unit, AOU Maggiore della Carità, Novara, Italy; ¹³Department of Precision and Translational Medicine, Sapienza University of Rome, Rome, Italy; ¹⁴Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy; ¹⁵Department of Gastroenterology, Università Politecnica delle Marche, Ancona, Italy; ¹⁶Studiosa Senior, University of Padua, Padua, Italy; ¹⁷Scientific Consultant, IRCCS Negrar, Verona, Italy; ¹⁸Department of Movement, Human and Health Science, University of Rome “Foro Italico”, Rome, Italy; ¹⁹Department of Mathematics, University of Rome Tor Vergata, Rome, Italy.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Marco Carbone, M.D., Ph.D.
Division of Gastroenterology
Center for Autoimmune Liver Diseases
Department of Medicine and Surgery
University of Milano-Bicocca

Via Cadore, 48
Monza 20900, Italy
E-mail: marco.carbone@unimib.it
Tel.: +393927333154

by international guidelines on PBC.⁽⁸⁻¹¹⁾ Noninvasive evaluation of LF with liver stiffness measurements (LSMs) by vibration-controlled transient elastography (VCTE) has been proved as a simple and reliable surrogate marker of fibrosis in several chronic liver diseases.^(12,13) In PBC, LSM by VCTE is currently recommended by European Association for the Study of the Liver (EASL) Clinical Practice Guidelines (CPGs) for disease staging at diagnosis and during follow-up.⁽⁹⁾ However, such a recommendation is based on a cross-sectional, single-center study including patients on treatment, at different times from diagnosis, and patients with overlap PBC/autoimmune hepatitis (AIH) under immunosuppressive therapy.⁽¹⁴⁾ Concomitant therapies and liver inflammation of a different disease phenotype (i.e., AIH) might have introduced an uncontrolled bias in this heterogeneous cohort. Thus, there is a need to identify and validate accurate cutoffs of LSM for LF assessment in a treatment-naïve population, at disease presentation, to implement early risk stratification in PBC.

We designed a study across six Italian liver centers with the aim to explore the diagnostic accuracy of VCTE compared with a reference standard based on histological evaluation of fibrosis, and to study the impact of body mass index (BMI), inflammation, and cholestasis on LSM readings. In addition, we aimed to identify LSM cutoffs for use in clinical practice.

Patients and Methods

STUDY DESIGN AND PARTICIPANTS

This is a diagnostic test accuracy study using data from the Italian PBC Registry, an ongoing, noninterventive, multicenter, prospective, observational cohort study that monitors patients with PBC across the country. From January 2006 to August 2019, all patients with a new diagnosis of PBC and naïve to specific therapy who underwent both VCTE and percutaneous LB within 12 weeks from each other and within 6 months from diagnosis were consecutively included in the study from six sites participating in the Italian PBC Registry (Ospedale San Gerardo, Monza; Ospedale San Giuseppe, Milan; Ospedale Maggiore della Carità, Novara; Fondazione IRCCS “Ca’ Granda” Ospedale Maggiore Policlinico, Milan;

Ospedale San Martino, Genova; and Policlinico Umberto I, Rome).

To confirm our results, we tested the diagnostic performance of cutoffs derived from the derivation cohort in an external validation cohort of 91 patients with the same inclusion and exclusion criteria, from an external site, the Policlinico Paolo Giaccone, Palermo.

The study was conducted in accordance with the guidelines of the Declaration of Helsinki and the principles of good clinical practice. All participants provided written informed consent. The study was approved by the University of Milan-Bicocca Research Ethics Committee (study name: PBC322), coordinator of the Italian National Registry, and by the research and development department of each collaborating hospital.

All authors had access to the study data and reviewed and approved the final manuscript. The STARD guidelines were followed to report the methods and results of this study.⁽¹⁵⁾

ENDPOINTS

The aim of the study was to assess the diagnostic accuracy of LSM by VCTE against liver histology at the time of diagnosis in PBC patients and evaluate the potential confounding effect of BMI, liver inflammation, and cholestasis in predicting fibrosis by LSM in patients naïve to therapy.

STUDY DEFINITIONS

Diagnosis of PBC was made based on elevated ALP and the presence of antimitochondrial antibodies (AMAs) at a titer >1:40 or specific antinuclear antibodies (ANAs) immunofluorescence (nuclear dots or perinuclear rims) or ELISA results (sp100, gp210) in AMA-negative patients. In patients negative to PBC-specific antibodies, diagnosis was made by histological evidence of inflammatory destructive cholangitis of bile ducts.

INCLUSION AND EXCLUSION CRITERIA

Patients aged ≥ 18 years, able to give written informed consent, were scheduled to have an LB in the context of the Italian PBC Registry for investigation of suspected PBC, search for overlap syndrome

with AIH and/or other chronic liver conditions (e.g., NASH), or disease staging within 3 months of VCTE examination.

We excluded from the study patients with other concomitant liver-related disease such as HBV or HCV, histological overlap syndrome with AIH and NASH, a history or an active alcohol abuse, and any other causes of liver injuries other than PBC.

DATA CAPTURE

Data on clinical, biochemical, histological features, and LSM values were collected prospectively into a bespoke database. The database is an electronic data capture system with an electronic case report form developed for the purpose of the Registry.

Baseline data were collected at diagnosis, before starting the UDCA. The following parameters were collected in the derivation cohort: age, sex, BMI, date of liver biopsy, date of diagnosis, LSM value, and liver function tests (LFTs) performed within 3 months from biopsy date and VCTE, that is, serum albumin, ALP, gamma-glutamyl-transferase, total bilirubin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and platelet count.

The following data were collected at baseline in the validation cohort: age, sex, BMI, date of liver biopsy, date of diagnosis, LSM, and LFTs performed within 3 months from biopsy date and VCTE, (i.e., ALP, bilirubin, ALT, AST, and platelets). Age, LFTs, and platelets were used to calculate Fibrosis-4 score (FIB-4) and AST to platelet ratio index (APRI) at diagnosis^(16,17) in both cohorts.

HISTOPATHOLOGICAL EVALUATION

Percutaneous LBs were performed according to the local standard procedure with a 16G needle in the right hepatic lobe. Only liver specimens with at least 10 complete portal tracts were considered eligible in this study. For all liver specimens, the following stainings were used for review: HE, Masson trichrome stain, and cytokeratin-7 stain. Slides were analyzed independently by two experienced pathologists (G.C. and N.Z.) who were blinded to each other's reading and to the patient's clinical data and LSM. In case of disagreement, consensus was obtained by joint review of sections until agreement. Fibrosis was staged according to Ludwig staging system on a 1-4 scale:

I = portal hepatitis with little or no periportal inflammation or piecemeal necrosis; II = periportal hepatitis with piecemeal necrosis; absence of bridging necrosis and of septal fibrosis; III = septal fibrosis or bridging necrosis, or both are present; and IV = cirrhotic stage: fibrous septa and regenerative nodules.⁽¹⁸⁾

For the purposes of this study, patients were categorized as having early stage (Ludwig = I or II) and advanced stage (Ludwig = III or IV).

VCTE (FIBROSCAN)

VCTE examination by FibroScan (Echosens, Paris, France) was performed in each center by physicians trained and certified by the manufacturer. An automatic probe selection tool was embedded in the device software that recommends the appropriate probe for each patient according to the real-time assessment of the skin-to-liver capsule distance. All LSMs included in the study has been performed after at least 4 hours of fast by scanning the right liver lobe through an intercostal space by an experienced operator with an experience of at least 500 examinations. LSMs are expressed in kilopascals (kPa), and the Boursier criteria performance as a quality control for FibroScan were evaluated in this cohort that was classified in three reliability categories: very reliable (interquartile range/median [IQR/M], ≤ 0.1); reliable (IQR/M > 0.1 and ≤ 0.3 or IQR/M > 0.3 with LSM < 7.1 kPa); and poorly reliable (IQR/M > 0.3 in patients with LSM > 7.1 kPa).⁽¹⁹⁾ Only examinations with at least 10 valid individual measurements and classified as reliable and very reliable according to the Boursier criteria were deemed valid.

STATISTICAL ANALYSIS

To account for interlaboratory variability, ALP, ALT, AST, and total bilirubin are expressed as a multiple of their respective upper limit of normal values (ULNs). Given that most variables showed skewed distributions with significant departures from the normal density, a nonparametric approach was preferred in the analysis. Continuous variables were summarized by median, first, and third quartiles. Histograms were used to describe distributions, and kernel-density estimates were overlaid. The Wilcoxon rank-sum test was used to compare groups. Categorical variables were described by absolute frequencies and percentages; to compare groups, we used the χ^2 test (or Fisher's exact test in the case of sparse tables).

Box-and-whisker plots were created for graphical comparison of empirical distributions.

Multivariable analysis was undertaken using logistic regression. For ALP, ALT, and bilirubin, the logarithmic transformation was considered to adjust for the extreme skewness of their distributions. Maximum likelihood estimates are reported. The Wald test was used to assess significance. Poorly predicted observations were identified by the standardized deviance residuals. Diagnostic accuracy was evaluated using receiving operator characteristic (ROC) curves. Nonparametric stratified bootstrapping was used to compute confidence bands for ROC curves. Area under the ROC curve (AUROC) is reported together with its 95% CI. ROC curves were compared using De Long's test.

Negative and positive predictive values (NPV, PPV), specificity and sensitivity, and positive and negative likelihood ratios (LR^+ , LR^-) are reported. In the case of a single cutoff, the optimal threshold was chosen maximizing the Youden index. In the dual

approach, criteria for choosing cut-off values were as follows: for the high confirmatory cutoff, specificity and PPV >0.90 , and for the low exclusionary cutoff, sensitivity and NPV >0.90 . If more than two cutoffs met these criteria, the additional requirement was to minimize the grey area.

Analyses were conducted using SAS (version 9.4; SAS Institute Inc., Cary, NC) and R software (version 3.4; R Foundation for Statistical Computing, Vienna, Austria).

Results

PATIENTS' CHARACTERISTICS

The study flow chart is represented in Fig. 1. One hundred sixty-seven patients with both VCTE and LB performed at diagnosis within 12 weeks from each other were consecutively enrolled in the study period.

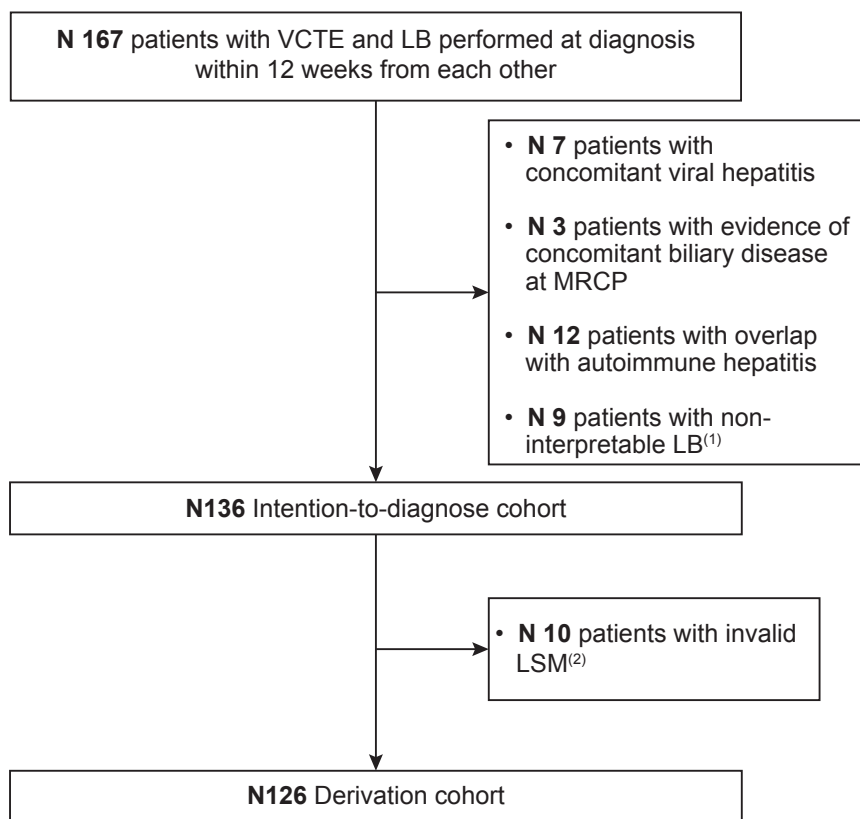


FIG. 1. Flow chart of the study. Notes: ⁽¹⁾We considered interpretable LB specimens those with at least 10 evaluable portal spaces. ⁽²⁾We considered valid LSM when 10 valid measurements were collected and classified as “very reliable” and “reliable” according to Boursier’s criteria. ⁽¹⁹⁾ Abbreviation: MRCP, magnetic resonance cholangiopancreatography.

Among them, 22 (13.2%) patients were excluded from the study for concomitant liver disease overlap. Biopsy failure or inadequate liver specimen was recorded in 9 patients (5.4%). Our intention-to-diagnose cohort consisted of 136 patients with LB and VCTE at diagnosis. Median time between VCTE and LB was 14.5 days (IQR, 0.0, 39.3). Consensus on the grade and stage of the biopsy sample was reached in all cases. Fifty-five patients (40.4%) had histological stage Ludwig I, 39 had Ludwig II (28.7%), 30 had Ludwig III (22.1%), and 13 had Ludwig IV (8.8%). Median age was 52 (IQR, 46, 58); 90.4% were females. Median ALP, ALT, and total bilirubin at baseline were $1.4 \times$ ULN (IQR, 1.0, 2.4), $1.3 \times$ ULN (IQR, 0.9, 2.0), and $0.6 \times$ ULN (IQR, 0.4, 0.8), respectively.

Ten patients had unreliable LSM (as defined below). The 126 remaining patients composed the derivation cohort leading to an applicability of 92.1%. Among these patients, 53 (42%) were classified as “very reliable” and 73 (58%) as “reliable” according to the Boursier criteria (Fig. 1).

DERIVATION COHORT

Prediction of Advanced Fibrosis

With the aim of predicting advanced fibrosis (AF), a per-protocol analysis was undertaken using data from the derivation cohort where valid measures of LSM were available. Advanced stage was defined as Ludwig stage III-IV, whereas early stage was defined

as Ludwig stage I-II. Ninety-one patients (72.2%) were in early stage, 35 (27.8%) in advanced stage.

Median biochemical values, LSM, APRI, and FIB-4, of patients in early and advanced stage are reported in Table 1. LSMs according to fibrosis stage by Ludwig are presented in Fig. 2.

In order to predict advanced stage, the logistic model was fitted to observed data considering LSM, BMI, ALP, ALT, bilirubin, albumin, and platelet count at diagnosis as potential predictive factors; in addition, the analysis was adjusted for age and sex. LSM was the only significant predictor of AF. None of the biochemical parameters nor BMI showed a significant, additive predictive contribution to LSM (Table 2). Predictive value of LSM in identifying AF, as measured from the AUROC, was 0.89 (CI, 0.83, 0.95; Supporting Fig. S1).

Figure 3 shows the relationship between predicted probabilities of AF and LSM values according to the fitted logistic model. The curve steeply increased in an interval of LSM between 7 and 11 kPa, which corresponds to an estimated probability of AF ranging from 0.21 to 0.75 (Fig. 3). In this range, patients with Ludwig I-II and patients with Ludwig III-IV have overlapping LSM. Thus, despite the good predictive capability of LSM, in this interval a reliable prediction of AF using a single-threshold approach appears unfeasible.

Indeed, the point where Youden's index is maximum was 7.0, with a sensitivity of 0.89 and a specificity of 0.79. However, whereas NPV was 0.95, PPV was only 0.62, with 19 patients falsely classified in advanced stage.

TABLE 1. Demographics and Clinical Characteristics at Diagnosis of the Derivation Cohort According to Ludwig Stage

	Early Stage (n = 91)		Advanced Stage (n = 35)	
	Median or N	Q1-Q3 or %	Median or N	Q1-Q3 or %
Age at diagnosis (years)	51	45-55	54	48-62
Female sex	84	92.3	29	82.8
ALP \times ULN	1.3	0.8-2.1	1.8	1.3-3.5
ALT \times ULN	1.1	0.8-2.0	1.9	1.4-2.3
AST \times ULN	1.0	0.9-1.5	1.4	1.0-2.0
Total bilirubin \times ULN	0.6	0.4-0.7	0.8	0.5-1.0
Albumin (g/dL)	4.3	4.0-4.4	4.2	4.0-4.4
PLT $\times 10^9/L$	253	213-304	203	160-245
LSM (kPa)	5.5	4.5-6.8	9.6	7.7-14.5
APRI	0.43	0.31-0.65	0.66	0.41-1.22
FIB-4	1.33	0.92-1.60	1.66	1.30-2.91
BMI	24	21-26	25	22-28

Abbreviation: PLT, platelet count.

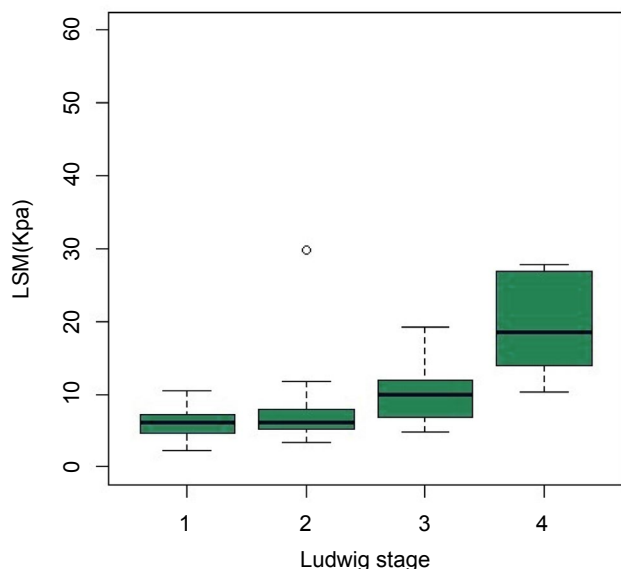


FIG. 2. Distribution of LSM according to histological stage by Ludwig in the derivation cohort. LSM increased significantly in fibrotic stages III and IV by the Ludwig system (Kruskal-Wallis test, $P < 0.00001$).

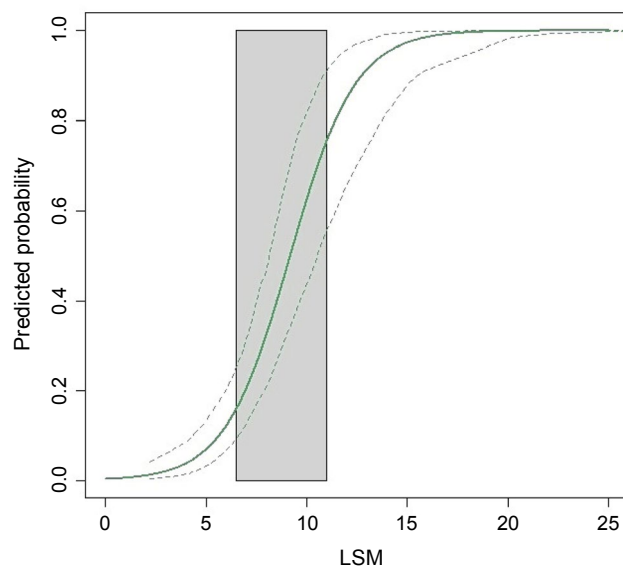


FIG. 3. Logistic curve of the relationship between predicted probabilities of AF and LSM. The grey area highlights the portion of the curve in which VCTE may not be reliable in predicting AF.

TABLE 2. Multivariable Logistic Model Fitted to Observed Data

	OR	95% CI	PValue
LSM (kPa)	1.76	(1.29, 2.41)	0.0004
Age (years)	1.04	(0.97, 1.11)	0.3153
Sex (female vs. male)	0.74	(0.10, 5.31)	0.7661
ALP \times ULN (log scale)	1.13	(0.45, 2.84)	0.7955
ALT \times ULN (log scale)	0.96	(0.35, 2.65)	0.9310
Total bilirubin \times ULN (log scale)	1.46	(0.44, 4.80)	0.5384
Albumin (g/dL)	1.10	(0.20, 6.00)	0.9117
PLT $\times 10^9/L$	1.00	(0.99, 1.01)	0.2954
BMI	1.00	(0.85, 1.19)	0.9854

Abbreviation: PLT, platelet count.

Thus, we explored the use of a dual cut-off approach with a lower and a higher threshold to define areas of accurate prediction and a grey area where VCTE may not provide a reliable prediction of AF (Fig. 4). The diagnostic accuracy of different possible high and low cut-off values is reported in Supporting Table S1. For the optimal lower threshold of 6.5 kPa, which defines the absence of AF, sensitivity and NPV were 0.91 and 0.96, respectively.

This threshold led to identifying 70 early-stage patients, of whom 67 (95.7%) were correctly predicted (Table 3). The 3 patients not correctly predicted

had Ludwig stage III, had an LSM of 4.3, 5.9, and 6.1 kPa, and none showed biochemical features of AF or cirrhosis. ALP was <1.5 ULN in all of them, and transaminases were <1.5 ULN. Only 1 patient had bilirubin levels of $1.1 \times$ ULN, with normal serum albumin and platelet count.

For the optimal higher threshold of 11.0 kPa, which defines the presence of AF, specificity and PPV were 0.99 and 0.94, respectively. This threshold identified 17 advanced-stage patients, of whom 16 (94.1%) were correctly predicted. The only patient not correctly predicted (LSM = 11.3 kPa, Ludwig I) had low platelet count ($120 \times 10^9/L$), ALP markedly increased ($\times 5.6$ ULN) along with a mild increase of transaminase. In this patient, APRI score was 1.72 and FIB-4 was 4.13 and were both consistent with AF.

Using this dual cut-off approach in the derivation cohort, the positive and negative likelihood ratio were 91.0 and 0.09, respectively, and the total error rate was 5.6%.

Comparison VCTE With FIB-4 and APRI Score

Performance of LSM was compared with FIB-4 and APRI score. FIB-4 and APRI score at diagnosis

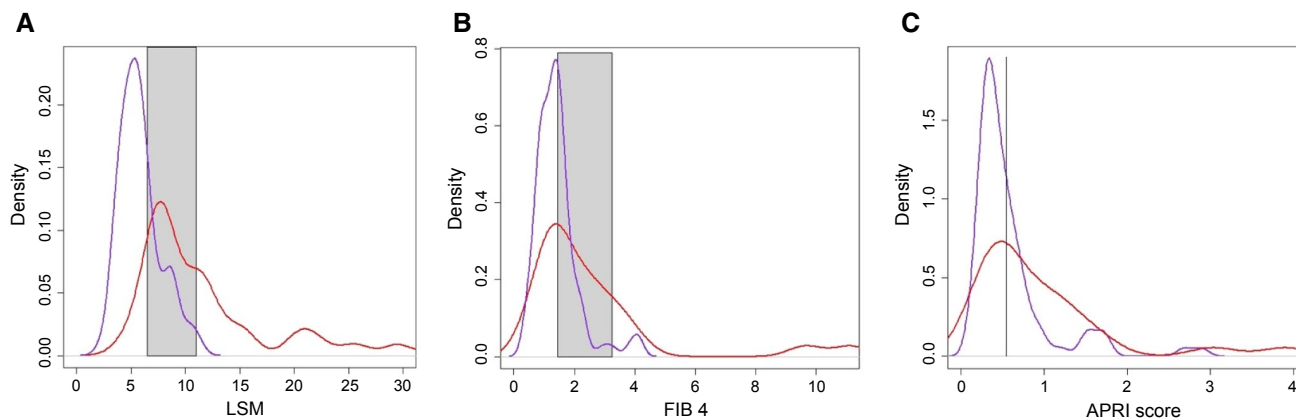


FIG. 4. Density plot of LSM (A), FIB-4 (B), and APRI score (C) in the derivation cohort. Patients with Ludwig stage I and II at liver biopsy are represented in purple lines, those with Ludwig stage III and IV in red lines. In the LSM density plot (A), the grey area highlights the interval of LSM in which TE is not reliable. In the APRI and FIB-4 density plots (B,C), the peak of density of patients in early and advanced stage are almost overlapped, which underlies the limits of these tools in PBC. Note: The grey area in the FIB-4 density plot (B) expresses the range of LSM in which FIB-4 as proposed by Sterling et al.⁽¹⁶⁾ The black straight line in the APRI score density plot (C) expresses the cutoff of 0.54 validated in PBC.⁽²⁰⁾ Extreme observations were excluded (4 cases).

TABLE 3. Ludwig Stage Stratified by Risk Class Prediction of Fibrosis in the Logistic Regression Model in the Intention-to-Diagnose Cohort.

	Early Stage (LSM ≤ 6.5 kPa) n (Column %)	Grey Area ($6.5 < \text{LSM} \leq 11.0$ kPa) n (Column %)	Advanced Stage (LSM > 11.0 kPa) n (Column %)	Unreliable LSM n (Column %)	Total n (Column %)
Ludwig stage I	40 (57.1)	13 (33.3)	1 (5.9)	1 (10.0)	55 (40.4)
Ludwig stage II	27 (38.6)	10 (25.6)	0 (0)	2 (20.0)	39 (28.7)
Ludwig stage III	3 (4.3)	15 (38.5)	6 (35.3)	6 (60.0)	30 (22.1)
Ludwig stage IV	0 (0)	1 (2.6)	10 (58.8)	1 (10.0)	12 (8.8)
Total	70	39	17	10	136

were available for 114 (90.5%) patients. AUROC for AF was 0.66 (CI, 0.54, 0.77) for FIB-4 and 0.64 (CI, 0.52, 0.76) for APRI. LSM outperformed both the alternative tests ($P = 0.0066$ and $P = 0.0037$, respectively; Supporting Fig. S2).

To illustrate the different discrimination power of the three tests, we estimated their empirical distributions stratified by presence or absence of histological AF. FIB-4, and APRI empirical densities overlap, with the distributions of cases with AF only slightly shifted to the right. A moderate overlap is shown also in LSM, which is, however, confined to the grey area (Fig. 4).

Using the dual cut-off approach validated and currently in use for FIB-4,⁽¹⁶⁾ among 60 patients with an FIB-4 < 1.45 , 51 were correctly predicted (sensitivity = 0.65; NPV = 0.85); among 10 patients with FIB-4 > 3.25 , 6 patients were correctly predicted (specificity = 0.96; PPV = 0.60). Among the 44

patients with FIB-4 measurements in the grey area (i.e., ≥ 1.45 and ≤ 3.25), 33 patients were in early stage and 11 were in advanced stage (Supporting Table S2).

When the single, validated cutoff of 0.54 was applied for APRI, we found specificity 0.65, sensitivity 0.58, NPV 0.84, and PPV 0.33, with 29 patients falsely classified in the advanced fibrotic stage.⁽²⁰⁾

Intention-to-Diagnose Analysis

Performance of the VCTE for discrimination of AF was tested also including patients with unreliable LSM results ($n = 10$). Among these patients, 4 had < 10 valid measurements (3 with eight measurements and 1 with nine measurements) and 6 were unreliable according to the Boursier criteria (median IQR/M, 0.33 [IQR, 0.32, 0.34]). Patients with invalid LSMs

had a significantly higher BMI than patients in the derivation cohort (27.3 [IQR, 24.3, 29.9] vs. 24.0 [IQR, 21.0, 27.0]; $P = 0.0475$). No other significant differences between these 10 patients and the derivation cohort in demographical and biochemical variables were found.

Patients with unreliable LSM were classified using two opposite extreme scenarios, either as were “all wrongly classified” (worst scenario) or as were “all correctly classified” (best scenario). In the worst and in the best scenario, we found a sensitivity of 0.76 and 0.93, a specificity of 0.96 and 0.99, a PPV of 0.80 and 0.96, an NPV of 0.87 and 0.96, an LR^+ of 19 and 93 and an LR^- of 0.25 and 0.07, respectively.

Considering the results of VCTE achieved in all 10 patients, 5 were correctly classified, 4 patients had LSM within the grey area, and 1 patient was wrongly classified in the early stage with Ludwig stage III at LB.

VALIDATION COHORT

The two cutoffs identified in the derivation cohort were validated in an external cohort of 91 PBC patients at disease onset, naïve to UDCA with a time span from LB and VCTE of 21.5 (IQR, 7.3, 52.3) days. There were no clinically meaningful differences between the two cohorts (Table 1 and Supporting Table S3). LSM versus fibrosis stage by Ludwig is presented as a box plot in Supporting Fig. S3.

The lower threshold of 6.5 kPa identified 40 early-stage patients, of whom 37 (92.5%) were correctly predicted (Supporting Table S4). The 3 patients not correctly predicted were all Ludwig stage III, LSM of 4.9, 5.6, and 6 kPa, and normal albumin, bilirubin, and platelet count; 2 patients had ALP and transaminases within $2 \times$ ULN, and 1 patient had ALP $8 \times$ ULN and ALT $4 \times$ ULN. Sensitivity and NPV were 0.89 and 0.93, respectively.

The higher threshold of 11.0 kPa identified 18 advanced-stage patients, of whom 16 (98.4%) were correctly predicted (Supporting Table S4).

Considering wrongly predicted cases, 1 patient had an LSM of 29.8 kPa and Ludwig stage II, with normal serum albumin, bilirubin level, and platelet count, and ALP $4 \times$ ULN and ALT $4 \times$ ULN; the second patient had an LSM 11.8 kPa and Ludwig stage II, with bilirubin $1.2 \times$ ULN, ALP $2.5 \times$ ULN and ALT $2.5 \times$ ULN, and normal albumin and platelet count.

Specificity and PPV were 0.97 and 0.89, respectively. Using the dual cutoff in the validation cohort, the LR^+ and LR^- were 29.67 and 0.11, respectively, and total error rate was 8.6%.

In the validation cohort, APRI and FIB-4 were available for 88 (96.7%) patients with median values of 0.45 (IQR, 0.30, 0.72) and 1.37 (IQR, 1.05, 1.97), respectively. To show the different discrimination power of the three tests in the validation cohort, we estimated the empirical distributions of LSM, FIB-4, and APRI stratified by presence or absence of histological AF (Supporting Fig. S4). Similarly to what we showed in the derivation cohort, the density peak of both groups was mostly overlapping, whereas in the LSM plot the overlap was confined to the grey area.

GREY AREA (OVERALL COHORT)

The overall number of patients in the grey area was 72 of 217 (33.2%): 68.1% were in early stage, 31.9% in advanced stage. Median values of LSM were 7.9 (IQR, 7.1, 8.7) and 7.9 kPa (IQR, 7.4, 9.0) for patients with Ludwig stage I-II and Ludwig stage III-IV, respectively. BMI, ALP, bilirubin, platelet count, and noninvasive scores of fibrosis (i.e., APRI and Fib-4) were not significantly different in patients in early and advanced stage (Supporting Table S5).

Discussion

Fibrosis is a major driver of clinical outcomes in PBC. The accurate staging of patients at disease presentation, ideally using noninvasive tests, is a major unmet clinical need in PBC. In this study, we confirmed the high performance of VCTE in predicting AF in a nation-wide cohort of treatment-naïve PBC patients and provided externally validated cut-off values for confirming or excluding fibrosis at diagnosis.

This study provides a pragmatic approach to threshold setting of noninvasive tests in PBC by creating three classes of risk: early stage, advanced stage, and a grey area of inaccurate discrimination. Indeed, the large number of falsely classified patients with a single cut-off approach, despite its good sensitivity and specificity, highlights the limits of the discriminating power of VCTE in a range of defined values of

stiffness. The proposed methodology showed a good predictive capability in per-protocol analysis that was confirmed also using an intention-to-diagnose approach.

Patients without relevant fibrosis at VCTE are more likely to respond to UDCA, have a lower risk of end-stage liver disease complication, and can therefore be de-escalated in the intensity of care.

On the other hand, the early identification of clinically relevant fibrosis at baseline would enhance patient management timeliness; this should be done through HCC surveillance and early (second-line) treatment escalation, particularly in those predicted at high risk of first-line treatment failure by the URS.⁽⁵⁾ Indeed, the URS and VCTE can be combined as noninvasive tools to implement baseline risk stratification in PBC by offering an estimated risk of treatment failure and disease stage, respectively (Fig. 5). Finally, an accurate, noninvasive staging at diagnosis would support patient selection in clinical trial design in PBC.

UDCA response and fibrosis stage at diagnosis are key parameters for risk stratification in PBC. Recently, two studies of the Globe PBC⁽⁶⁾ and UK-PBC⁽⁷⁾ study groups independently showed that the assessment of fibrosis stage at diagnosis grants prognostic value beyond biochemical treatment response. This highlights the need to incorporate fibrosis stage,

or a reliable noninvasive surrogate, in individual risk stratification in patients with PBC.

Currently, liver biopsy has a marginal role for diagnosis, and it is not recommended for disease staging at diagnosis. VCTE by FibroScan is considered the best surrogate markers for the detection of severe fibrosis or cirrhosis in patients with PBC. There is a critical need in clinical practice and clinical research to define an accurate cutoff of LSM. A seminal French study in PBC by Corpechot et al. (n = 150) demonstrated the high specificity and sensitivity (>90%) of VCTE in distinguishing the fibrotic stages.⁽¹⁴⁾ However, the prediction of intermediate fibrosis was dismal (LSM = 8.8 kPa for fibrosis F2; sensitivity, 0.67; specificity, 1.0). Based only on this study, the EASL CPGs recommend the use of VCTE for disease staging at baseline and during follow-up.⁽⁹⁾ However, this study, though relevant, had some methodological flaws: The cohort was cross-sectional with patients at different phases of the disease course (mean time from diagnosis = 6.7 years), and only 11% of patients were assessed at diagnosis and naïve to therapy. Moreover, 14% of patients had histologically proven PBC-AIH overlap syndrome, and 18% of patients were receiving additional corticosteroids and/or mycophenolate mofetil; more important, this was a single-center study lacking an external validation cohort.⁽¹⁴⁾

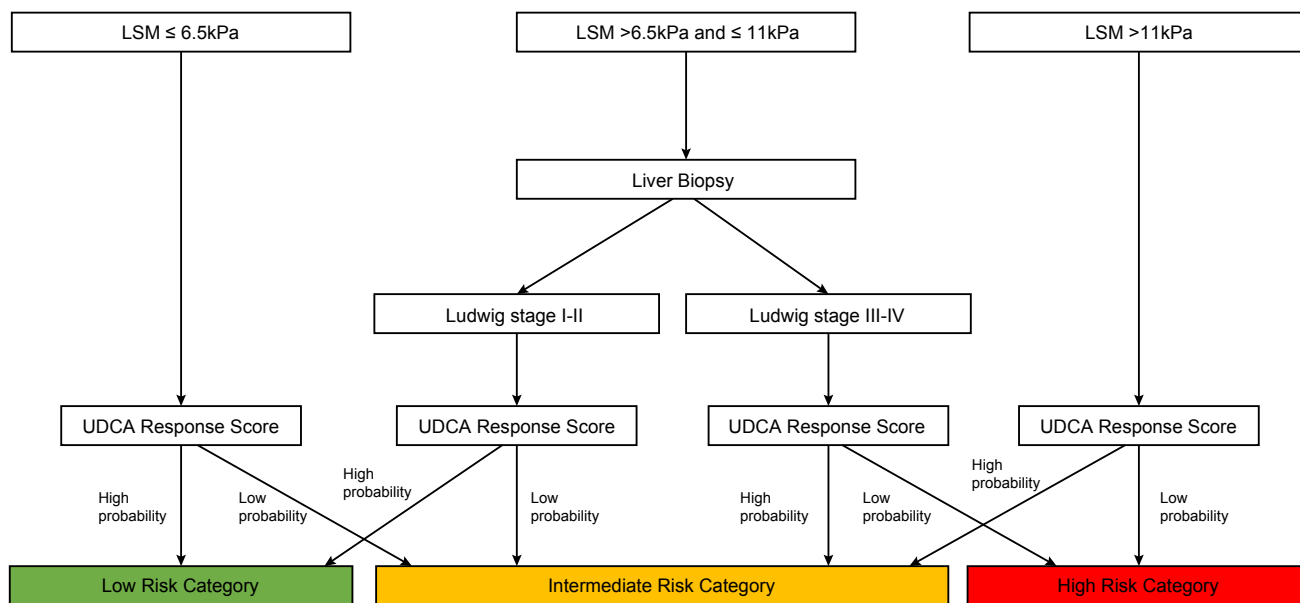


FIG. 5. Proposed algorithm for risk stratification at diagnosis in PBC patients.

The diagnostic performance of VCTE and cut-offs for staging fibrosis in our study are broadly in keeping with data from the French study and others, which showed a mean LSM value for fibrosis F3-F4 by Metavir (comparable to Ludwig stage III-IV) of 10.9 (CI, 10.7,11.5).^(14,21) However, the exclusion of fibrosis could not be compared given that in previous studies, the evaluation of LSM in nonfibrotic patients was overlooked.^(14,21,22)

The dual cut-off approach is not unfamiliar. This has been proposed in hepatitis B patients by Viganò et al. and recently in HBV-HIV-coinfected patients by Sterling et al.^(23,24) It highlights a grey area of inaccurate prediction, which is inherent to the device, known to outperform with the extreme readings and fail with intermediate ones.^(23,25)

With the same intent, we used the Ludwig system for disease staging, rather than Metavir or Ishak systems, to identify clinically relevant fibrosis (Ludwig stage \geq III), rather than intermediate fibrotic stages. Unsurprisingly, APRI and FIB-4 do not provide help in fibrosis discrimination even in this subset. For this reason, in patients with intermediate LSM readings, liver biopsy can be a justified approach for accurate disease staging to guide further management.

It is reported that liver inflammation and cholestasis may influence VCTE accuracy for the noninvasive evaluation of LF.⁽²⁶⁻³¹⁾ Hepatic inflammation in particular has been identified as a potential confounder that may lead to false-positive LSMs even when transaminases are not markedly elevated.^(32,33) In our study, surrogate markers of hepatic inflammation and cholestasis (i.e., transaminases, ALP, and bilirubin) were explored in their relationship with LSM, and no significant influence of their effect was observed at diagnosis. This is consistent with the recent study on VCTE in AIH. Hartl et al. showed that the diagnostic accuracy of VCTE for staging fibrosis was not different in patients after immunosuppressant treatment achieving biochemical remission compared to those without biochemical remission.^(34,35) Likewise, in our cohort, the hepatic inflammation linked to PBC may not be severe enough to impair LSM. Similar results have been recently published by Eddowes et al. in a prospective study on NAFLD, in which they found no significant influence of ALT values on LSM for each fibrosis stage.

Regarding cholestasis, the only evidence of impaired accuracy of LSM derives from a small cohort of patients ($n = 15$) with obstructive jaundice before endoscopic

retrograde cholangiopancreatography.⁽³⁶⁾ In line with our results, Corpechot et al., in a study conducted on 66 patients with primary sclerosing cholangitis with median ALP values at baseline of $2.2 \times$ ULN and median bilirubin values of $20.9 \mu\text{mol/L}$, showed that the only parameter associated with LSM was the stage of fibrosis.⁽³⁷⁾

Our study has several strengths. The study cohort was represented by a naïve cohort of patients at disease presentation; liver biopsies underwent centralized digital pathology review with double-blind reading; the identified cutoffs underwent validation in an independent cohort.

We acknowledge some limitations of the study. Liver biopsy in PBC is recommended by EASL CPGs only in patients with suspicion of overlap with other conditions (e.g., AIH or NASH), given that it is not necessary for diagnosis. This might have introduced a selection bias in our cohort (e.g., enrichment of severe cases). However, the cohort characteristics show that many patients had indolent disease (i.e., low values of ALP and transaminases). This can be explained, in some cases, by the diagnostic purpose of the biopsy in AMA-negative patients and by the historical experience in viral hepatitis in Italy, which might have made physicians more prone to stage chronic disease by liver biopsy. Furthermore, we did not establish whether repeat VCTE examination in the grey area would have generated consistent readings. Last, although $>40\%$ of patients in our cohort are overweight, the median BMI is not generalizable to other populations (e.g., United States) in which BMI is higher. However, as shown in other studies,^(12,38) BMI seems to affect more the quality of the measurement than the measurement itself (stiffness value) reducing the number of reliable results. When strictly quality criteria (i.e., >10 valid measurements, adequate fasting, application of the Boursier criteria, and availability of XL probe) are applied, providing reliable reading, the diagnostic accuracy of VCTE is held in obese patients. Thus, we anticipate that the cutoffs derived in our study may be applied in other populations with PBC with greater BMI, although a larger number of invalid measures are expected.

In conclusion, this study confirms the high applicability of VCTE with a dual cut-off approach in a naïve cohort of patients with PBC at diagnosis, and demonstrates that LSM readings are not influenced by BMI and biochemical markers of cholestasis and

liver inflammation. Additional studies are required to identify alternative methods for disease staging of patients in the grey area and evaluate LSM progression after therapy.

Acknowledgment: L.C., M.C., P.I., D.D., A.G., S.E.O., V.R., F.M., and M.L. are members of the European Reference Network on Hepatological Diseases (ERN RARE-LIVER). L.C., M.C., and P.I. thank AMAF Monza ONLUS and AIRCS for the unrestricted research funding. MIUR Department of Excellence project awarded to the Department of Mathematics, University of Rome Tor Vergata.

Author Contributions: L.C., M.C., and A.N. conceptualized the project. M.C., P.I., and G.C. provided funding acquisition. M.C. and A.N. supervised the project. All authors were involved in the acquisition of data. A.N. and L.C. performed the formal analysis. A.N., L.C., M.C., and M.V. performed interpretation of data and drafting of the manuscript. All authors revised the manuscript for important intellectual content. All authors approved the final version of the article, including the authorship list.

REFERENCES

- Carey EJ, Ali AH, Lindor KD. Primary biliary cirrhosis. *Lancet* 2015;386:1565-1575.
- Lammers WJ, Hirschfield GM, Corpechot C, Nevens F, Lindor KD, Janssen HLA, et al. Development and validation of a scoring system to predict outcomes of patients with primary biliary cirrhosis receiving ursodeoxycholic acid therapy. *Gastroenterology* 2015;149:1804-1812.e4.
- Carbone M, Sharp SJ, Flack S, Paximadas D, Spiess K, Adgey C, et al. The UK-PBC risk scores: derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. *HEPATOLOGY* 2016;63:930-950.
- Carbone M, Mells GF, Pells G, Dawwas MF, Newton JL, Heneghan MA, et al. Sex and age are determinants of the clinical phenotype of primary biliary cirrhosis and response to ursodeoxycholic acid. *Gastroenterology* 2013;144:560-569.e7.
- Carbone M, Nardi A, Flack S, Carpino G, Varvaropoulou N, Gavrilu C, et al. Pretreatment prediction of response to ursodeoxycholic acid in primary biliary cholangitis: development and validation of the UDCA Response Score. *Lancet Gastroenterol Hepatol* 2018;3:626-634.
- Murillo Perez CF, Hirschfield GM, Corpechot C, Floreani A, Mayo MJ, van der Meer A, et al. Fibrosis stage is an independent predictor of outcome in primary biliary cholangitis despite biochemical treatment response. *Aliment Pharmacol Ther* 2019;50:1127-1136.
- Carbone M, D'Amato D, Hirschfield GM, Jones DEJ, Mells GF. Letter: histology is relevant for risk stratification in primary biliary cholangitis. *Aliment Pharmacol Ther* 2020;51:192-193.
- Lindor KD, Bowlus CL, Boyer J, Levy C, Mayo M. Primary biliary cholangitis: 2018 practice guidance from the American Association for the Study of Liver Diseases. *HEPATOLOGY* 2019;69:394-419.
- Hirschfield GM, Beuers U, Corpechot C, Invernizzi P, Jones D, Marziani M, et al. EASL Clinical Practice Guidelines: the diagnosis and management of patients with primary biliary cholangitis. *J Hepatol* 2017;67:145-172.
- Hirschfield GM, Dyson JK, Alexander GJM, Chapman MH, Collier J, Hübscher S, et al. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. *Gut* 2018;67:1568-1594.
- Working Subgroup (English version) for Clinical Practice Guidelines for Primary Biliary Cirrhosis. Guidelines for the management of primary biliary cirrhosis. *Hepatol Res* 2014;44(Suppl. 1):71-90.
- Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, 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.
- Friedrich-Rust M, Ong M, Martens S, Sarrazin C, Bojunga J, Zeuzem S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. *Gastroenterology* 2008;134:960-974.e8.
- Corpechot C, Carrat F, Poujol-Robert A, Gaouar F, Wendum D, Chazouillères O, et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. *HEPATOLOGY* 2012;56:198-208.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *HEPATOLOGY* 2006;43:1317-1325.
- Wai CT, Greenon JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *HEPATOLOGY* 2003;38:518-526.
- Ludwig J, Dickson ER, McDonald GSA. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch A Pathol Anat Histol* 1978;379:103-112.
- Boursier J, Zarski JP, de Ledinghen V, Rousselet MC, Sturm N, Lebaill B, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *HEPATOLOGY* 2013;57:1182-1191.
- Trivedi PJ, Bruns T, Cheung A, Li KK, Kittler C, Kumagi T, et al. Optimising risk stratification in primary biliary cirrhosis: AST/platelet ratio index predicts outcome independent of ursodeoxycholic acid response. *J Hepatol* 2014;60:1249-1258.
- Floreani A, Cazzagon N, Martines D, Cavalletto L, Baldo V, Chemello L. Performance and utility of transient elastography and noninvasive markers of liver fibrosis in primary biliary cirrhosis. *Dig Liver Dis* 2011;43:887-892.
- Gómez-domínguez E, Mendoza J, García-buey L, Trapero M, Gisbert JP, Jones EA, et al. Transient elastography to assess hepatic fibrosis in primary biliary cirrhosis. *Aliment Pharmacol Ther* 2008;27:441-447.
- Viganò M, Paggi S, Lampertico P, Fraquelli M, Massironi S, Ronchi G, et al. Dual cut-off transient elastography to assess liver fibrosis in chronic hepatitis B: a cohort study with internal validation. *Aliment Pharmacol Ther* 2011;34:353-362.
- Sterling RK, King WC, Wahed AS, Kleiner DE, Khalili M, Sulkowski M, et al. Evaluating noninvasive markers to identify

- advanced fibrosis by liver biopsy in HBV/HIV co-infected adults. *HEPATOLOGY* 2020;71:411-421.
- 25) Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan WK, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol Hepatol* 2020;5:362-373.
 - 26) Tapper EB, Cohen EB, Patel K, Bacon B, Gordon S, Lawitz E, et al. Levels of alanine aminotransferase confound use of transient elastography to diagnose fibrosis in patients with chronic hepatitis C virus infection. *Clin Gastroenterol Hepatol* 2012;10:932-937.e1.
 - 27) Coco B, Oliveri F, Maina AM, Ciccorossi P, Sacco R, Colombatto P, et al. Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat* 2007;14:360-369.
 - 28) Romanque P, Stickel F, Dufour JF. Disproportionally high results of transient elastography in patients with autoimmune hepatitis. *Liver Int* 2008;28:1177-1178.
 - 29) Cobbold JFL, Taylor-Robinson SD. Transient elastography in acute hepatitis: all that's stiff is not fibrosis. *HEPATOLOGY* 2008;47:370-372.
 - 30) 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.
 - 31) Arena U, Vizzutti F, Corti G, Ambu S, Stasi C, Bresci S, et al. Acute viral hepatitis increases liver stiffness values measured by transient elastography. *HEPATOLOGY* 2008;47:380-384.
 - 32) Dhaliwal HK, Hoeroldt BS, Dube AK, McFarlane E, Underwood JCE, Karajeh MA, et al. Long-term prognostic significance of persisting histological activity despite biochemical remission in autoimmune hepatitis. *Am J Gastroenterol* 2015;110:993-999.
 - 33) Lüth S, Herkel J, Kanzler S, Frenzel C, Galle PR, Dienes HP, et al. Serologic markers compared with liver biopsy for monitoring disease activity in autoimmune hepatitis. *J Clin Gastroenterol* 2008;42:926-930.
 - 34) **Hartl J, Ehlken H**, Sebode M, Peiseler M, Krech T, Zenouzi R, et al. Usefulness of biochemical remission and transient elastography in monitoring disease course in autoimmune hepatitis. *J Hepatol* 2018;68:754-763.
 - 35) Hartl J, Denzer U, Ehlken H, Zenouzi R, Peiseler M, Sebode M, et al. Transient elastography in autoimmune hepatitis: timing determines the impact of inflammation and fibrosis. *J Hepatol* 2016;65:769-775.
 - 36) Millonig G, Reimann FM, Friedrich S, Fonouni H, Mehrabi A, Büchler MW, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *HEPATOLOGY* 2008;48:1718-1723.
 - 37) Corpechot C, Gaouar F, El Naggar A, Kemgang A, Wendum D, Poupon R, et al. 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;146:970-979.
 - 38) Chen J, Yin M, Talwalkar JA, Oudry J, Glaser KJ, Smyrk TC, et al. Diagnostic performance of MR elastography and vibration-controlled transient elastography in the detection of hepatic fibrosis in patients with severe to morbid obesity. *Radiology* 2017;283:418-428.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31810/supinfo.