

DR. LAURA CRISTOFERI (Orcid ID : 0000-0001-9846-7474)

DR. DILETTA OVERI (Orcid ID : 0000-0003-3561-8903)

DR. ELISABETTA DEGASPERI (Orcid ID : 0000-0002-9899-4669)

DR. ANDREA PALERMO (Orcid ID : 0000-0001-8057-9398)

DR. ALESSIO GERUSSI (Orcid ID : 0000-0002-5086-0514)

DR. GUIDO CARPINO (Orcid ID : 0000-0001-8570-2519)

Article type : Original Article

**Manuscript ID HEP-20-2710.R1**

**Accuracy of Transient Elastography in assessing fibrosis at diagnosis in naïve patients with Primary Biliary Cholangitis: a dual cut-off approach**

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

<sup>1</sup> *Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy*

<sup>2</sup> *European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/HEP.31810](https://doi.org/10.1002/HEP.31810)

This article is protected by copyright. All rights reserved

<sup>3</sup> *Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy.*

<sup>4</sup> *Section of Gastroenterology and Hepatology, PROMISE, University of Palermo, Palermo, Italy*

<sup>5</sup> *Department of Anatomical, Histological, Forensic Medicine and Orthopedics Sciences; Sapienza University of Rome, Rome, Italy*

<sup>6</sup> *Division of Hepatology, Ospedale San Giuseppe, University of Milan, Milan, Italy*

<sup>7</sup> *Department of Translational Medicine, Università degli Studi del Piemonte Orientale "A. Avogadro", Novara, Italy*

<sup>8</sup> *CRC "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*

<sup>9</sup> *Department of Medico-Surgical Sciences and Biotechnologies, Polo Pontino "Sapienza" University of Rome, Latina, Italy*

<sup>10</sup> *Department of Internal Medicine, University of Genova, Genova, Italy*

<sup>11</sup> *San Gerardo Hospital, Pathology department, Monza, Italy*

<sup>12</sup> *Histopathology Unit, AOU Maggiore della Carità, Novara, Italy.*

<sup>13</sup> *Department of Precision and Translational Medicine, Sapienza University of Rome, Rome, Italy*

<sup>14</sup> *Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy.*

<sup>15</sup> *Università Politecnica delle Marche", Department of Gastroenterology, Ancona, Italy.*

<sup>16</sup> *Studiosa Senior, University of Padova, Italy*

<sup>17</sup> *Scientific Consultant, IRCCS Negrar, Verona, Italy*

<sup>18</sup> *Department of Movement, Human and Health Science, University of Rome "Foro Italico", Rome, Italy.*

<sup>19</sup> *Department of Mathematics, University of Rome Tor Vergata, Rome, Italy*

\*These authors contributed equally and are joint first authors

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

**Keywords:** primary biliary cholangitis, transient elastography, fibrosis, risk stratification, diagnostic accuracy.

**Corresponding author:**

Marco Carbone, MD PhD

Division of Gastroenterology, Center for Autoimmune Liver Diseases

Department of Medicine and Surgery, University of Milano-Bicocca

Monza, Italy

Telephone number: +393927333154

Email: marco.carbone@unimib.it

**List of Abbreviations:**

Primary biliary cholangitis (PBC)

Alkaline phosphatase (ALP)

Ursodeoxycholic Acid (UDCA)

UDCA Response Score (URS)

Liver Biopsy (LB)

Liver Stiffness Measurement (LSM)

Vibration-Controlled transient elastography (VCTE)

Autoimmune hepatitis (AIH)

Body Mass Index (BMI)

Fibrosis-4 score (FiB-4)

AST to platelet ratio index (APRI)

Upper limit of normal (ULN)

Area under the receiver operating characteristic curves (AUROCs)

Positive likelihood ratio (LR+)

Negative likelihood ratio (LR-)

Negative predictive values (NPV)

Positive predictive value (PPV)

**Financial support:** 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 Number GR-2018-12367794) and “the Role of Auto-reactive Hepatic Natural Killer Cells in the Pathogenesis of Primary Biliary Cholangitis” (Grant/Award Number PE-2016-02363915).

**Conflict of Interest:** None of the authors has relevant or material financial interests that relate to the research described in this paper.

**Acknowledgements:** LC, MC, PI, DD, AG, SEO, VR, FM and ML are members of the European Reference Network on Hepatological Diseases (ERN RARE-LIVER). LC, MC, and PI thank AMAF Monza ONLUS and AIRCS for the unrestricted research funding. MIUR excellence Department project awarded to the Department of Mathematics, Univeristy of Rome Tor Vergata.

**Author contributions:**

LC, MC, AN conceptualized the project. MC, PI and GC provided funding acquisition. MC and AN supervised the project. All authors were involved in the acquisition of data. AN and LC performed the formal analysis. AN, LC, MC, and MV 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.

**Abstract**

**Background & Aims:** Liver fibrosis holds a relevant prognostic meaning in primary biliary cholangitis (PBC). Non-invasive 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 at disease presentation in PBC.

**Approach & 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 Ludwig system. Diagnostic accuracy was estimated using the area under the receiver operating characteristic curves(AUROC) for advanced fibrosis (Ludwig stage $\geq$ III). The effects of biochemical and clinical parameters on liver stiffness measurement (LSM) were appraised.

Derivation cohort consisted of 126 patients with valid LSM and LB, VCTE identified patients with advanced fibrosis with AUROC of 0.89. LSM cut-offs  $\leq$ 6.5kPa and  $>$ 11.0kPa enabled to exclude and confirm, respectively, advanced fibrosis (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 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 cut-offs (LSM $\leq$ 6.5kPa and $>$ 11.0kPa) able to discriminate at diagnosis the presence or the absence, respectively, of advanced fibrosis in PBC patients, with external validation. In patients with LSM between these two cut-offs, VCTE is not reliable and liver biopsy should be evaluated for accurate disease staging. BMI and liver biochemistry did not affect LSMs.

## 1. Introduction

Primary biliary cholangitis (PBC) is an autoimmune liver disease characterised by destructive cholangitis affecting the small intrahepatic 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 characterised 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 as 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 one year; this highlighted the need to incorporate liver fibrosis 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 liver fibrosis staging in chronic liver disease. However, due to its invasiveness and potential sampling error it is not recommended for staging purposes at diagnosis by international guidelines on PBC (8–11). Non-invasive evaluation of liver fibrosis with liver stiffness measurements (LSM) 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 (CPG) for disease staging at diagnosis and during follow-up (9). However, such recommendation is based on a cross-sectional, single-centre study including patients on treatment, at different time 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 cut-offs of LSM for liver fibrosis assessment in a treatment naïve population, at disease presentation, to implement early risk stratification in PBC.

We designed a study across six Italian liver centres 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 cut-offs for use in clinical practice.

## **2. Patients and methods**

### **2.1 Study design and participants**

This is a diagnostic test accuracy study using data from the *Italian PBC Registry*, an ongoing, non-interventional, multicentre, 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 at 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; Policlinico Umberto I, Roma).

To confirm our results, we tested the diagnostic performance of cut-offs 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 the results of this study (15).

### **2.2 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 to evaluate potential confounding effect of BMI, liver inflammation and cholestasis in predicting fibrosis by LSM in patient naïve to therapy.

### **2.3 Study definitions**

Diagnosis of PBC was made based on elevated ALP and the presence of anti-mitochondrial antibodies (AMA) at a titre >1:40 or specific antinuclear antibodies (ANA) 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.

### **2.4 Inclusion and exclusion criteria**

Patients  $\geq$  18 years of age, able to give written informed consent, were scheduled to have a 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. non-alcoholic steatohepatitis, 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.

### **2.5 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 (EDC) system with e-CRF 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, gender, BMI, date of liver biopsy, date of diagnosis, LSM value, liver function tests (LFTs) performed within three months from biopsy date and VCTE, i.e. serum albumin, ALP, gamma-glutamyl-transferase (GGT), total bilirubin, ALT, AST, and platelet count.

The following data were collected at baseline in the validation cohort: age, gender, BMI, date of liver biopsy, date of diagnosis, LSM, and LFTs performed within three months from biopsy date

and VCTE, i.e. ALP, bilirubin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), platelet. Age, LFTs and platelet were used to calculate Fibrosis-4 score (FIB-4), and AST to platelet ratio index (APRI) at diagnosis (16,17) in both cohorts.

## **2.6 Histopathologic Evaluation**

Percutaneous LB were performed according to the local standard procedure with 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 staining were used for review: haematoxylin and eosin, Masson trichrome stain, cytokeratin 7 stain. Slides were analysed independently by two experienced pathologists (GC and NZ) 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*; IV= *cirrhotic stage: fibrous septa and regenerative nodules* (18).

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

## **2.7 Vibration-controlled transient elastography (FibroScan®)**

VCTE examination by FibroScan® (Echosens, Paris, France) was performed in each centre by physicians trained and certified by 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 LSM 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. LSM were expressed in kiloPascal (kPa) and 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<7.1kPa), 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 Boursier criteria were deemed valid.

## 2.8 Statistical Analysis

To account for inter-laboratory variability, ALP, ALT AST and total bilirubin were expressed as a multiple of their respective upper limit of normal values (ULNs). Since most variables showed skewed distributions with significant departures from the normal density, a non-parametric 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 over-imposed. Wilcoxon rank-sum test was used to compare groups. Categorical variables were described by absolute frequencies and percentages; to compare groups, we used the  $\chi^2$  test (or Fisher 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 were reported. Wald test was used to assess significance. Poorly predicted observations were identified by the standardized deviance residuals. Diagnostic accuracy was evaluated using receiving operator (ROC) curves. Non-parametric stratified bootstrapping was used to compute confidence bands for ROC curves. Area under the ROC curve (AUROC) was reported together with its 95% CI. ROC curves were compared using De Long test.

Negative and positive predictive values (NPV, PPV), specificity and sensitivity, positive and negative likelihood ratios (LR+, LR-) were reported. In case of a single cut-off, 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 cut-off specificity and PPV greater than 0.90 and for the low exclusionary cut-off sensitivity and NPV greater than 0.90. If more than two cut-offs met these criteria, the additional requirement was to minimize the grey area.

Analyses were conducted using SAS (version 9.4) and R (version 3.4).

### 3. Results

#### 3.1 Patients characteristics

The study flow chart is represented in **Figure 1**. One hundred and sixty-seven patients with both VCTE and LB performed at diagnosis within 12 weeks from each other were consecutively enrolled in the study period. 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. The median time between VCTE and LB was 14.5 days (IQR 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, total bilirubin at baseline were 1.4xULN (IQR 1.0,2.4), 1.3xULN (IQR 0.9,2.0), and 0.6xULN (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 Boursier criteria. (**Figure 1**).

#### 3.2 Derivation cohort

##### 3.2.1 Prediction of advanced fibrosis

With the aim of predicting advanced fibrosis, 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*’ 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 stage* and *advanced stage* are reported in **Table 1**. LSMs according to fibrosis stage by Ludwig are presented in **Figure 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 advanced fibrosis. None of the biochemical parameters nor BMI showed a significant, additive predictive contribution to LSM (**Table 2**). Predictive value of LSM in identifying advanced fibrosis as measured from the AUROC was 0.89 (CI 0.83,0.95) (**Supplementary Figure 1**).

**Figure 3** shows the relationship between predicted probabilities of advanced fibrosis and LSM values according the fitted logistic model. The curve steeply increased in an interval of LSM between 7kPa to 11kPa, which corresponds to an estimated probability of advanced fibrosis ranging from 0.21 to 0.75 (**Figure 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 advanced fibrosis 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, while NPV was 0.95, PPV was only 0.62 with 19 patients falsely classified in *advanced stage*.

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 reliable prediction of advanced fibrosis (**Figure 4**). The diagnostic accuracy of different possible high and low cut-off values is reported in **Supplementary Table 1**. For the optimal lower threshold of 6.5 kPa, which defines the absence of advanced fibrosis, sensitivity and NPV were 0.91 and 0.96, respectively.

This threshold led to identify 70 *early stage* patients, of whom 67 (95.7%) being correctly predicted (**Table 3**). The three patients not correctly predicted had Ludwig stage III, had a LSM of 4.3, 5.9 and 6.1 kPa and none of them showed biochemical features of advanced fibrosis or cirrhosis. ALP was <1.5ULN in all of them and transaminases were <1.5ULN. Only one patient had bilirubin levels of 1.1xULN, with normal serum albumin and platelet count.

For the optimal higher threshold of 11.0 kPa, which defines the presence of advanced fibrosis, specificity and PPV were 0.99 and 0.94, respectively. This threshold identified 17 *advanced stage* patients, of whom 16 (94.1%) being 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) and mild increase of transaminase. In this patient APRI score was 1.72 and Fib-4 was 4.13 and were both consistent with advanced fibrosis.

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%.

### **3.2.2 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 were available for 114 (90.5%) patients. AUROC for advanced fibrosis 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 (**Supplementary Figure 2**).

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

Using the dual cut-off approach validated and currently in use for Fib-4 (16), among 60 patients with a 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.  $\geq 1.45$  and  $\leq 3.25$ ), 33 patients were in early stage and 11 were in advanced stage (**Supplementary Table 2**).

When the single, validated cut-off 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 advanced fibrotic stage (20).

### 3.2.3 Intention-to-diagnose analysis

The performance of the VCTE for discrimination of advanced fibrosis was tested also including patients with unreliable LSM results ( $n=10$ ). Among these patients, 4 had less than 10 valid measurements (3 with 8 measurement and 1 with 9 measurement) and 6 were unreliable according to Boursier criteria (median IQR/M 0.33 [IQR 0.32,0.34]). Patients with invalid LSMs had a significantly higher BMI than patients in 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, a NPV of 0.87 and 0.96, a LR+ of 19 and 93 and 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 one patient was wrongly classified in the early stage with Ludwig stage III at LB.

### **3.3 Validation cohort**

The two cut-offs identified in the derivation cohort were validated in an external cohort of 91 PBC patients at disease onset, naïve to UDCA with 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, Supplementary Table 3**). LSM vs fibrosis stage by Ludwig is presented as a boxplot in **Supplementary Figure 3**.

The lower threshold of 6.5 kPa identified 40 *early stage* patients, of whom 37(92.5%) being correctly predicted (**Supplementary Table 4**). The three patients not correctly predicted were all Ludwig stage III, LSM of 4.9, 5.6 and 6 kPa, normal albumin, bilirubin and platelet count; two patients had ALP and transaminases within 2xULN, one patient had ALP 8x ULN and ALTx4 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%) being correctly predicted (**Supplementary Table 4**).

Considering wrongly predicted cases, one patient had LSM of 29.8 kPa and Ludwig stage II, with normal serum albumin, bilirubin level and platelet count, ALPx4 ULN and ALTx4 ULN; the second patient had LSM 11.8kPa and Ludwig stage II, with bilirubin 1.2xULN, ALPx2.5 ULN and ALTx2.5 ULN and normal albumin and platelet count.

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

In 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 advanced fibrosis (**Supplementary Figure 4**). Similarly to what we showed in the derivation cohort, the density peak of both groups is mostly overlapping, whereas in LSM plot the overlap is confined to the *grey area*.

### **3.4 Grey Area (Overall cohort)**

The overall number of patients in grey area was 72 out of 217 (33.2%): 68.1% were in early stage, 31.9% in advanced stage. The median values of LSM were 7.9kPa (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 non-invasive scores of fibrosis, i.e. APRI and Fib-4, were not significantly different in patients in *early stage* and *advanced stage* (**Supplementary Table 5**).

#### 4. Discussion

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

This study provides a novel, pragmatic approach to threshold setting of non-invasive 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 timely; this should be done through hepatocellular carcinoma (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, URS and VCTE can be combined as non-invasive tools to implement baseline risk stratification in PBC by offering an estimated risk of treatment failure and disease stage, respectively (**Figure 5**). Finally, an accurate, non-invasive 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 the UK-PBC (7) study groups independently showed as 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 non-invasive surrogate of it, 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 accurate cut-off of LSM. A seminal French study in PBC

by Corpechot et al. (n=150) demonstrated for the first time 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, EASL CPG recommends the use of VCTE for disease staging at baseline and during follow-up (9). However, this study, while 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 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 importantly, this was a single-centre study lacking an external validation cohort (14).

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) (21,22). However, the exclusion of fibrosis could not be compared since in previous studies the evaluation of LSM in non-fibrotic patients was overlooked (14,22,23).

The dual cut-off approach is not novel. This has been proposed in hepatitis B patients by Viganò et al. and recently in HBV-HIV co-infected patients by Sterling et al (24,25). It highlights a *grey area* of inaccurate prediction which is inherent to the device, known to outperform with the extreme readings and to fail with intermediate ones (24,26).

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 non-invasive evaluation of liver fibrosis (27–32). 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 (33,34). 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 J. 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 (35,36). Likewise, in our cohort, the hepatic inflammation linked to PBC may be not enough severe to impair LSM. Similar results have been recently published by Eddowes et al. in a prospective study on non-alcoholic fatty liver disease (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 (37). 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 2.2xULN and median bilirubin values of 20.9 $\mu$ mol/L showed that the only parameter associated with LSM was the stage of fibrosis (38).

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

We acknowledge some limitations of the study. Liver biopsy in PBC is recommended by EASL CPG only in patients with suspicion of overlap with other conditions, e.g. AIH or NASH, as 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 more than 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 (12,39), 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 Boursier criteria, availability of XL probe) are applied, providing reliable reading, the diagnostic accuracy of VCTE is held in obese patients. Thus, we anticipate that the

cut-offs derived in our study may be applied in other population 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 demonstrate 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 to evaluate the LSM progression after therapy.

## References

1. Carey EJ, Ali AH, Lindor KD. Primary biliary cirrhosis [Internet]. Vol. 386, The Lancet.;2015. p.1565–75.
2. 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 Dec;149(7):1804-1812.e4.
3. 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 Mar;63(3):930–50.
4. 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 Mar;144(3):560-569.e7.
5. 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(9):626–34.
6. 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 Nov 1;50(10):1127–36.
7. Carbone M, D'Amato D, Hirschfield GM, Jones DEJ, Mells GF. Letter: histology is

- relevant for risk stratification in primary biliary cholangitis. Vol. 51, *Alimentary Pharmacology and Therapeutics*. 2020. p. 192–3.
8. 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(1):394–419.
  9. Hirschfield GM, Beuers U, Corpechot C, Invernizzi P, Jones D, Marzioni M, et al. EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. *J Hepatol*. 2017 Jul 1;67(1):145–72.
  10. 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(9):1568–94.
  11. Biliary WS (English version) for CPG for P, Cirrhosis. Guidelines for the management of primary biliary cirrhosis. *Hepatology Res*. 2014;44(suppl 1):71–90.
  12. 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(6):1717–30.
  13. Friedrich-Rust M, Ong MF, 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(4).
  14. 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(1):198–208.
  15. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015 Oct 28;351.
  16. 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 Jun;43(6):1317–25.
  17. Wai CT, Greenson 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 Aug 1;38(2):518–26.

18. Ludwig J, Dickson ER, McDonald GSA. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch A Pathol Anat Histol*. 1978 Jun; 379(2):103–12.
19. Boursier J, Zarski JP, de Ledinghen V, Rousselet MC, Sturm N, Lebail B, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology*. 2013;57(3):1182–91.
20. Trivedi PJ, Bruns T, Cheung A, Li K-K, 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 Jun;60(6):1249–58.
21. 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 Jul;56(1):198–208.
22. 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(11):887–92.
23. Gómez-Dominguez 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(5):441–7.
24. 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 Aug;34(3):353–62.
25. 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 Feb 1;71(2):411–21.
26. 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 Apr 1;5(4):362–73.
27. 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(8):932-937.e1.

28. 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 May;14(5):360–9.
29. Romanque P, Stickel F, Dufour JF. Disproportionally high results of transient elastography in patients with autoimmune hepatitis [Internet]. Vol. 28, *Liver International*. *Liver Int*; 2008. p. 1177–8.
30. Cobbold JFL, Taylor-Robinson SD. Transient elastography in acute hepatitis: All that's stiff is not fibrosis. Vol. 47, *Hepatology*. *Hepatology*; 2008. p. 370–2.
31. Sagir A, Erhardt A, Schmitt M, Häussinger D. Transient elastography is unreliable for detection of cirrhosis in patients with acute liver damage. *Hepatology*. 2008 Feb;47(2):592–5.
32. 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(2):380–4.
33. 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 Jul8;110(7):993–9.
34. 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 Sep;42(8):926–30.
35. 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(4):754–63.
36. 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 Oct1;65(4):769–75.
37. 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 Nov;48(5):1718–23.
38. Corpechot C, Gaouar F, Naggar A El, 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 Apr;146(4):970-9

39. 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 May 1;283(2):418–28.

Author names in **bold** designate shared co-first authorship

## Figures legends

### Figure 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 measurement were collected and classified as “very reliable” and “reliable” according to Boursier Criteria (19).

*Abbreviations:* AIH: Autoimmune Hepatitis; LB: Liver Biopsy; LSM: Liver Stiffness Measurement; MRCP: Magnetic Resonance Cholangiopancreatography; VCTE: Vibration Controlled Transient Elastography.

**Figure 2.** Distribution of LSM according to histological stage by Ludwig in derivation cohort. LSM increase significantly in the fibrotic stages III and IV by Ludwig system (Kruskal-Wallis test  $p < 0.00001$ ).

*Abbreviations:* LSM, liver stiffness measurement

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

*Abbreviations:* LSM, liver stiffness measurement

**Figure 4.** Density plot of LSM (panel A), Fib-4 (panel B) and APRI score (panel 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 (panel A), the grey area highlights the interval of LSM in which TE is not reliable. In the APRI and Fib-4 density plots (panel B-C) the peak of density of patients in early stage 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 (panel 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 (panel C) express the cut-off of 0.54 validated in PBC (20). Extreme observations were excluded (4 cases).

*Abbreviations:* APRI, AST to Platelet Index; Fib-4, Fibrosis 4 score; LSM, Liver Stiffness Measurement.

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

*Abbreviations:* UDCA, Ursodeoxycholic acid; LSM, Liver Stiffness Measurement.

**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 gender	84	92.3	29	82.8
ALP xULN	1.3	0.8-2.1	1.8	1.3-3.5
ALT xULN	1.1	0.8-2.0	1.9	1.4-2.3
AST xULN	1.0	0.9-1.5	1.4	1.0-2.0
Total Bilirubin xULN	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 x10 <sup>9</sup> /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

Abbreviations: Alanine Aminotransferase (ALT), Alkaline Phosphatase (ALP), Aspartate Aminotransferase (AST), AST to platelet index (APRI), Body Mass Index (BMI), Fibrosis 4 score (FIB-4), Platelet count (PLT), Upper Limit of Normal (ULN).

**Table 2.** Multivariable logistic model fitted to observed data.

	<b>Odds ratio</b>	<b>95% CI</b>	<b>p-value</b>
LSM (kPa)	1.76	(1.29, 2.41)	0.0004
Age (years)	1.04	(0.97, 1.11)	0.3153
Gender (female vs male)	0.74	(0.10, 5.31)	0.7661
ALP x ULN (log scale)	1.13	(0.45, 2.84)	0.7955
ALT x ULN (log scale)	0.96	(0.35, 2.65)	0.9310
Total Bilirubin x ULN (log scale)	1.46	(0.44, 4.80)	0.5384
Albumin (g/dl)	1.10	(0.20, 6.00)	0.9117
PLT x10 <sup>9</sup> /L	1.00	(0.99, 1.01)	0.2954
BMI	1.00	(0.85, 1.19)	0.9854

Abbreviations: ALP: Alkaline Phosphatase, ALT: Alanine Aminotransferase, BMI: Body Mass Index, kPa: kilopascal, LSM: Liver Stiffness Measurement, PLT: platelet count, ULN: Upper Limit of Normal.

**Table 3.** Ludwig stage stratified by risk class prediction of fibrosis in the logistic regression model in the intention-to-diagnose cohort.

	<b>Early stage (LSM ≤ 6.5kPa)</b> n (column %)	<b>Grey area (6.5 &lt;LSM≤11.0kPa)</b> n (column %)	<b>Advanced stage (LSM &gt; 11.0kPa)</b> n (column %)	<b>Unreliable LSM</b> n (column %)	<b>Total</b> n (column %)
<b>Ludwig stage I</b>	40 (57.1)	13 (33.3)	1 (5.9)	1 (10.0)	55 (40.4)
<b>Ludwig stage II</b>	27 (38.6)	10 (25.6)	0 (0)	2 (20.0)	39 (28.7)
<b>Ludwig stage III</b>	3 (4.3)	15 (38.5)	6 (35.3)	6 (60.0)	30 (22.1)
<b>Ludwig stage IV</b>	0 (0)	1 (2.6)	10 (58.8)	1 (10.0)	12 (8.8)
<b>Total</b>	70	39	17	10	136

**N 167** patients with VCTE and LB performed at diagnosis within 12 weeks from each other

- **N 7** patients with concomitant viral hepatitis
- **N 3** patients with evidence of concomitant biliary disease at MRCP
- **N 12** patients with overlap with autoimmune hepatitis
- **N 9** patients with non-interpretable LB<sup>(1)</sup>

**N136** Intention-to-diagnose cohort

- **N 10** patients with invalid LSM<sup>(2)</sup>

**N126** Derivation cohort







