Research Article
DILI, Autoimmune, Cholestatic and Genetic Diseases

JOURNAL OF HEPATOLOGY

Liver stiffness measurement by vibration-controlled transient elastography improves outcome prediction in primary biliary cholangitis

Christophe Corpechot[1,*], Fabrice Carrat[2], Farid Gaouar[1], Frederic Chau[2], Gideon Hirschfield[3], Aliya Gulamhusein[3], Aldo J. Montano-Loza[4], Ellina Lytvyak[4], Christoph Schramm[5], Albert Pares[6], Ignasi Olivas[6], John E. Eaton[7], Karim T. Osman[7], George Dalekos[8], Nikolaos Gatselis[8], Frederik Nevens[9], Nora Cazzagon[10], Alessandra Zago[10], Francesco Paolo Russo[10], Nadir Abbas[11], Palak Trivedi[11], Douglas Thorburn[12], Francesca Saffioti[12], Laszlo Barkai[12], Davide Roccarina[12], Vicenza Calvaruso[13], Anna Fichera[13], Adèle Delamarre[14], Esli Medina-Morales[15], Alan Bonder[15], Vilas Patwardhan[15], Cristina Rigamonti[16], Marco Carbone[17], Pietro Invernizzi[17], Laura Cristoferi[17], Adriaan van der Meer[18], Rozanne de Veer[18], Ehud Zigmond[19], Eyal Yehezkel[19], Andreas E. Kremer[20], Ansgar Deibel[20], Jérôme Dumortier[21], Tony Bruns[22], Karsten Große[22], Georges-Philippe Pageaux[23], Aaron Wetten[24], Jessica Dyson[24], David Jones[24], Olivier Chazouillères[1], Bettina Hansen[3], Victor de Lédinghen [14], on behalf of the Global & ERN Rare-Liver PBC Study Groups

[1]Reference Center for Inflammatory Biliary Diseases and Autoimmune Hepatitis, European Reference Network on Hepatological Diseases (ERN Rare-Liver), Saint-Antoine Hospital, Assistance Publique - Hôpitaux de Paris, Inserm UMR_S938, Saint-Antoine Research Center, Sorbonne University, Paris, France; [2]Public Health Unit, Saint-Antoine Hospital, Assistance Publique – Hôpitaux de Paris, Pierre Louis Institute of Epidemiology and Public Health, Sorbonne University, Paris, France; [3]Toronto Centre for Liver Disease, University Health Network, University of Toronto, Toronto, Canada; [4]Division of Gastroenterology and Liver Unit, University of Alberta, Edmonton, Canada; [5]Department of Medicine I and Martin Zeitz Center for Rare Diseases, European Reference Network on Hepatological Diseases (ERN Rare-Liver), University Medical Center Hamburg–Eppendorf, Hamburg, Germany; [6]Liver Unit, Hospital Clínic, University of Barcelona, The August Pi i Sunyer Biomedical Research Institute, Biomedical Research Networking Center in Hepatic and Digestive Diseases, European Reference Network on Hepatological Diseases (ERN Rare-Liver), Barcelona, Spain; [7]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, USA; [8]Department of Medicine and Research Laboratory of Internal Medicine, National Expertise Center of Greece in Autoimmune Liver Diseases, European Reference Network on Hepatological Diseases (ERN Rare-Liver), General University Hospital, Larissa, Greece; [9]Division of Hepatology and Liver Transplantation, European Reference Network on Hepatological Diseases (ERN Rare-Liver), University Hospitals KU, Leuven, Belgium; [10]Department of Surgery, Oncology and Gastroenterology, European Reference Network on Hepatological Diseases (ERN Rare-Liver), University of Padova, Padova, Italy; [11]Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom; [12]University College London Institute for Liver and Digestive Health, Royal Free Hospital, London, United Kingdom; [13]Section of Gastroenterology and Hepatology, PROMISE, University of Palermo, Palermo, Italy; [14]Department of Hepatology, University Hospitals of Bordeaux, Pessac, France; [15]Department of Medicine, Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, USA; [16]Department of Internal Medicine, Università del Piemonte Orientale, Novara, Italy; [17]Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, European Reference Network on Hepatological Diseases (ERN Rare-Liver), University of Milano-Bicocca, Monza, Italy; [18]Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, The Netherlands; [19]The Research Center for Digestive Tract and Liver Diseases, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; [20]Department of Gastroenterology and Hepatology, University Hospital Zürich, Zürich, Switzerland; [21]Department of Gastroenterology and Hepatology, Edouard Herriot Hospital, Hospices Civils de Lyon, Claude Bernard University, Lyon, France; [22]Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany; [23]Department of Hepatology and Liver Transplantation, University Hospital, Montpellier, France; [24]Department of Hepatology and Liver Transplantation, Newcastle upon Tyne Hospitals, Newcastle University, United Kingdom

Background & Aims: Liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE) has been shown to predict outcomes of patients with primary biliary cholangitis (PBC) in small-size studies. We aimed to validate the prognostic value of LSM in a large cohort study.

Methods: We performed an international, multicentre, retrospective follow-up study of 3,985 patients with PBC seen at 23 centres in 12 countries. Eligibility criteria included at least 1

Keywords: PBC; FibroScan; Mortality; Transplantation; Prognosis.
Received 11 March 2022; received in revised form 25 May 2022; accepted 13 June 2022; available online xxx

* Corresponding author. Address: Reference center for inflammatory biliary diseases and autoimmune hepatitis, Saint-Antoine Hospital, Assistance Publique – Hôpitaux de Paris; Inserm UMR_S938, Saint-Antoine Research Center, Sorbonne University; 184 rue du faubourg Saint-Antoine, 75571 Paris, Cedex 12, France.
E-mail address: christophe.corpechot@aphp.fr (C. Corpechot).

https://doi.org/10.1016/j.jhep.2022.06.017

reliable LSM by VCTE and a follow-up ≥ 1 year. Independent derivation (n = 2,740) and validation (n = 568) cohorts were built. The primary endpoint was time to poor clinical outcomes defined as liver-related complications, liver transplantation, or death. Hazard ratios (HRs) with CIs were determined using a time-dependent multivariable Cox regression analysis.

Results: LSM was independently associated with poor clinical outcomes in the derivation (5,324 LSMs, mean follow-up 5.0 ± 3.1 years) and validation (1,470 LSMs, mean follow-up 5.0 ± 2.8 years) cohorts: adjusted HRs (95% CI) per additional kPa were 1.040 (1.026–1.054) and 1.042 (1.029–1.056), respectively (p <0.0001 for both). Adjusted C-statistics (95% CI) at baseline were 0.83 (0.79–0.87) and 0.92 (0.89–0.95), respectively. Between 5 and 30 kPa, the log-HR increased as a monotonic function of LSM. The predictive value of LSM was stable in time. LSM improved the prognostic ability of biochemical response criteria, fibrosis scores, and prognostic scores. The 8 kPa and 15 kPa cut-offs optimally separated low-, medium-, and high-risk groups. Forty percent of patients were at medium to high risk according to LSM.

Conclusions: LSM by VCTE is a major, independent, validated predictor of PBC outcome. Its value as a surrogate endpoint for clinical benefit in PBC should be considered.

Lay summary: Primary biliary cholangitis (PBC) is a chronic autoimmune disease, wherein the body's immune system mistakenly attacks the bile ducts. PBC progresses gradually, so surrogate markers (markers that predict clinically relevant outcomes like the need for a transplant or death long before the event occurs) are often needed to expedite the drug development and approval process. Herein, we show that liver stiffness measurement is a strong predictor of clinical outcomes and could be a useful surrogate endpoint in PBC trials.

Introduction

Primary biliary cholangitis (PBC) is a slowly progressive, autoimmune cholestatic liver disease that, if undertreated, eventually leads to cirrhosis and its life-threatening complications.[1] Ursodeoxycholic acid (UDCA), the current universal standard-of-care therapy for PBC,[2–4] is associated with prolonged liver transplantation (LT)-free survival, regardless of the disease stage and the observed biochemical response.[5] Nevertheless, patients with PBC with no complete response to UDCA (approximately one-third), mainly due to persistently elevated levels of alkaline phosphatase (ALP) or bilirubin after 12 months of treatment, remain at high risk of poor long-term clinical outcome.[6,7] These patients are currently considered candidates for second-line therapies, whether available in routine care (on-label [obeticholic acid] or off-label [fibrates]) or still under clinical investigation in trials, in association with UDCA therapy.[8,9]

Because PBC is a very slow chronic disease, liver biochemical tests are used as primary surrogate endpoints for mortality or need for LT in most ongoing clinical trials.[10] Yet, the predictive ability of such biochemical endpoints is only partial and their prognostic value in the specific context of combination therapies is not fully established. Therefore, the development and validation of new surrogate markers of the disease that could be used in combination with biochemical response as substitutes for clinical outcomes appear as a necessary condition for a better evaluation of therapeutic strategies for PBC.[11,12] In this regard, the need to incorporate fibrosis stage into PBC risk stratification systems has recently been proposed since assessment of fibrosis stage was shown to grant prognostic value beyond biochemical response to treatment.[13] These findings strongly support the inclusion of non-invasive markers of fibrosis in all future prognostic tools developed for PBC.

The introduction of vibration-controlled transient elastography (VCTE) using FibroScan® in 2003 has revolutionized the evaluation of liver diseases.[14] Today, of the various alternative techniques of liver elastography, VCTE remains one of the most popular and widely used worldwide.[15] Liver stiffness measurement (LSM) by VCTE has indeed been shown as a very simple and reliable means for diagnosing cirrhosis or advanced fibrosis in many chronic liver diseases, including PBC.[16,17] It has further been linked to the risk of portal hypertension, cirrhotic decompensation, hepatocellular carcinoma, and liver-related mortality in different liver conditions, including PBC.[16,18–20] That said, although LSM is currently recommended by clinical practice guidelines for the assessment and monitoring of PBC,[21–23] its usefulness as a reliable predictor of clinical outcome in this disease remains to be proven. In the present large-scale, international follow-up study of patients with PBC, our aim was to assess and validate the prognostic value of LSM using FibroScan® alongside those of pre-established biochemical response criteria and prognostic scores.

Patients and methods

Study population and design

Between November 2019 and November 2021, we conducted an international, multicentre, retrospective cohort study under the auspices of the Global PBC Study Group and the European Reference Network (ERN) for rare hepatological diseases (RareLiver). A total of 23 tertiary centres from 12 countries in Europe, North America and the Middle East participated in the study. Unrelated derivation and validation cohorts were consecutively built, including a total of 3,284 and 701 patients with PBC, respectively. The contribution of each participating centre to the constitution of these original data sources is shown in Table S1.

All patients had a diagnosis of PBC established according to international guidelines.[2,4] To be eligible for analysis, patients had to have at least 1 reliable VCTE measurement using FibroScan® and a minimum of 1 year of subsequent follow-up. All available LSMs in a single patient could be collected, regardless of the probe (M or XL) used. A reliable LSM was defined by at least 10 valid measurements with an interquartile/median ratio ≤30% according to the manufacturer's recommendations and usual definition.[14,24] All unreliable LSMs were excluded from analysis. Patients with a history of LT or cirrhotic complications before first LSM were excluded, as well as those with a history of autoimmune hepatitis overlap syndrome or long-term corticosteroid and/or immunosuppressive therapy.

Blood test results within 2 months of LSMs were collected when available. We considered as explanatory variables that were previously shown to predict survival in PBC and those judged to be clinically relevant, including age, sex, serum bilirubin and albumin, ALP, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), platelet count, international normalised ratio, serum creatinine, histological stage, and concomitant treatment with UDCA, obeticholic acid, and/or fibrates. The Globe, UK-PBC, model for end-stage liver disease (MELD), and fibrosis-4 (FIB-4)

scores were calculated whenever possible.[25–28] Because most patients were included several years after the start of treatment, the Globe and UK-PBC scores were calculated considering LSM date as at 1 year of treatment. Biochemical response to treatment was defined according to the Paris-2, Toronto, and Globe criteria using the same rule.[6,25,29]

The study entry was the date of the first reliable LSM available. The time between the date of entry and the occurrence of a clinically significant event was calculated. Poor clinical outcomes were defined as death, LT or registration on the liver transplant waiting list, or complication of cirrhosis, including variceal bleeding, ascites, hepatic encephalopathy, or hepatocellular carcinoma, when no death or LT was reported at time of last follow-up. The patients who did not experience any of these events were censored at the date of last known visit.

The study was conducted in accordance with the Declaration of Helsinki. It was a retrospective observational analysis based on previously collected routine care data with no opposition by patients. The protocol was approved by the institutional research board of each participating centre in accordance with their local regulations.

Statistical analysis

We conducted a multivariable survival analysis using LSM and biochemical covariates as time-dependent variables. The primary endpoint of the study was survival without poor clinical outcomes, including death, LT, or liver complications as defined above. This endpoint was assessed in the derivation and validation cohorts using a time-dependent, multivariable-adjusted Cox proportional-hazards model. For each covariate analysed, the association with the primary endpoint was expressed as a hazard ratio (HR) with a 95% CI. To explore the shape or log-linearity of the association between LSM and the HR of poor clinical outcome, we estimated a Cox model with a smooth transformation of LSM at baseline using cubic B-spline with 4 equally spaced knots. Several sensitivity analyses on different subpopulations of patients were performed: i) patients treated with UDCA monotherapy beyond 12 months (exclusion of patients who received second-line therapy, including fibrates, obeticholic acid, or corticosteroids; exclusion of LSMs performed before 12 months of UDCA); ii) patients assessed using FibroScan® M probe only (exclusion of LSMs performed with XL probe). The predictive performance of LSM was evaluated at baseline using Harrell's C-index and time-dependent receiver-operating characteristic curve analysis.[30] Inverse probability of censoring weighting was used to calculate area under the receiver operating characteristic (ROC) curves as a function of time. Using bivariate models, the prognostic ability of LSM was individually faced with those of the Paris-2, Toronto, and Globe criteria of biochemical response, as well as those of the Globe, UK-PBC, MELD, APRI, and FIB-4 scores. Comparison of Harrell's C-indices was performed using a non-parametric bootstrap approach with 500 replicates.[31] The thresholds of LSM and Globe score that best discriminated patients into low-, medium-, and high-risk groups for poor clinical outcomes were determined from the derivation cohort using an unbiased grid search method with 2 cut-offs and optimal model selection based on log-likelihood.[32] For simplicity, the LSM thresholds were rounded to the nearest whole number. Risk heat maps for poor clinical outcomes were generated at different time points of follow-up from LSM and Globe score values at baseline. The goodness-of-fit of the prognostic model was assessed using a Poisson regression model applied to the validation cohort.[33] Survival rates were determined using the Kaplan-Meier or Cox estimates. *P* values <0.05 were considered significant. All statistical analyses were performed using SAS version 9.4 (SAS corporation).

Results

Baseline characteristics

The flow chart of the study is shown in Fig. S1. Of the 3,985 patients initially registered, of whom 3,284 and 701 were in the derivation and validation cohorts, respectively, 3,308 (83%) were eligible for analysis, of whom 2,740 (83%) were in the derivation cohort and 568 (81%) in the validation cohort. Of the 6,859 and 1,656 LSMs registered in the derivation and validation cohorts, respectively, 1,044 (15.2%) and 148 (8.9%) were missing (supposedly failed) or unreliable. The total number of LSMs available for analysis was 5,324 (median per patient 2, range 1–5) in the derivation cohort and 1,470 (median per patient 2, range 1–8) in the validation cohort. In patients with ≥2 LSMs, the mean time between consecutive LSMs was 2.3 ± 1.5 years and 1.7 ± 1.4 years, respectively. The entry characteristics of patients in both cohorts are shown in Table 1. Disease characteristics, such as serum levels of total bilirubin and transaminases, LSM, APRI and FIB-4 scores, or the percentage of patients with cirrhosis or advanced fibrosis, indicated more severe disease at baseline in the validation cohort than in the derivation cohort.

Treatment and disease outcome

As expected, the vast majority of patients (>90%) were treated with UDCA (more than 50% at entry received UDCA for >1 year), while second-line therapy with obeticholic acid or fibrates at any time of follow-up was reported in approximately 20% (18.0% in the derivation cohort, 22.2% in the validation cohort). The mean follow-up was 5.0 ± 3.1 years (median 4.1 years, IQR 2.5–6.9 years, range 1.0–16.4 years) in the derivation cohort and 5.0 ± 2.8 years (median 4.5 years, IQR 2.7–6.3 years, 1.0–15.7 years) in the validation cohort. The rates of all-cause and liver-related deaths and of liver complications without subsequent death/LT at time of last follow-up were comparable between cohorts, but LT was more frequently observed in the validation cohort than in the derivation cohort, which is consistent with more advanced disease in the former cohort (Table 1). In total, 274 (10.1%) patients in the derivation cohort and 82 (14.4%) in the validation cohort reached the primary endpoint. The 5-, 10-, and 15-year survival rates without LT or liver complications were 93.5%, 81.3%, and 49.3%, respectively in the derivation cohort, and 87.5%, 75.3%, and 42.6%, respectively in the validation cohort (*p* value for log-rank test <0.0001, Fig. 1).

Association of LSM with disease outcome

In a time-dependent univariate analysis, LSM was significantly associated with poor clinical outcomes in both the derivation (HR per kPa added 1.065, 95% CI 1.057–1.074, p <0.0001) and validation (HR per kPa added 1.061, 95% CI 1.050–1.071, p <0.0001) cohorts. Between 5 and 30 kPa, the log hazard of clinical outcomes increased as a monotonic function of LSM (Fig. 2A). In a time-dependent, multivariable analysis within the derivation cohort, including age, sex, and known prognostic factors such as total bilirubin, albumin, ALP, AST, and platelet count, LSM was the covariate most strongly associated with poor clinical outcomes (Table 2). This result was replicated in the

Table 1. Entry and follow-up characteristics of patients.

Characteristics	Derivation cohort (n = 2,740)	Validation cohort (n = 568)
Year at entry	2013.9 ± 3.3	2012.5 ± 5.2
Age at entry (year)	57.9 ± 11.7	55.3 ± 12.3
Age at diagnosis (year)	51.8 ± 12.0	52.1 ± 12.1
Female sex (yes)	2,496 (91.1%)	509 (89.6%)
BMI ($kg.m^{-2}$)	26.2 ± 6.3	25.1 ± 4.7
Obesity (yes)	546 (19.9%)	80 (14.1%)
Total bilirubin (μmol/L)	11.4 ± 13.1	15.5 ± 22.0
ALP (xULN)	1.8 ± 8.2	1.9 ± 1.4
ALT (xULN)	1.2 ± 1.1	1.7 ± 1.9
Albumin (g/L)	41.8 ± 3.8	41.8 ± 4.3
Platelet count (G/L)	252.8 ± 79.3	243.8 ± 88.5
LSM (kPa)	9.5 ± 9.0	11.7 ± 11.9
XL probe (yes)	351 (12.8%)	74 (13.0%)
APRI score	0.55 ± 0.57	0.73 + 0.77
FIB-4 score	1.64 ± 1.25	1.87 ± 1.73
Advanced fibrosis* (yes)	660 (24.1%)	183 (32.3%)
Cirrhosis* (yes)	434 (15.8%)	130 (22.9%)
UDCA (at any time)	2,704 (98.7%)	550 (96.8%)
UDCA at entry ≥1 year (yes)	1,696 (61.9%)	302 (53.2%)
Obeticholic acid (at any time)	208 (7.6%)	43 (7.6%)
Fibrates (at any time)	362 (13.2%)	90 (15.9%)
Follow-up time (year)	5.0 ± 3.1	5.0 ± 2.8
No. of LSM per patient	2.1 ± 1.1	2.4 ± 1.3
All-cause death (yes)	145 (5.3%)	26 (4.6%)
Liver-related death (yes)	62 (2.3%)	13 (2.3%)
Liver transplantation (yes)	56 (2.0%)	39 (6.9%)
Liv. compl. wo LT/death (yes)	77 (2.8%)	17 (3.0%)

ALP, alkaline phosphatase; APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; FIB-4, fibrosis-4; LSM, liver stiffness measurement; UDCA, ursodeoxycholic acid.
*Advanced fibrosis was defined by histological stage 3-4 or LSM >10.7 kPa.[16] Cirrhosis was defined by histological stage 4, LSM >16.9 kPa,[16] or platelet count <150,000 G/L.

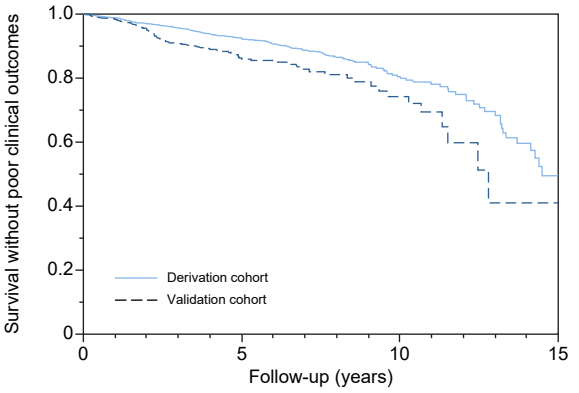
validation cohort (Table S2) and remained true even when the analysis was restricted to patients treated with UDCA only and to LSMs beyond 1 year of treatment (Table S3). Excluding the measurements obtained with the XL probe (n = 682) did not affect the results either (Table S4).

Predictive ability of LSM

At baseline, the adjusted Harrell's C-index (95% CI) for LSM was 0.8305 (0.7933–0.8677) in the derivation cohort and 0.9181 (0.8895–0.9467) in the validation cohort. When analysed using a time-dependent ROC curve analysis, the predictive ability of baseline LSM for poor clinical outcomes was stable over years (Fig. 2B). The prognostic performance at baseline of the biochemical response criteria (Paris-2, Toronto, and Globe), fibrosis scores (APRI, FIB-4) and prognostic scores (Globe, UK-PBC, and MELD) were determined both individually and alongside LSM in the derivation (Table 3) and validation (Table S5) cohorts. LSM significantly added to the performance of these predictors at baseline. In a time-dependent bivariate analysis, LSM showed independent predictive value compared to each predictor (Table S6). As the Globe score had the second highest predictive value after LSM, it was chosen as the main competitor of LSM for the rest of the study.

LSM and biochemical predictors interactions

Interactions between LSM and biochemical predictors were evaluated both semi-quantitatively and continuously. In the derivation cohort, the thresholds that optimally separated patients at baseline into low-, medium-, and high-risk groups were 8 kPa and 15 kPa for LSM, and 0.5 and 1.8 for the Globe score (p <0.0001 for both, Fig. 3A). The discriminative ability of these thresholds was confirmed in the validation cohort (p <0.0001 for both, Fig. 3B). The distributions of low-, medium-, and high-risk groups based on LSM were 61.9%, 26.5%, and 11.6% in the derivation cohort and 52.1%, 29.1%, and 18.8% in the validation cohort. Those based on the Globe score were 70.5%, 24.9%, and 4.6%, and 65.8%, 26.9%, and 7.3%, respectively. Compared to low-risk group, the HRs associated with medium- and high-risk groups were 3.496 (95% CI 2.412–5.069) and 15.677 (95% CI



Patients at risk				
Derivation cohort	2,740	1,096	217	19
Validation cohort	568	240	39	2

Fig. 1. Probability of survival without poor clinical outcomes. Probability of survival without poor clinical outcomes of patients with primary biliary cholangitis in the derivation (n = 2,740, solid line) and validation (n = 568, dotted line) cohorts. Survival rates between both cohorts were significantly different (log-rank test, p <0.0001).

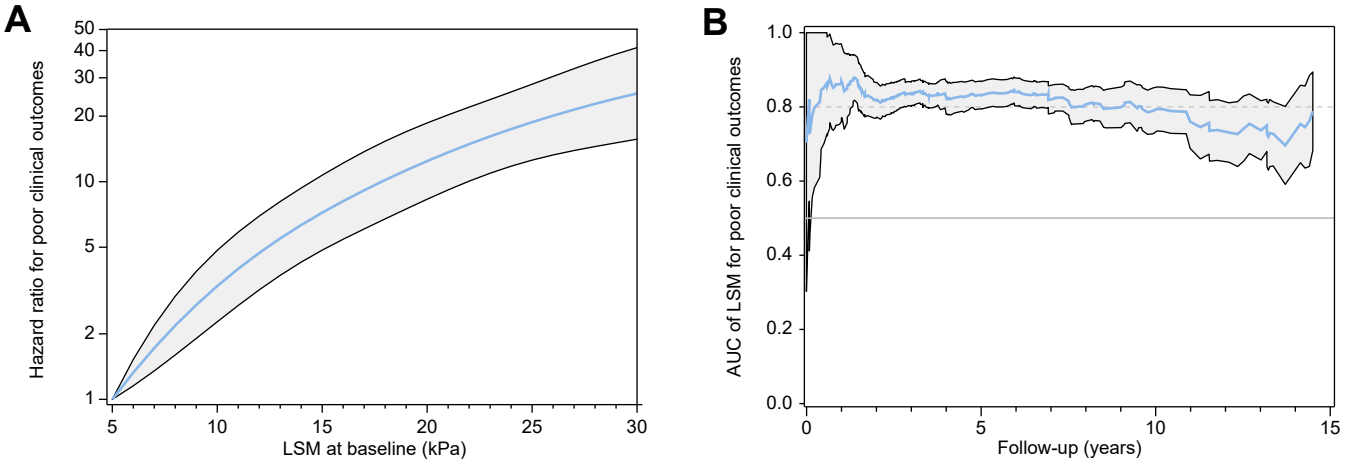


Fig. 2. Performance of LSM in predicting poor clinical outcomes. (A) The log hazard of poor clinical outcomes as estimated using a cubic spline function of LSM at baseline. (B) Predictive ability of baseline LSM for poor clinical outcomes as evaluated using a time-dependent area under ROC curve analysis. LSM, liver stiffness measurement.

10.878–22.592), respectively for LSM and 4.157 (95% CI 3.013–5.734) and 24.720 (95% CI 16.773–36.430), respectively for the Globe score. LSM thresholds were able to separate patients into risk groups irrespective of age category (≤45 years *vs.* >45 years; Fig. S2). Considered together, LSM and Globe score thresholds provided a more accurate stratification of patients into risk subgroups (Fig. 4A). When a commonly used binary criterion for biochemical response, such as the Paris-2 definition, was used instead of the Globe score, LSM thresholds continued to discriminate patients into significant risk groups regardless of therapeutic response (Fig. 4B). Finally, risk heat maps for poor clinical outcomes at 1, 3, 5, and 10 years of follow-up were derived from the continuous values of LSM and Globe score at baseline (Fig. 5). The calibration of the model was adequate with a standardized incidence ratio of observed *vs.* expected events of 1.15 (95% CI 0.92–1.45) (Table S7). The predicted probabilities of poor outcomes and 95% CIs can be calculated using a web tool available at the following internet address: https://triton.iplesp.upmc.fr/webcalculator/.

Discussion

In this study of >3,000 patients with PBC without prior cirrhosis-related complications followed-up for an average of 5 years over the past 15 years in tertiary centres from Europe, North America or the Middle East, elevated values of LSM (as assessed by VCTE) were unequivocally associated with increased risk of poor clinical outcomes, including death, LT or liver complications. This association was independent of baseline and temporal prognostic factors and scores, treatment type and response, and duration of follow-up. These results were replicated in an external validation cohort. The combination of LSM with established biochemical criteria of therapeutic response or prognostic scores improved prognostic prediction for patients. Therefore, these findings confirm that LSM could be useful to evaluate and monitor patients with PBC in routine care and therapeutic trials in combination with biochemical results.

The use of LSM in medical practice has developed considerably over the last 15 years.[15,34] Among the different techniques available, VCTE (FibroScan®) was the first to be marketed, has been the most extensively evaluated, and is currently one of the most popular and commonly used, especially in hepatology units.[14,35] Prior studies in patients with PBC have shown an association of LSM with histological fibrosis stage, degree of portal hypertension, and risk of death or need for LT.[16,17,19,20,36–38] However, these results are still supported by limited data and a few small-size single-centre studies. The present study is the largest ever study assessing the utility of LSM in PBC. It shows that between 5 kPa and 30 kPa, namely the range within which most patients with PBC are seen in routine, the risk of poor clinical outcomes increases as a monotonic function of LSM, a characteristic shared with serum levels of ALP and bilirubin.[7] This indicates that the lower the LSM, the higher the survival time without poor clinical outcomes. In addition, our results show that the prognostic ability of LSM is sustained in time and across specific high-risk groups such as inadequate responders or relatively young patients. Taken

Table 2. Time-dependent, multivariable-adjusted Cox regression analysis for poor clinical outcomes (derivation cohort, 2,740 patients, 5,324 LSMs).

Parameter	Estimate	Chi^2	HR (95% CI)	*p* value
LSM (kPa)	0.0394 ± 0.0068	33.60	1.040 (1.026 – 1.054)	<0.0001
ALP (xULN)	0.1996 ± 0.0461	18.77	1.221 (1.116 – 1.336)	<0.0001
Platelets (G/L)	-0.0059 ± 0.0016	14.71	0.994 (0.991 – 0.997)	0.0001
Age (year)	0.0313 ± 0.0093	11.42	1.032 (1.013 – 1.051)	0.0007
Albumin (g/L)	-0.0471 ± 0.0140	11.39	0.954 (0.928 – 0.980)	0.0007
T. bilirubin (μM/L)	0.1458 ± 0.0680	4.60	1.157 (1.013 – 1.322)	0.03
Male sex (yes)	0.3966 ± 0.2515	2.49	1.486 (0.908 – 2.433)	0.11
AST (xULN)	0.0316 ± 0.0955	0.11	1.032 (0.856 – 1.245)	0.74

Parameters are sorted by decreasing order of statistical significance. Poor clinical outcomes are defined by death, LT or liver complications.
ALP, alkaline phosphatase; AST, aspartate aminotransferase; LSM, liver stiffness measurement; LT, liver transplantation.

Table 3. Added value of LSM to prognostic scores, fibrosis scores, and biochemical response criteria in predicting poor clinical outcomes in PBC.

	Harrell's C-index (95% CI) unadjusted for LSM	Harrell's C-index (95% CI) adjusted for LSM	*p* value
Globe score	0.7977 (0.7603–0.8351)	0.8322 (0.7952–0.8692)	0.0004
UK-PBC score	0.7469 (0.7028–0.7910)	0.7956 (0.7527–0.8385)	<0.0001
MELD score	0.7774 (0.7347–0.8201)	0.8314 (0.7959–0.8669)	0.0334
APRI score	0.7767 (0.7401–0.8134)	0.8253 (0.7904–0.8602)	<0.0001
FIB-4 score	0.7885 (0.7528–0.8242)	0.8312 (0.7961–0.8663)	<0.0001
Globe resp. (no)	0.7168 (0.6811–0.7525)	0.8178 (0.7802–0.8554)	<0.0001
Toronto resp. (no)	0.6591 (0.6258–0.6924)	0.8073 (0.7712–0.8434)	<0.0001
Paris-2 resp. (no)	0.6553 (0.6234–0.6872)	0.8105 (0.7744–0.8466)	<0.0001

Harrell's C-index reflects the probability a randomly selected patient who experienced a poor clinical outcome (death, liver transplantation, or liver complications) had a higher risk score or poorer response than a patient who had not experienced the event. Comparison of Harrell's C-indices was performed using a non-parametric bootstrap approach with 500 replicates.

APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis-4; LSM, liver stiffness measurement; MELD, model for end-stage liver disease; PBC, primary biliary cholangitis.

together, these results strongly support the use of LSM by VCTE in the assessment and monitoring of PBC, both in routine care and clinical trials, in conjunction with established prognostic biochemical indices.

It has been shown that LSM values >9.6 kPa at baseline could distinguish patients with PBC who are at higher risk of clinical outcomes.[16] The 8 kPa and 15 kPa thresholds that were determined using an unbiased method in this study provide greater

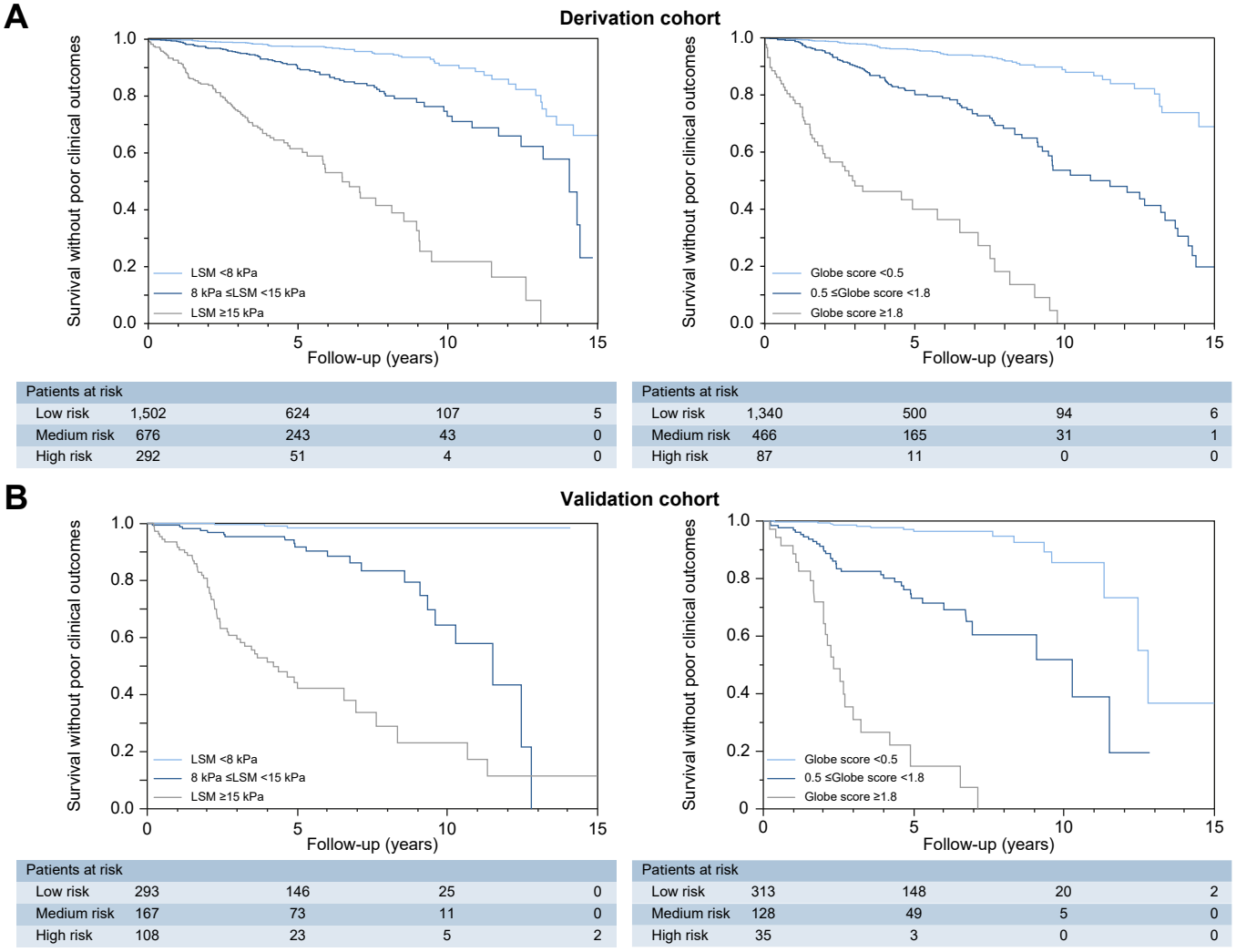


Fig. 3. Probability of survival without poor clinical outcomes according to LSM and Globe score risk groups at baseline. (A,B) Probability of survival without poor clinical outcomes of patients with primary biliary cholangitis as determined in the derivation (A) and validation (B) cohorts according to low- (light blue line), medium- (dark blue line), and high- (grey line) risk groups defined at baseline according to LSM (left diagrams) or Globe score (right diagrams) thresholds. The LSM and Globe score thresholds were determined using a grid search unbiased method. Survival probabilities were calculated using Kaplan-Meier estimates. LSM, liver stiffness measurement.

JOURNAL
OF HEPATOLOGY

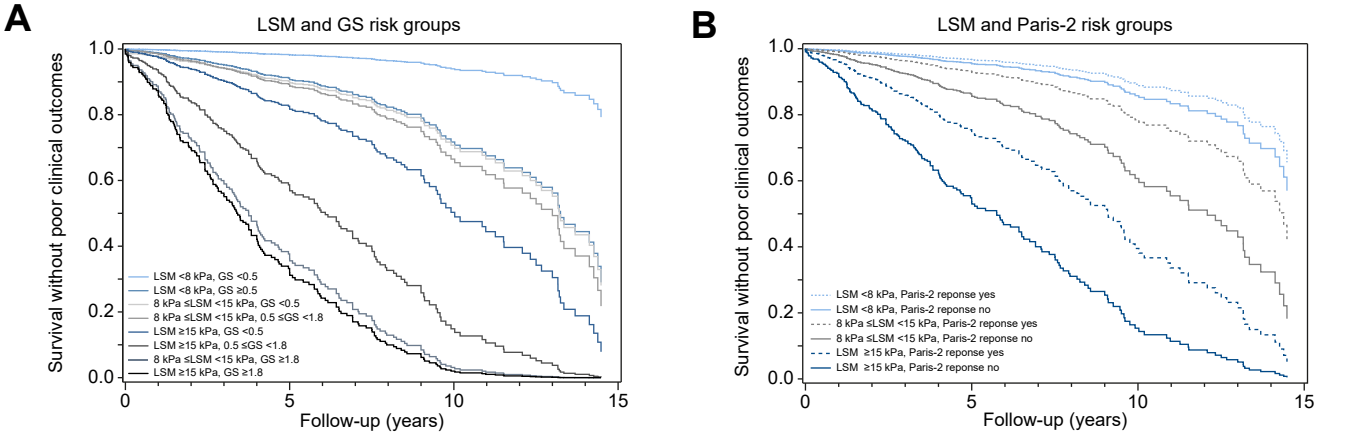


Fig. 4. Probability of survival without poor clinical outcomes according to LSM and biochemical (GS, Paris-2 criteria) combined risk groups at baseline. (A,B) Probability of survival without poor clinical outcomes of patients with primary biliary cholangitis as determined in the derivation cohort according to LSM and GS combined risk groups (A) or LSM and Paris-2 combined risk groups (B) assessed at baseline. Survival probabilities were calculated using Cox regression estimates. GS, Global score; LSM, liver stiffness measurement.

specificity for distinguishing between low-, medium- and high-risk patients. These cut-offs correspond virtually to the diagnostic thresholds for significant fibrosis and cirrhosis, respectively, which is clinically relevant and consistent with the established prognostic value of fibrosis stage in chronic liver diseases, especially PBC.[13] The 15 kPa threshold is also in line with the Baveno VII recommendations, which suggest using >15 kPa as the threshold above which compensated advanced chronic liver disease should be strongly suspected, regardless of aetiology.[23] Medium- and high-risk patients, who should be prioritised for clinical trials, represented approximately 40% of the whole cohort. These patients showed approximately 4-fold and 16-fold higher risk of poor clinical outcomes, respectively, compared to low-risk patients. While the 10-year rate of clinical events in low-risk patients did not exceed 20%, it varied between 20% and 50% in medium-risk patients, and between 50% and 90% in high-risk patients, particularly in relation to the biochemical response.

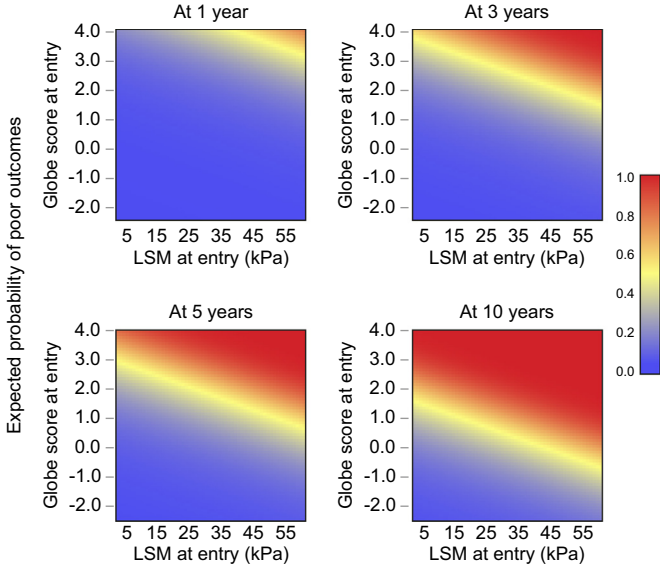


Fig. 5. Risk heat maps for poor clinical outcomes as derived at 1, 3, 5, and 10 years from the baseline values of LSM and Globe score. The expected probabilities of poor clinical outcomes, *i.e.* liver complications, liver transplantation, or death, were calculated at different time points (1, 3, 5, and 10 years) based on the baseline values of LSM and Globe score. A continuous colour scale (y-axis shown on the left) was used to visually represent the probability level. Probabilities and 95% CIs can be determined using a web calculator available at the following internet address: https://triton.iplesp.upmc.fr/webcalculator/. LSM, liver stiffness measurement.

Second-line therapy for PBC (obeticholic acid or fibrates) is currently recommended in patients with an inadequate biochemical response to UDCA, regardless of disease stage.[2–4] Over the last 10 years, several definitions of inadequate biochemical response have been proposed, of which the Paris-2, Toronto, and Globe score-based definitions are among the best known.[6,25,29] The present study shows that LSM provides prognostic information in PBC beyond biochemical response, regardless of the definition used. This should prompt future recommendations for PBC therapy to integrate LSM in decision-making algorithms. Indeed, even adequate responders to UDCA could be legitimate candidates for second-line therapy if a medium or high probability of clinical events is predicted by LSM, considering that any 10-year rate of clinical events above 20% in a middle-aged population is a poor outcome.

LSM was the variable most strongly associated with long-term outcomes in PBC in the present study, ahead of ALP and bilirubin levels, suggesting that it might be used in therapy trials as a surrogate endpoint for survival. This is supported by the results of the recent phase III BEZURSO (bezafibrate for PBC with inadequate response to UDCA) trial, a 2-year double-blind, randomized study in which LSM was shown to increase significantly in patients assigned to the placebo group, whereas it remained stable or even decreased in those assigned to the bezafibrate group.[9] This suggests 2 key points: i) progression of LSM is expected in UDCA-resistant patients; ii) a pharmacological treatment can slow, stop, or even reverse this course. Consequently, a primary composite endpoint defined by a clinically relevant biochemical response (ideally, normalisation of ALP and bilirubin) and no concomitant increase in LSM might appear to be a realistic target that could become the new standard for PBC trials in UDCA-resistant patients. However, further studies are needed to characterise the temporal evolution of LSM in different risk

groups of patients with PBC, and specifically the inter- and within-patient variability of this measure over time. These repeated measures analyses are currently underway.

There are some limitations to the present study. The derivation and validation cohorts were created consecutively from different periods and centres. This could explain the differences observed in the severity of the 2 cohorts at baseline, as well as in long-term clinical outcomes, despite the same selection and exclusion criteria used. Finding additional LSM data for validation was difficult because the derivation cohort was assembled from the main leading PBC expert centres equipped with FibroScan®. This may have introduced a selection bias in the validation cohort in that, especially in smaller, less PBC-focused centres, FibroScan® may be used primarily in patients with more advanced disease. The study was not conducted on an intent-to-predict basis and was limited to LSM results considered reliable and interpretable according to the manufacturer's recommendations and usual definition.[14,24] On the overall cohorts, 1,192 (13.9%) out of 8,515 LSMs were unreliable or failed, which is consistent with data from other large studies in other chronic liver diseases.[39,40] The rate of unreliable or failed results (~10%-15%) with this elastography technique (VCTE) should be kept in mind, especially if LSM is considered as an endpoint for assessing treatment strategies. The use and follow-up policies of LSM were not uniform across centres and were probably biased by the fact that clinicians are more likely to assess and repeat LSM in symptomatic patients and/or those with more advanced disease. However, we used a time-dependent multivariable-adjusted analysis in order to minimise the potential impact of population and follow-up heterogeneity between centres, and a large number of patients with low LSM at baseline was followed over time with repeated measurements.

In conclusion, in the largest ever study of LSM in patients with PBC, we show that LSM assessed by VCTE improves prediction of survival beyond biochemical response, established prognostic scores, and age categories, independent of time. A significant proportion (40%) of patients with PBC are at medium or high risk of poor clinical outcome according to LSM. This robust analysis suggests that LSM can reasonably be regarded as a useful prognostic marker and a potential surrogate endpoint in PBC clinical trials.

Abbreviations

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; HR, hazard ratio; LSM, liver stiffness measurement; LT, liver transplantation; MELD, model for end-stage liver disease; PBC, primary biliary cholangitis; ROC, receiver operating characteristic; UDCA, ursodeoxycholic acid; VCTE, vibration-controlled transient elastography.

Financial support

The authors received no financial support to produce this manuscript.

Conflict of interest

Dr. Corpechot reports receiving grants from Arrow and Intercept France, consulting fees from Intercept France, Inventiva Pharma, Cymabay and Genkyotex, and fees for teaching from Intercept France and GlaxoSmithKline France; Dr. Chazouillères, receiving grant support from Aptalis, fees for teaching from Mayoly Spindler, consulting fees from Genfit, and fees for teaching and consulting fees from Intercept; Dr. Schramm, receiving lecture fees from Falk Pharma; Dr. Dumortier, receiving consulting and teaching fees from Intercept France; Dr. Parés, receiving grant funding, speaking fees, and advisory board fees from Intercept, advisory board fees and speaking fees from Novartis, and speaking fees from CymaBay and Inova Diagnostics; Dr. Bruns reports receiving advisory board fees from Intercept, Grifols and Sobi and receiving speaking fees from Falk Foundation, Abbvie, CSL Behring, Intercept, Merck and Gilead. No other potential conflict of interest relevant to this article was reported.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Dr. C. Corpechot is acting as the submission's guarantor. CC, study co-designer, coordinating investigator, data acquisition, data analysis and interpretation, drafting manuscript; FC: statistical analysis, data interpretation, critical revision; VL, study co-designer, data acquisition, critical revision; Remaining authors: data acquisition, critical revision. All authors approved the final version of the manuscript.

Data availability statement

The dataset generated during this study is available from the corresponding author upon reasonable request.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2022.06.017.

References

Author names in bold designate shared co-first authorship

[1] Lleo A, Wang GQ, Gershwin ME, Hirschfield GM. Primary biliary cholangitis. Lancet 2020;396:1915–1926.
[2] EASL Clinical Practice Guidelines. The diagnosis and management of patients with primary biliary cholangitis. J Hepatol 2017;67:145–172.
[3] Hirschfield GM, Dyson JK, Alexander GJM, Chapman MH, Collier J, Hubscher S, et al. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. Gut 2018;67:1568–1594.
[4] 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.
[5] Harms MH, van Buuren HR, Corpechot C, Thorburn D, Janssen HLA, Lindor KD, et al. Ursodeoxycholic acid therapy and liver transplant-free survival in patients with primary biliary cholangitis. J Hepatol 2019;71:357–365.
[6] Corpechot C, Chazouilleres O, Poupon R. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. J Hepatol 2011;55:1361–1367.
[7] Lammers WJ, van Buuren HR, Hirschfield GM, Janssen HL, Invernizzi P, Mason AL, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study. Gastroenterology 2014;147:1338–1349.
[8] Nevens F, Andreone P, Mazzella G, Strasser SI, Bowlus C, Invernizzi P, et al. A placebo-controlled trial of obeticholic acid in primary biliary cholangitis. N Engl J Med 2016;375:631–643.
[9] Corpechot C, Chazouilleres O, Rousseau A, Le Gruyer A, Habersetzer F, Mathurin P, et al. A placebo-controlled trial of bezafibrate in primary biliary cholangitis. N Engl J Med 2018;378:2171–2181.
[10] Silveira MG, Brunt EM, Heathcote J, Gores GJ, Lindor KD, Mayo MJ. American Association for the Study of Liver Diseases endpoints conference: design and endpoints for clinical trials in primary biliary cirrhosis. Hepatology 2010;52:349–359.

[11] Harms MH, van Buuren HR, van der Meer AJ. Improving prognosis in primary biliary cholangitis - therapeutic options and strategy. Best Pract Res Clin Gastroenterol 2018;34–35:85–94.
[12] Corpechot C. Clinical trials in PBC going forward. Semin Liver Dis 2019;39:e1–e6.
[13] 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.
[14] Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705–1713.
[15] Friedrich-Rust M, Poynard T, Castera L. Critical comparison of elastography methods to assess chronic liver disease. Nat Rev Gastroenterol Hepatol 2016;13:402–411.
[16] Corpechot C, Carrat F, Poujol-Robert A, Gaouar F, Wendum D, Chazouilleres O, et al. Noninvasive elastography-based assessment of liver fibrosis progression and prognosis in primary biliary cirrhosis. Hepatology 2012;56:198–208.
[17] **Cristoferi L**, **Calvaruso V**, Overi D, Vigano M, Rigamonti C, Degasperi E, et al. Accuracy of Transient Elastography in assessing fibrosis at diagnosis in naive patients with Primary Biliary Cholangitis: a dual cut-off approach. Hepatology 2021;74:1496–1508.
[18] Rigamonti C, De Benedittis C, Azzolina D, Costanzo M, Pirovano A, Croce A, et al. Liver stiffness dynamics predict survival in patients with primary biliary cholangitis treated with ursodeoxycholic acid. Hepatology 2019;70:803A.
[19] Moctezuma-Velazquez C, Saffioti F, Tasayco-Huaman S, Casu S, Mason A, Roccarina D, et al. Non-invasive prediction of high-risk varices in patients with primary biliary cholangitis and primary sclerosing cholangitis. Am J Gastroenterol 2019;114:446–452.
[20] **Osman KT**, **Maselli DB**, Idilman IS, Rowan DJ, Viehman JK, Harmsen WS, et al. Liver stiffness measured by either magnetic resonance or transient elastography is associated with liver fibrosis and is an independent predictor of outcomes among patients with primary biliary cholangitis. J Clin Gastroenterol 2021;55:449–457.
[21] European Association for the Study of the Liver. Electronic address eee, Clinical Practice Guideline P, Chair, representative EGB, Panel m. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol 2021;75:659–689.
[22] Corpechot C, Heurgue A, Tanne F, Potier P, Hanslik B, Decraecker M, et al. Non-invasive diagnosis and follow-up of primary biliary cholangitis. Clin Res Hepatol Gastroenterol 2022;46:101770.
[23] de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno VIIF. Baveno VII - renewing consensus in portal hypertension. J Hepatol 2022;76:959–974.
[24] 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;57:1182–1191.
[25] Lammers WJ, Hirschfield GM, Corpechot C, Nevens F, Lindor KD, Janssen HL, 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.
[26] 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.
[27] Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology 2000;31:864–871.
[28] 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.
[29] Kumagi T, Guindi M, Fischer SE, Arenovich T, Abdalian R, Coltescu C, et al. Baseline ductopenia and treatment response predict long-term histological progression in primary biliary cirrhosis. Am J Gastroenterol 2010;105:2186–2194.
[30] Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005;61:92–105.
[31] Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. Stat Med 2015;34:685–703.
[32] Lerman PM. Fitting segmented regression models by grid search. Appl Stat 1980;29:77–84.
[33] Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res 2016;25:1692–1706.
[34] Mueller S. Liver elastography. In: Mueller S, editor. Clinical Use and Interpretation. Cham: Springer; 2020. p. XIX, 737.
[35] 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;134:960–974.
[36] Corpechot C, El Naggar A, Poujol-Robert A, Ziol M, Wendum D, Chazouilleres O, et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. Hepatology 2006;43:1118–1124.
[37] 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.
[38] Zachou K, Lygoura V, Arvaniti P, Giannoulis G, Gatselis NK, Koukoulis GK, et al. FibroMeter scores for the assessment of liver fibrosis in patients with autoimmune liver diseases. Ann Hepatol 2021;22:100285.
[39] Castera L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. Hepatology 2010;51:828–835.
[40] Cassinotto C, Boursier J, de Ledinghen V, Lebigot J, Lapuyade B, Cales P, et al. Liver stiffness in nonalcoholic fatty liver disease: a comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. Hepatology 2016;63:1817–1827.