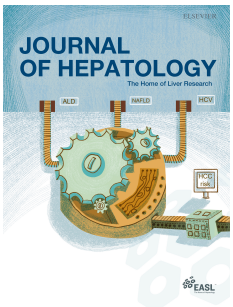
Journal Pre-proof



Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease

Margarita Papatheodoridi, Jean Baptiste Hiriart, Monica Lupsor-Platon, Fabrizio Bronte, Jerome Boursier, Omar Elshaarawy, Fabio Marra, Maja Thiele, Georgios Markakis, Audrey Payance, Edgar Brodkin, Laurent Castera, George Papatheodoridis, Aleksander Krag, Umberto Arena, Sebastian Mueller, Paul Cales, Vincenza Calvaruso, Victor de Ledinghen, Massimo Pinzani, Emmanuel A. Tsochatzis

PII: S0168-8278(20)33838-1

DOI: https://doi.org/10.1016/j.jhep.2020.11.050

Reference: JHEPAT 8057

To appear in: *Journal of Hepatology*

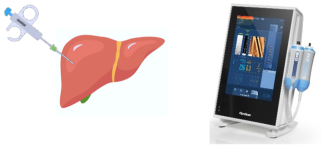
Received Date: 1 June 2020

Revised Date: 19 November 2020

Accepted Date: 22 November 2020

Please cite this article as: Papatheodoridi M, Hiriart JB, Lupsor-Platon M, Bronte F, Boursier J, Elshaarawy O, Marra F, Thiele M, Markakis G, Payance A, Brodkin E, Castera L, Papatheodoridis G, Krag A, Arena U, Mueller S, Cales P, Calvaruso V, de Ledinghen V, Pinzani M, Tsochatzis EA, Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease, *Journal of Hepatology* (2021), doi: https://doi.org/10.1016/j.jhep.2020.11.050.

Over 5,600 patients with liver disease from 10 European centres



Underwent liver biopsy and transient elastography within 6 months

To evaluate **compensated advanced chronic liver disease (cACLD),** defined as fibrosis stage ≥ F3

Validation of the Baveno VI LS dual cut-off for cACLD

Excluding cACLD using low cut-off **LS <10 kPa**

Sensitivity 75%

Diagnosing cACLD using high cut-off **LS >15 kPa**

Specificity 96%

Optimal LS dual cut-off for cACLD

Excluding cACLD using low cut-off* **LS <8 kPa**

Sensitivity 91%

**Low cut-off LS <7 kPa for patients with viral hepatitis.*

Diagnosing cACLD using high cut-off **LS >12 kPa**

Specificity 92%

For **unclassified patients**, cACLD risk calculated by:

Pred. Probability of cACLD= e^a/ (1+ e^a)

with a=-5.264 +0.181***Age** - 0.62 for **male sex** - 0.081* **BMI** +0.415* **LS** + 0.126* **ALT**+ 0.257* **AST** +0.129* **GGT** -0.289* **Platelets** +0.335 for **Type II Diabetes**

Title: **Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease**

Margarita Papatheodoridi[1], Jean Baptiste Hiriart[2], Monica Lupsor-Platon[3], Fabrizio Bronte[4], Jerome Boursier[5], Omar Elshaarawy[6], Fabio Marra[7], Maja Thiele[8], Georgios Markakis[9], Audrey Payance[10], Edgar Brodkin[1], Laurent Castera[10], George Papatheodoridis[9], Aleksander Krag[8], Umberto Arena[7], Sebastian Mueller[6], Paul Cales[5], Vincenza Calvaruso[4], Victor de Ledinghen[2], Massimo Pinzani[1*], Emmanuel A. Tsochatzis[1*]

*These two authors had equal contribution to the manuscript and are joint senior authors.

1. UCL Institute for Liver and Digestive Health, Royal Free Hospital and UCL, London, UK; 2. INSERM U1053, Bordeaux University, Bordeaux, France; 3. Department of Medical Imaging, Regional Institute of Gastroenterology and Hepatology, Prof. Dr. Octavian Fodor", University of Medicine and Pharmacy "Iuliu Hatieganu" Cluj-Napoca, Romania; 4. Gastroenterology and Hepatology Unit, Department of Health Promotion Sciences Maternal and Infantile Care, Internal Medicine and Medical Specialities, PROMISE. University of Palermo; 5. Liver-Gastroenterology Department, University Hospital, Angers, France; 6. Center for Alcohol Research, University of Heidelberg, Heidelberg, Germany; 7. Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy; 8. Department of Gastroenterology and Hepatology, Odense University Hospital, Odense, Denmark; 9. Academic Department of Gastroenterology, Medical School of National and Kapodistrian University of Athens, Laiko General Hospital, Athens,

Greece; 10. Service d'Hépatologie, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, Clichy, France

Corresponding author:

Prof. Emmanuel A. Tsochatzis
Sheila Sherlock Liver Unit and UCL Institute for Liver and Digestive Health,
Royal Free Hospital and UCL, London, UK.
Email: e.tsochatzis@ucl.ac.uk

Keywords

Cirrhosis; Fibroscan; FIB-4; NAFLD; Alcoholic liver disease; viral hepatitis; portal hypertension

Word count: 5979

Number of tables and figures: 2 tables and 3 figures

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request

Conflicts of Interest Statement:

Victor de Ledinghen reports consultancy for Echosens and SuperSonic Imagine. All other authors have nothing to disclose.

Financial Support Statement:

The study in Odense University Hospital received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 668031, the Challenge Grant "MicrobLiver" number NNF15OC0016692 from the Novo Nordisk Foundation, Innovation Fund Denmark and the free research funds for Odense University Hospital and Region of Southern Denmark.

Author Contributions:

Margarita Papatheodoridi: Data analysis, study design, primary authorship

Jean Baptiste Hiriart: Data collection

Monica Lupsor-Platon: Data collection, revision for important intellectual concepts.

Fabrizio Bronte: Data collection

Jerome Boursier: Data collection, revision for important intellectual concepts.

Omar Elshaarawy: Data collection

Fabio Marra: Data collection, revision for important intellectual concepts.

Maja Thiele: Data collection, revision for important intellectual concepts.

Georgios Markakis: Data collection

Audrey Payance: Data collection

Edgar Brodkin: Data collection

Laurent Castera: Data collection, revision for important intellectual concepts.

George Papatheodoridis: Data collection, revision for important intellectual concepts.

Aleksander Krag: Data collection, revision for important intellectual concepts.

Umberto Arena: Data collection, revision for important intellectual concepts.

Sebastian Mueller: Data collection, revision for important intellectual concepts.

Paul Cales: Data collection, revision for important intellectual concepts.

Vincenza Calvaruso: Data collection, revision for important intellectual concepts.

Victor de Ledinghen: Data collection, revision for important intellectual concepts.

Massimo Pinzani; Study concept and design, revision for important intellectual concepts

Emmanuel Tsochatzis: Data analysis, study concept and design, revision for important intellectual concepts

Abbreviations:

ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUC, area under the curve; BMI, body mass index; cACLD, compensated advanced chronic liver disease; CI, confidence interval; CSPH, clinically significant portal hypertension; GGT, gamma-glutamyl transferase; HVPG, hepatic venous pressure gradient; IQR, interquartile range; LS, liver stiffness; ROC, receiver operating curve; TE, transient elastography

ABSTRACT

Background: The Baveno VI consensus proposed a dual liver stiffness (LS) by transient elastography threshold of <10 and >15 kPa for excluding and diagnosing compensated advanced chronic liver disease (cACLD), in the absence of other clinical signs. We validated these criteria in a real-world multicentre study.

Methods: We included 5648 patients (mean age 51±13 years, 53% males) from ten European liver centres who had a liver biopsy and LS measurement within 6 months. We included patients with chronic hepatitis C (CHC, n=2913, 52%), non-alcoholic fatty liver disease (NAFLD, n=1073, 19%), alcohol-related liver disease (ALD, n=946, 17%) or chronic hepatitis B (CHB, n=716, 13%). cACLD was defined as fibrosis stage ≥F3.

Results: Overall, 3606 (66%) and 987 (18%) patients had LS <10 and >15 kPa, respectively, while cACLD was histologically confirmed in 1772 (31%) patients. The cut-offs of <10 and >15 kPa showed 75% sensitivity and 96% specificity to exclude and diagnose cACLD, respectively. Examining the ROC curve, a more optimal dual cut-off at <7 and >12 kPa, with 91% sensitivity and 92% specificity for excluding and diagnosing cACLD (AUC=0.87, 95%CI:0.86-0.88, P<0.001) was derived. Specifically for ALD and NAFLD, a low cut-off of 8 kPa can be used (sensitivity=93%). For the unclassified patients, we derived a risk model based on common patient characteristics with better discrimination than LS alone (AUC=0.74 vs. 0.69, P<0.001).

Conclusions: Instead of the Baveno VI proposed <10 and >15 kPa dual cut-offs, we found that the <8 kPa (or <7 kPa for viral hepatitis) and >12 kPa dual cut-offs have better diagnostic accuracy in cACLD.

LAY SUMMARY

Compensated advanced chronic liver disease (cACLD) is a term introduced by the Baveno VI consensus in 2015 to describe the spectrum of advanced fibrosis and cirrhosis in asymptomatic patients with chronic liver disease. The same consensus also proposed that cACLD could be ruled out by transient elastography measurements with liver stiffness (LS) values <10 kPa or diagnosed by LS values >15 kPa. We assessed the performance of these LS cut-offs in over 5,000 patients and found that only 75% of the patients could be excluded with the low cut-off of <10 kPa, while 95% of them would be correctly diagnosed using the >15 kPa cut-off. To improve these rates, we propose that the cut-offs of 7 and 12 kPa can correctly exclude or diagnose ~90% of the patients. In patients with non-alcoholic or alcohol related fatty liver disease, 8 kPa can be safely used as a low cut-off. The findings of this study have important implications for patients with chronic liver disease, as the use of liver biopsy can be safely reduced.

INTRODUCTION

The Baveno VI consensus introduced the term “compensated advanced chronic liver disease (cACLD)” to describe the spectrum of advanced fibrosis and cirrhosis in asymptomatic patients, who are at risk of developing clinically significant portal hypertension (CSPH)(1). Despite the absence of symptoms and/or clinical signs, patients with cACLD are at high risk of future liver-related morbidity and mortality(1). Consequently, they may have improved outcome of their liver disease with early diagnosis and subsequent prompt interventions(2).

In order to diagnose cACLD in the large group of patients with asymptomatic liver disease, the Baveno VI consensus suggested that transient elastography (TE) is sufficient to suspect cACLD, since it has good diagnostic accuracy and is also a safe, painless, fast and relatively low-cost non-invasive diagnostic method(1). Liver stiffness (LS) values <10 kPa were proposed as a safe cut-off for excluding the presence of cACLD and LS values >15 kPa as highly suggestive of cACLD.

The dual cut-off approach for TE was seldom used until the Baveno VI recommendations. TE has been mostly studied as a single cut-off diagnostic test, with optimal cut-offs derived in most studies from post-hoc analyses(3). Therefore, reported cut-offs for the diagnosis of specific fibrosis stages vary, sometimes significantly, amongst studies(4,5). Moreover, it is unlikely that the same cut-offs for a specific fibrosis stage apply to different liver disease aetiologies(4,6). The use of dual cut-offs may overcome both these limitations and introduce uniformity in the diagnosis of chronic liver disease. The Baveno VI criteria on screening for varices needing treatment including LS values <20 kPa and platelet count >150×10^9 cells/L(1) have already been validated in several small to medium independent

cohorts(7–10) and in a recent large meta-analysis of 30 studies(11). However, the criteria to rule in and out cACLD have not been externally validated in a real-world population until now.

This study aimed to assess the diagnostic accuracy of LS dual cut-off (<10 and >15 kPa) as a standalone test for the exclusion and diagnosis of cACLD (defined as the presence of advanced fibrosis or ≥F3), as proposed by the Baveno VI criteria(1), in a multicentre validation study of real-world data. Secondary aims were to explore optimal alternative rule in/rule out cut-offs with a target specificity and sensitivity of ≥90% and to derive a risk model for predicting cACLD in unclassified patients.

PATIENTS AND METHODS

Study population

This study included 5,648 adult patients followed in 10 liver centres in Europe [Bordeaux n=1335 (24%), Cluj n=1180 (21%), Palermo n=808 (14%), Angers, n=698 (12%), Heidelberg 450 (8%), Firenze n=334 (6%), Odense, n=316 (6%), London n=303 (5%), Athens n=154 (3%), Beaujon n=70 (1%)]. Study population included patients with all fibrosis stages and no previous decompensation who had chronic hepatitis B (CHB), chronic hepatitis C (CHC), non-alcoholic fatty liver disease (NAFLD) or alcohol-related liver disease (ALD). Patients who had co-infection with hepatitis B and C (n=21) were included in the group of CHC patients for the analysis of this study.

Patients were included in the study if (i) they had undergone a liver biopsy and a TE measurement with an interval of ≤6 months, (ii) the M probe was used for the TE measurement and (iii) they were fasting >4 hours before TE was performed.

Exclusion criteria for the study were: a) absence of liver biopsy and/or TE measurement within 6 months, b) TE measurement with XL probe, c) Meal ingestion <2 hours prior to the TE measurement, d) Patients with HCV and previous SVR, e) Patients with HBV under stable longstanding antiviral treatment.

Data on basic demographic and laboratory characteristics were collected for all patients. All fibrosis staging in patients with viral hepatitis was adapted to the METAVIR(12) staging system, to maintain consistency for the analysis (**Suppl. Table 1**). Fibrosis was staged according to the Brunt staging system for ALD and NAFLD(13). cACLD was defined as ≥F3 (in Brunt for ALD/NAFLD or METAVIR in viral hepatitis) and cirrhosis as F4 fibrosis stage.

All TE examinations were performed by experienced operators with more than 500 procedures. For TE measurements, median LS values, as well as the number of valid measurements and interquartile range (IQR) were recorded. Ten valid LS measurements were obtained and reported as a median value in kPa. Adequacy of the measurement was assessed by the TE device and the interquartile range/median ratio (IQR/M) was used for the assessment of TE reliability: “very reliable” (IQR/M≤0.10), “reliable” (0.10<IQR/M≤0.30, or IQR/M>0.30 with LS median <7.1 kPa), and “poorly reliable” (IQR/M>0.30 with LS median ≥7.1 kPa)(14).

The study was approved by the local ethics committee in each of the participating centres.

Study endpoints

The primary endpoint was to validate the diagnostic accuracy of the LS cut-off values of <10 and >15 kPa for ruling out and diagnosing cACLD (≥F3 fibrosis stage), respectively(1).

The secondary study endpoints were to:

1. Explore optimal alternative rule in/rule out LS thresholds with specificity and sensitivity of >90%, respectively.
2. Evaluate risk factors and derive a diagnostic model for cACLD for unclassified patients with indeterminate LS results.

We performed sensitivity analyses and explored the diagnostic accuracy of established and alternative LS cut-offs according to:

a. Reliability of TE results.

b. Aetiology of liver disease.

c. Biopsy sample length (<15 and ≥15 mm) and portal tract number (≤10 and >10).

d. Body mass index (BMI) (<30 and ≥30 kg/m^2).

e. Alanine aminotransferase (ALT) levels (<100, ≥100 and <200, and ≥200 IU/L).

f. Participating centre.

Finally, we explored independent variables associated with the discrepancies between LS and histology.

Statistical analysis

Data were analyzed either using the SPSS statistical package (IBM, Chicago, IL, 2019) or the R platform with the packages ggplot2(15), ROCR(16) and rms(17). Parametric and non-parametric quantitative variables were presented by their mean values ± standard deviation (SD) or median values [IQR], respectively. Comparisons between two patient groups were performed by the t-test or non-parametric Mann-Whitney test. The Kolmogorov-Smirnov goodness-of-fit test was used to assess distribution of each variable. Categorical variables were summarized as frequencies and percentages. The corrected chi-squared or two-sided Fisher's exact test was used to test for association between two categorical variables. Missing values in the sensitivity analyses were handled with case-wise deletion and all analyses were based on patients with available data.

Sensitivity (95% confidence intervals [CI]) and specificity (95%CI) levels of the LS cut-offs were compared as percentages (%) using the "N-1" Chi-squared test as recommended by Campbell(18) and Richardson(19). Their 95%CI was calculated according to Altman(20). Positive predictive (PPV) and negative predictive values

(NPV) of the LS cut-offs for the identification of the true positive (cACLD histological diagnosis) and true negative (absence of cACLD histological diagnosis) were calculated, respectively.

Optimal LS thresholds and their accuracy performance was evaluated in a receiver-operating characteristics (ROC) analysis. For the unclassified patients, univariate and multivariate binary logistic regression analyses were performed using patient characteristics in order to identify risk factors for cACLD. For the regression analysis, continuous variables were grouped in 5 quintiles to avoid predictors on different scales. Risk models using combination of the identified predictive factors were constructed and the predictive probability >0.5 or ≤0.5 was classified as positive or negative for cACLD, respectively. Different combinations of the predictive factors were tested, but the model including all independent factors had the best discriminatory ability and is presented in this paper. This model was internally validated and corrected for optimism with bootstrapping(21). To this end, 200 test data sets of the same size as the analysis dataset were created by random selection with replacement from the original dataset. To examine the performance of the risk prediction model, we assessed the discrimination and calibration of the model(22) and a nomogram was constructed based on the corrected logistic regression.

Finally, correctly classified patients using the LS dual cut-off were considered those where cACLD was correctly excluded or diagnosed by the low or high LS cut-off, respectively. For the calculation of the correct classification rate using the LS dual cut-offs and the risk model, the number of unclassified patients who were correctly predicted using the model was added to the rate of correctly classified patients using the LS cut-offs.

RESULTS

Patient characteristics

Our study included 5648 patients from ten centres; mean age was 51±13 years and 3016 (53%) were males. Main patient characteristics are presented in **Table 1.** The most common cause of liver disease was CHC (52%) followed by NAFLD (19%), ALD (17%) and CHB (13%). The majority (66%) of the patients with valid TE measurements (n=5483) had LS <10 kPa, while 18% of patients had LS >15 kPa. Furthermore, 7% of the TE results were evaluated as "poorly reliable" and 27% were considered as "very reliable". Fibrosis stages F0, F1, F2, F3 and F4 (METAVIR or Brunt) were present in 12%, 29%, 28%, 16% and 15% of patients, respectively. In total, 1772 (31%) of 5648 cases had histologically confirmed advanced fibrosis ≥F3, representing cACLD. Presence of portal hypertension had excellent concordance with presence of cACLD **(Suppl Index 1).** Median LS values differed significantly among patients with different fibrosis stages (**Suppl. Figure 1)**.

Patient characteristics differed significantly among patients with different aetiologies of liver disease. Patients with CHB were younger with mean age 44±14 years (P<0.001) and had the highest proportion of males (66%, P<0.001, compared to patients of all other liver disease aetiologies). On the other hand, NAFLD patients had the highest mean BMI (31±5 kg/m^2) and more frequently type II diabetes (43%) (always P<0.001, compared to patients with all other liver disease causes). In addition, NAFLD patients had the highest rate of poorly reliable TE results (9%) and the lowest rate of very reliable results (23%); this could be potentially explained by the higher median BMI and the fact that TE was performed with M-probe only. Finally, ALD patients had more often LS values >15 kPa (32%) and more frequently cACLD (38%) (**Table 1).**

Performance of the BAVENO VI proposed LS cut-offs for cACLD

The performance of the proposed LS cut-offs at <10 and >15 kPa was validated in 5483 patients (**Suppl. Table 2)**. A total of 165 patients did not have valid Fibroscan measurements and were excluded from the analysis. Setting the low LS cut-off at 10 kPa resulted in a sensitivity of 75% and a NPV of 88%, while setting the high cut-off at 15 kPa offered a specificity of 96% and PPV of 83% for diagnosing cACLD. The reliability of TE measurements affected the specificity of LS measurements; "poorly reliable" TE examinations were associated with >10% decrease in specificity (85% vs. 96%) and PPV (71% vs. 85%) as well as 3-fold lower positive likelihood ratio (LR+) (3.9 vs. 12.5), in comparison with "reliable" and "very reliable" TE measurements (always P<0.001). Patients with "poorly reliable" measurements had significantly higher prevalence of cACLD and significantly higher TE measurements (**Suppl. Material**). There was no significant difference of diagnostic accuracy between "reliable" and "very reliable" TE measurements. (**Suppl. Table 2, Figure 1**).

Among patient groups with different liver disease aetiology, the sensitivity of the cut-off <10 kPa to rule out cACLD was lower in patients with CHB vs. CHC [60% (52-66%) vs. 72% (69-75%), P<0.001) and highest for NAFLD [79% (74-84%)] compared to any other patient group (always P<0.001). The specificity of the high cut-off >15 kPa was 96% (96-97%) in CHC and 98% (96-99%) in CHB (P=0.113), while it was lower in NAFLD [94% (92-96%), P<0.001 compared to CHB or CHC) and lowest in ALD [92% (92-94%), P<0.001 compared to any other patient group] (**Suppl. Table 3, Figure 1).**

Similarly, the specificity of the high cut-off was significantly higher in patients with BMI <30 than ≥30 kg/m^2 [97% (94-100%) vs. 91% (84-98%)] as well as in patients with ALT levels <100 than ≥100 and <200 [97% (93-100%) vs. 92% (85-100%)] or than ≥200 IU/L [92% (85-100%) vs. 87% (72-100%)] (for the above comparisons all P-values <0.001). The specificity was also significantly higher in patients with liver biopsy sample length >15 mm. Conversely, the sensitivity of the low cut-offs was higher in patients with BMI≥30 kg/m^2 and with ALT levels ≥100. All comparisons between groups are shown in **Suppl. Table 4**.

Evaluation of optimal LS cut-offs to exclude or diagnose cACLD

Since the low cut-off of <10 kPa had moderate sensitivity and the high cut-off of >15 kPa unnecessarily high specificity, we evaluated alternative optimal dual cut-offs with ≥90% sensitivity and specificity values. Examination of the coordinate points of ROC curve showed that the cut-offs of <7 kPa and >12 kPa offered overall better diagnostic accuracy: sensitivity 91% and NPV 94% at <7 kPa and specificity 92% and PPV 78% at >12 kPa (AUC: 0.87, 95%CI: 0.86-0.88, P<0.001) (**Suppl. Figure 2**).

Among patients with different liver disease aetiology, the cut-off of >12 kPa had significantly higher specificity in patients with viral hepatitis, either CHC or CHB, [CHC vs. CHB: 94% (92-95%) vs. 94% (92-96%), P=0.839)] compared to patients with NAFLD [88% (86-91%), P<0.001] or those with ALD [89% (86-91%), P<0.001]. In contrast, patients with NAFLD and ALD were more accurately ruled out from having cACLD using the <7 kPa cut-off [NAFLD: 96% (95%CI 93-98%) vs. ALD: 96% (95%CI 94-98%), P=0.438] in comparison to patients with CHC [89% (95%CI 87-

91%), P<0.001] or CHB [83% (95%CI 77-88%) P<0.001] (**Suppl. Table 6 Figure 2)**. The cut-off of <8 kPa had an overall sensitivity of 86% (CHC 85%, CHB 73%, ALD 94%, NAFLD 93%), therefore it could be potentially used for ruling out cACLD in ALD or NAFLD but not in CHC or CHB.

Diagnostic accuracy performance of the new dual cut-offs of <7 and >12 kPa for different patient subgroups are presented in detail in **Suppl. Tables 6 and 7**. Sensitivity analysis of the performance of LS cut-offs for cACLD per participating centre of the study is shown in **Suppl. Table 8** and factors associated with discrepancies between histological findings and LS measurements are presented in **Suppl. Table 9**.

Diagnostic accuracy of the Baveno VI consensus cut-offs in cirrhosis

The diagnostic accuracy of LS measurements for diagnosing cirrhosis was excellent (AUC: 0.92 (95%CI: 0.91-0.93, P<001). The sensitivity/NPV was 92%/98%, 96%/99% and 98%/99% for the <10, <8 and <7 kPa low cut cut-offs, while the specificity/PPV was 92%/62% and 87%/52% for the >15 and >12 kPa high cut offs, respectively **(Suppl. Table 10)**. Diagnostic accuracies across aetiologies are also shown in **Suppl. Table 10**.

Evaluating cACLD in unclassified patients with LS values 7-12 kPa

The use of dual cut-offs to predict cACLD generates inevitably a group of unclassified patients. Applying the Baveno VI[1] recommended cut-offs resulted in 891 (16%) unclassified patients, while applying the new cut-offs of <7 and >12 kPa resulted in a larger number of unclassified patients (n=1797, 32%), which could be

reduced to 1647 patients (30%), if the cut-off <8 kPa was used for patients with NAFLD/ALD. In order to diagnose cACLD in the 1797 unclassified patients, we initially examined the diagnostic accuracy of the FIB-4 score. However, FIB-4 had poor discriminative ability (AUC:0.65), whereas its established cut-off points of 1.3 and 2.67 had sensitivity of 77% and specificity of 87%, respectively.

Therefore, we identified risk factors of cACLD, using a binary logistic regression, with the aim of constructing a new specific risk model for unclassified patients. Accordingly, in such unclassified patients with available clinical data (n=1097), cACLD risk was independently associated with age (adjusted OR per quintile: 1.19, 95%CI: 1.07-1.32, P=0.005), male sex (adjusted OR: 1.44, 95%CI: 1.09-1.90, P=0.020), BMI (adjusted OR per quintile: 0.89, 95%CI: 0.81-0.99, P=0.045), LS values (adjusted OR per kPa from 7-12kPa: 1.51, 95%CI: 1.38-1.66, P=0.005), AST (adjusted OR per quintile: 1.36, 95%CI: 1.15-1.59, P=0.005), platelet count (adjusted OR per quintile: 0.82, 95%CI: 0.74-0.90, P=0.005) and type II diabetes (adjusted OR: 1.47, 95%CI: 1.05-2.06, P=0.020) but not with ALT or GGT serum levels (**Table 2**).

Using the above predictive factors, we constructed a risk model for the unclassified patients. The risk model equation containing all variables of the model from the unadjusted logistic regression is presented in the **Supplementary Index 2** and the risk model equation using the adjusted coefficients from the multivariable logistic regression is the following:

$$\text{Pred. Probability of cACLD} = e^{a}/(1+ e^{a})$$

with a=-5.264 +0.171*age +0.364 for male sex -0.113*BMI +0.415*LS +0.304*AST -0.198*platelets +0.384 for type II diabetes

(age in years per quintile, BMI in kg/m^2 per quintile, LS in kPa per quintile, ALT in IU/L per quintile, platelets per mm^3 per quintile). Quintiles are described in the footnote of **Table 2**.

The discrimination of the risk model was better than using LS measurements alone (AUC: 0.74, 95%CI: 0.71-0.77 vs. 0.69, 95%CI: 0.65-0.72, P<0.001) in the unclassified patients with LS values between 7 and 12 kPa (**Suppl. Figure 3**). Of note, the discriminative ability of this model, based on the AUC, did not differ significantly between patient groups with different liver disease aetiology (**Suppl. Figure 4**). **Suppl. Figure 5** shows the representation based on LS values, calibration plot and nomogram of the risk model.

In total, using the new LS cut-offs of <7/<8 (for patients with viral hepatitis/ NAFLD or ALD) and >12 kPa and then the risk model for the unclassified patients, a classification rate of 84% could be reached, instead of 73% that would be achieved with the original Baveno VI[1] criteria. The classification rates with the Baveno VI criteria and the new proposed algorithm are depicted in **Figure 3**.

DISCUSSION

This is the first large study to validate the LS thresholds recommended by the Baveno VI consensus(1) for diagnosing or ruling out cACLD in asymptomatic patients with chronic liver disease. We included over 5,600 patients from 10 centres across Europe. Our study showed that the proposed cut-offs of <10 and >15 kPa have moderate sensitivity (75%) and very high specificity (96%) for ruling out and diagnosing cACLD respectively. Thus, we propose the adoption of lower dual cut-offs at <7 and >12 kPa, which offer optimal sensitivity for ruling out and specificity for diagnosing cACLD of 91% and 92%, respectively. However, the inherent weakness of the dual cut-off principle, which is the unclassified patient subgroup left with indeterminate LS values, remained and increased when applying the new cut-off points (although the misclassification rate was substantially lower). To counter this limitation, we may first increase the low LS cut-off to 8 kPa for ruling out cACLD specifically for patients with ALD and NAFLD. Moreover, using readily available variables, we constructed a risk model that could predict the presence of cACLD in >70% of the unclassified patients. Eventually, we reached a correct classification rate of 84%, in comparison to 73% achieved by the original Baveno VI proposed LS thresholds(1), with a significantly lower false negative rate. The original Baveno VI thresholds can be used for ruling out or diagnosing cirrhosis (rather than cACLD), where the sensitivity and specificity is above 90%. It is important to underline that these findings apply to untreated patients with viral hepatitis, and not to patients with CHC who have achieved sustained virological response or to CHB patients on long term antiviral treatment.

Although considerable research efforts have been focused on excluding advanced fibrosis with the use of non-invasive markers, including TE(3,23), there is no real

world data to validate the specific LS dual cut-offs for excluding or diagnosing cACLD as suggested by the Baveno VI[1] consensus. The dual cut-off of <10 and >15 kPa was recommended after reviewing the available literature at that time and might not have been optimal. Numerous studies proposing LS cut-offs for advanced fibrosis have yielded a wide range of thresholds with varied accuracy performance according to liver disease aetiology(4,24–26), partially due to spectrum bias. Additionally, a large meta-analysis of 10 studies including 760 ALD patients with fibrosis stage F3 or worse, which represents cACLD according to the new definition, reported that cACLD could be diagnosed by LS values between 8 to 17 kPa(5). However, considerable debate exists on whether one single LS threshold can accurately diagnose advanced fibrosis or cACLD or a dual one is required. It was recently shown that sequential use of non-invasive fibrosis tests performs better that single tests when a single cut-off is used(27).

The approach of a dual cut-off(1) allows for better classification in terms of higher NPV and PPV of the two cut-offs. Thus, a dual cut-off undoubtedly offers more optimal performance than a single threshold, which would misclassify more patients. Moreover, it allows the use of uniform cut-offs for all major liver disease aetiologies, as the diagnostic performance was similar, particularly for the high cut-off values. On the other hand, we must not disregard the major disadvantage of the dual threshold principle in patient stratification, which lied in the proportion of patients that remained unclassified in the grey zone. In order to rectify for these patients, we provided a novel non-proprietary risk model based on readily available patient characteristics.

Although the high cut-off of >12 kPa performed uniformly well across all liver disease aetiologies, the low cut-off of <7 kPa performed relatively worse in chronic viral hepatitis and considerably worse in CHB. The inclusion of almost exclusively

patients with HBeAg-negative CHB, which has a distinctive natural history with a non-linear fibrosis progression during flares, could be potentially responsible for the worse performance of TE in these patients. Since there was a difference across aetiologies for the low cut-offs, we chose to propose <7 kPa rather than <8 kPa, which had a worse overall sensitivity of 86%. However, specifically for patients with NAFLD or ALD, who are the majority of the patients evaluated these days, a cut-off of <8 kPa for ruling out cACLD has a sensitivity of >90% and could be safely used. It should also be noted that high ALT levels and obesity influence the diagnostic accuracy of the criteria, with lower specificity (and higher sensitivity) of the high and low cut-offs, respectively.

Finally, we show that the quality criteria for TE can be simplified to “reliable” and “poorly reliable”, since there was no difference between “very reliable” and “reliable” measurements in terms of diagnostic accuracy. This simplification is important for clinical practice, as TE is increasingly used outside high volume hepatology services.

This study has some limitations. Firstly, diagnosis of cACLD was based on histological assessment of fibrosis stage, which is considered an imperfect gold standard that can be affected by non-uniform distribution of histological lesions throughout the liver parenchyma and ascertainment variation. Liver biopsies were not analysed by a central pathologist, but the relevant reports from each centre where collected and analysed in this patient database. Although all the pathologists evaluating the biopsies were highly experienced, potential inter-observer variability could not be eliminated. Having said that, the lesions of interest were bridging fibrosis (F3) and cirrhosis, where interobserver variability is lower. In addition, such an independent evaluation of histological lesions reflects better routine clinical practice. Moreover, we chose to include all liver biopsies that histopathologists

deemed adequate for staging, irrespective of the length and number of portal tracts. We do provide however sensitivity analyses based on the quality of liver biopsies. Secondly, TE was performed in all patients using M probe, regardless of their BMI, which might have led to less accurate measurements in obese patients, although the overall diagnostic performance for cACLD was not reduced. Similarly, patients with ALD were not necessarily abstinent from alcohol for at least one week, which could have optimised the TE performance(25). Thirdly, this was a cross sectional analysis, therefore further longitudinal studies with liver-related outcomes could validate the utility of these criteria. Finally, external validation of the diagnostic accuracy of the whole approach is warranted. However, overall, we believe that the proposed algorithm seems to allow for more efficient identification of cACLD in a large diverse patient population than the current standard of care.

In summary, the recommended Baveno VI(1) LS thresholds of <10 and >15 kPa seem to be less sensitive and more specific than required to rule out or diagnose cACLD, as demonstrated in a large real-world population. Therefore, we propose a lower dual cut-off of <7 and >12 kPa to optimise the overall predictive performance of TE for assessing the presence of cACLD. Specifically for ALD and NAFLD, a cut-off of <8 kPa can be safely used for ruling out cACLD. In addition, we developed a risk model to predict the probability of having cACLD in the unclassified patients with indeterminate LS range. In this way, by using a two-step approach and combining TE with readily available patient characteristics, we propose an algorithm that can correctly classify more than 80% of patients with chronic liver disease.

REFERENCES

Author names in bold designate shared co-first authorship.

1. de Franchis R, Baveno VI faculty. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. J Hepatol. 2015;63:743–52.

2. Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. Lancet. 2014;383:1749–61.

3. Castera L, Chan H, Arrese M, Afdhal N, Bedossa P, Friedrich-Rust M, et al. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol. 2015 Jul;63:237–64.

4. Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR, Burroughs AK. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: A meta-analysis of diagnostic accuracy. J Hepatol. 2011;54:650–9.

5. Pavlov CS, Casazza G, Nikolova D, Tsochatzis E, Burroughs AK, Ivashkin VT, et al. Transient elastography for diagnosis of stages of hepatic fibrosis and cirrhosis in people with alcoholic liver disease. Cochrane Database Syst Rev. 2015;1:CD010542.

6. Mueller S, Englert S, Seitz HK, Badea RI, Erhardt A, Bozaari B, et al. Inflammation-adapted liver stiffness values for improved fibrosis staging in patients with hepatitis C virus and alcoholic liver disease. Liver Int Off J Int Assoc Study Liver. 2015;35:2514–21.

7. Maurice JB, Brodkin E, Arnold F, Navaratnam A, Paine H, Khawar S, et al. Validation of the Baveno VI criteria to identify low risk cirrhotic patients not

requiring endoscopic surveillance for varices. J Hepatol. 2016;65:899–905.

8. Augustin S, Pons M, Maurice JB, Bureau C, Stefanescu H, Ney M, et al. Expanding the Baveno VI criteria for the screening of varices in patients with compensated advanced chronic liver disease. Hepatology. 2017;66:1980–8.

9. Thabut D, Bureau C, Layese R, Bourcier V, Hammouche M, Cagnot C, et al. Validation of Baveno VI Criteria for Screening and Surveillance of Esophageal Varices in Patients With Compensated Cirrhosis and a Sustained Response to Antiviral Therapy. Gastroenterology. 2019;156:997-1009.e5.

10. Merchante N, Saroli Palumbo C, Mazzola G, Pineda JA, Téllez F, Rivero-Juárez A, et al. Prediction of Esophageal Varices by Liver Stiffness and Platelets in Persons With Human Immunodeficiency Virus Infection and Compensated Advanced Chronic Liver Disease. Clin Infect Dis. 2019; doi: 10.1093/cid/ciz1181. Online ahead of print.

11. Stafylidou M, Paschos P, Katsoula A, Malandris K, Ioakim K, Bekiari E, et al. Performance of Baveno VI and Expanded Baveno VI Criteria for Excluding High-Risk Varices in Patients With Chronic Liver Diseases: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2019;17:1744-1755.e11.

12. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. J Hepatol. 2007;47:598–607.

13. Brunt EM, Tiniakos DG. Histopathology of nonalcoholic fatty liver disease. World J Gastroenterol. 2010;16:5286.

14. 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;57:1182–91.

15. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Berlin: Springer; 2009.

16. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21:3940–1.

17. Harrell FE. Regression Modeling Strategies with Applications to Linear Models, Logistic and Ordinal Regression and Survival Analysis. 2015.

18. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med. 2007;26:3661–75.

19. Richardson JTE. The analysis of 2 x 2 contingency tables - Yet again. Stat Med. 2011;30:890.

20. Altman DG, Machin D, Bryant TN, Gardner MJ. Statistics with confidence. 2nd ed. MJ Books; 2000. 49.

21. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. 2008.

22. Steyerberg EW, Harrell FE, Borsboom GJJ., Eijkemans MJ., Vergouwe Y, Habbema JDF. Internal validation of predictive models. J Clin Epidemiol. 2001;54:774–81.

23. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodríguez-Perálvarez M, et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. Health Technol Assess (Rockv). 2015 Jan;19:1–410.

24. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2019;156:1717–30.

25. Thiele M, Madsen BS, Hansen JF, Detlefsen S, Antonsen S, Krag A. Accuracy of the Enhanced Liver Fibrosis Test vs FibroTest, Elastography, and Indirect Markers in Detection of Advanced Fibrosis in Patients With Alcoholic Liver Disease. Gastroenterology. 2018;154:1369–79.

26. Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. J Hepatol. 2016;65:570–8.

27. Majumdar A, Campos S, Gurusamy K, Pinzani M, Tsochatzis EA. Defining the Minimum Acceptable Diagnostic Accuracy of Noninvasive Fibrosis Testing in Cirrhosis: A Decision Analytic Modeling Study. Hepatology. 2020;71:627–42.

Table 1. Main characteristics of patients with different aetiologies of liver disease. ANOVA tests were performed to compare values of patient subgroups with different aetiologies. P-values for the comparisons are presented in the last column of the table.

Patient characteristics	All patients, n=5648	CHC, n=2913	NAFLD, n=1073	ALD, n=946	CHB, n=716	P-value
Age, years	50.7±12.5	50.6±11.5	52.6±12.9	55.2±10.5	43.6±13.8	<0.001
Male gender, n (%)	3016 (53)	1463 (50.2)	567 (52.8)	514 (54.3)	472 (65.9)	<0.001
Body mass index, kg/m^2	27.0±6.5	25.8±7.6	30.6±5.4	26.9±5.1	25.3±4.1	<0.001
ALT, IU/L	63 [62]	71 [65]	63 [53]	35 [31]	63 [69]	<0.001
AST, IU/L	45 [36]	48 [41]	42 [27]	37 [35]	42 [35]	0.003
GGT, IU/L	60 [89]	54 [72]	80 [109]	91 [206]	35 [35]	<0.001
Platelets, x10^3/mm^3	210±74	204±70	220±71	233±91	198±71	<0.001
Diabetes mellitus, yes (%)	830 (14.6)	274 (10.4)	442 (43.2)	71 (17.7)	43 (6.5)	<0.001
Liver stiffness, kPa	7.7 [6.1]	7.5 [5.1]	8.5 [6.3]	8.5 [12.0]	6.8 [4.1]	<0.001
Liver stiffness, n (%) <10	3606 (65.7)	1966 (68.6)	602 (61.2)	516 (55.2)	522 (74.1)	<0.001

>15	987 (17.9)	442 (15.5)	167 (17.0)	299 (32.1)	79 (11.2)	<0.001
10-15 kPa	891 (16.3)	456 (15.9)	214 (21.8)	118 (12.7)	103 (14.6)	<0.001
Fibrosis stage, n (%) F0	681 (12.1)	182 (6.2)	199 (18.5)	215 (22.7)	85 (11.9)	<0.001
F1	1644 (29.1)	885 (30.4)	332 (30.9)	182 (19.2)	245 (34.2)	
F2	1551 (27.5)	955 (32.8)	234 (21.8)	189 (20.0)	173 (24.2)	
F3	907 (16.1)	477 (16.4)	179 (16.7)	131 (13.8)	120 (16.8)	
F4	865 (15.3)	414 (14.2)	129 (12.0)	229 (24.2)	93 (13.0)	
cACLD, n (%)	1772 (31.4)	891 (30.6)	308 (28.7)	360 (38.0)	213 (29.7)	<0.001
Liver biopsy samples, n (%)						
> 15mm length	2201/4400 (50%)	755/1943 (39%)	737/986 (75%)	439/928 (47%)	270/523 (52%)	<0.001
>10 portal tracts	1412/2432 (58%)	829/1468 (57%)	379/559 (68%)	15/30 (50%)	182/355 (51%)	<0.001
TE results, n (%)						0.001
“very reliable”[#]	1351 (26.8)	736 (28.3)	202 (22.5)	251 (28.7)	162 (24.6)	
“reliable”[#]	3338 (66.4)	1689 (65.0)	617 (68.8)	573 (65.5)	461 (70.1)	
“poorly reliable”[#]	338 (6.7)	173 (6.7)	78 (8.7)	52 (5.9)	35 (5.3)	

CHC, chronic hepatitis C; NAFLD, Non-alcoholic fatty liver disease; ALD, Alcohol-related liver disease; CHB, chronic hepatitis B; cACLD, compensated advanced chronic liver disease defined as ≥F3 Metavir fibrosis stage; BMI, Body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase

Quantitative variables are presented as mean ± standard deviation (SD) or median values [IQR]. Categorical variables are presented as frequency (valid percentage, %)

[#]TE reliability by interquartile range/median ratio (IQR/M): “very reliable” (IQR/M ≤ 0.10), “reliable” (0.10 < IQR/M ≤ 0.30, or IQR/M > 0.30 with LS median < 7.1 kPa), and “poorly reliable” (IQR/M > 0.30 with LS median ≥7.1 kPa).

Table 2. Predictive factors for compensated advanced chronic liver disease (cACLD) in 1797 unclassified patients with liver stiffness values between 7 and 12 kPa. All reported odds ratios (OR) are based on the bootstrapped binary logistic regression analysis for n=200 test data sets.

Factors	OR (95%CI)	P-value	Adj. OR (95% CI)	P-value
Age per quintile[1]	**1.20 (1.12-1.29)**	**0.005**	**1.19 (1.07-1.32)**	**0.005**
Sex, male vs female	0.94 (0.77-1.15)	0.522	**1.44 (1.09-1.90)**	**0.020**
BMI, per quintile[2]	0.92 (0.85-1.00)	0.060	**0.89 (0.81-0.99)**	**0.045**
Liver stiffness, per kPa	**1.52 (1.41-1.64)**	**<0.001**	**1.51 (1.38-1.66)**	**0.005**
ALT levels, per quintile[3]	**1.13 (1.05-1.22)**	**0.005**	0.97 (0.83-1.14)	0.721
AST levels, per quintile[4]	**1.29 (1.20-1.40)**	**0.005**	**1.36 (1.15-1.59)**	**0.005**
GGT levels, per quintile[5]	**1.14 (1.06-1.23)**	**0.010**	0.95 (0.86-1.05)	0.335
Platelet count, per quintile[6]	**0.75 (0.69-1.81)**	**0.005**	**0.82 (0.74-0.90)**	**0.005**
Type II diabetes, yes vs no	**1.40 (1.07-1.83)**	**0.020**	**1.47 (1.05-2.06)**	**0.020**

CI, confidence interval; Adj., Adjusted; BMI, Body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase.

[1]Age quintiles: ≤40/41-48/49-54/55-61/≥62 years

[2]BMI quintiles: <23/23-24.9/25-27.9/28-30.9/≥31 kg/m^2

[3]ALT quintiles: <40/40-57/58-81/82-124/≥125 IU/L

[4]AST quintiles: <32/32-41/42-54/55-79/≥80 IU/L

[5]GGT quintiles: <32/32-50/51-82/83-142/≥143 IU/L

[6]Platelets quintiles: <157/157-187.9/188-219.9/220-253.9/≥254 x10$^{9/}$mm^3

Figures

Figure 1. Bar chart of specificity and sensitivity levels of the high (>15 kPa) and low (<10 kPa) liver stiffness (LS) cut-offs proposed by the Baveno VI consensus in order to diagnose or rule out compensated advanced chronic liver disease (cACLD), respectively, in different patient groups. CHC, chronic hepatitis C; CHB, chronic hepatitis B; NAFLD, non-alcoholic fatty liver disease; ALD, alcohol-related liver disease; BMI: body mass index (kg/m^2); ALT, alanine aminotransferase.

Figure 2. Bar chart of specificity and sensitivity levels of new high (>12 kPa) and low (<7 kPa) liver stiffness (LS) cut-offs in order to diagnose or rule out compensated advanced chronic liver disease (cACLD), respectively, in different patient groups of the study cohort based on the aetiology of liver disease. CHC, chronic hepatitis C; CHB, chronic hepatitis B; NAFLD, non-alcoholic fatty liver disease; ALD, alcohol-related liver disease; BMI: body mass index (kg/m^2); ALT, alanine aminotransferase.

Figure 3. Classification of patients with chronic liver disease in order to diagnose or rule out compensated advanced chronic liver disease (cACLD) by liver stiffness (LS) cut-offs proposed by the Baveno VI consensus, the newly-proposed LS thresholds of <7 for viral hepatitis or <8 kPa for non-alcoholic fatty liver disease (NAFLD) and alcohol-related liver disease (ALD) and >12 kPa and a risk model for the unclassified patients. ALT, alanine aminotransferase; AST aspartate aminotransferase; BMI: body mass index; GGT, gamma-glutamyl transferase.

Fig 1

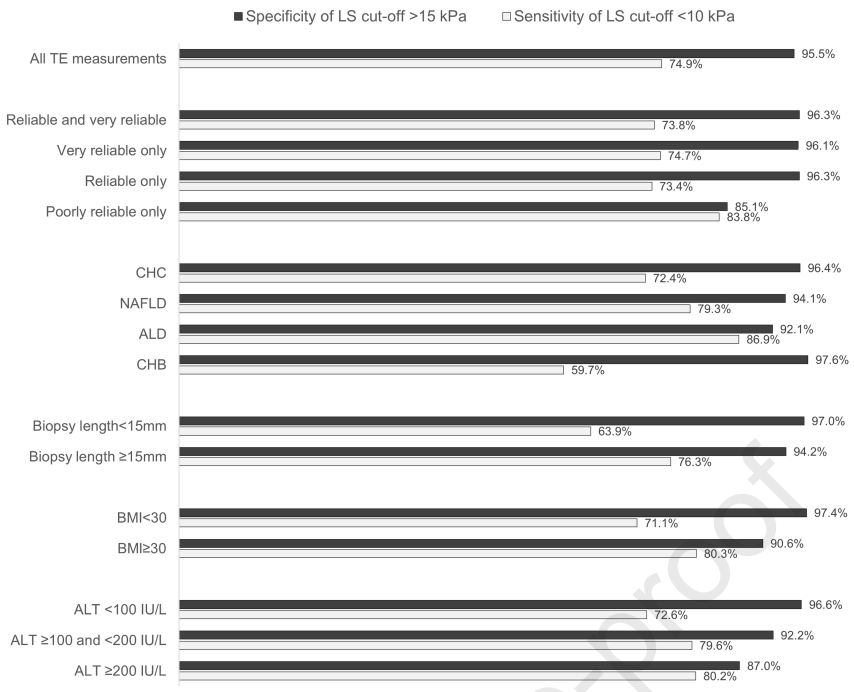


Fig 2

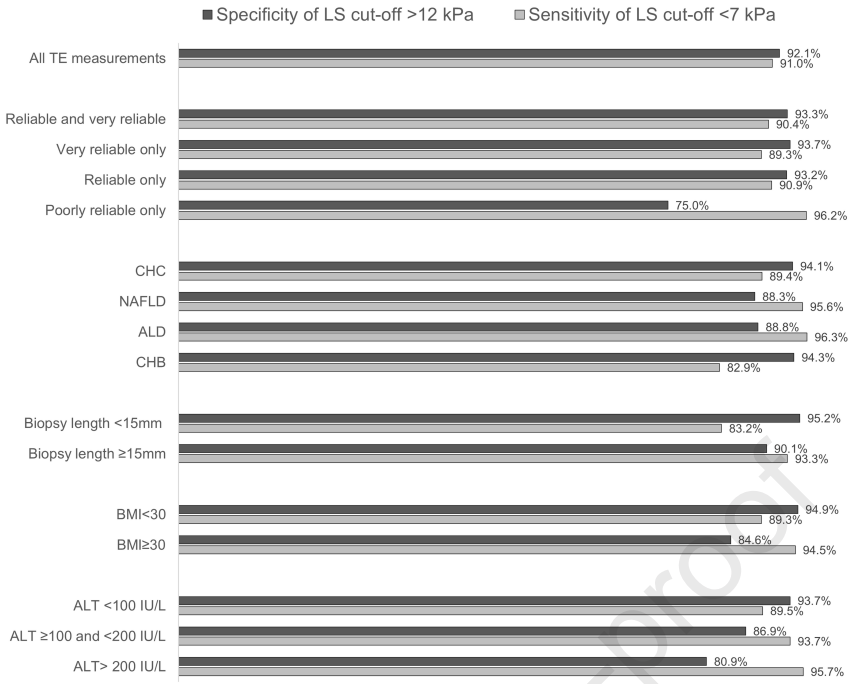
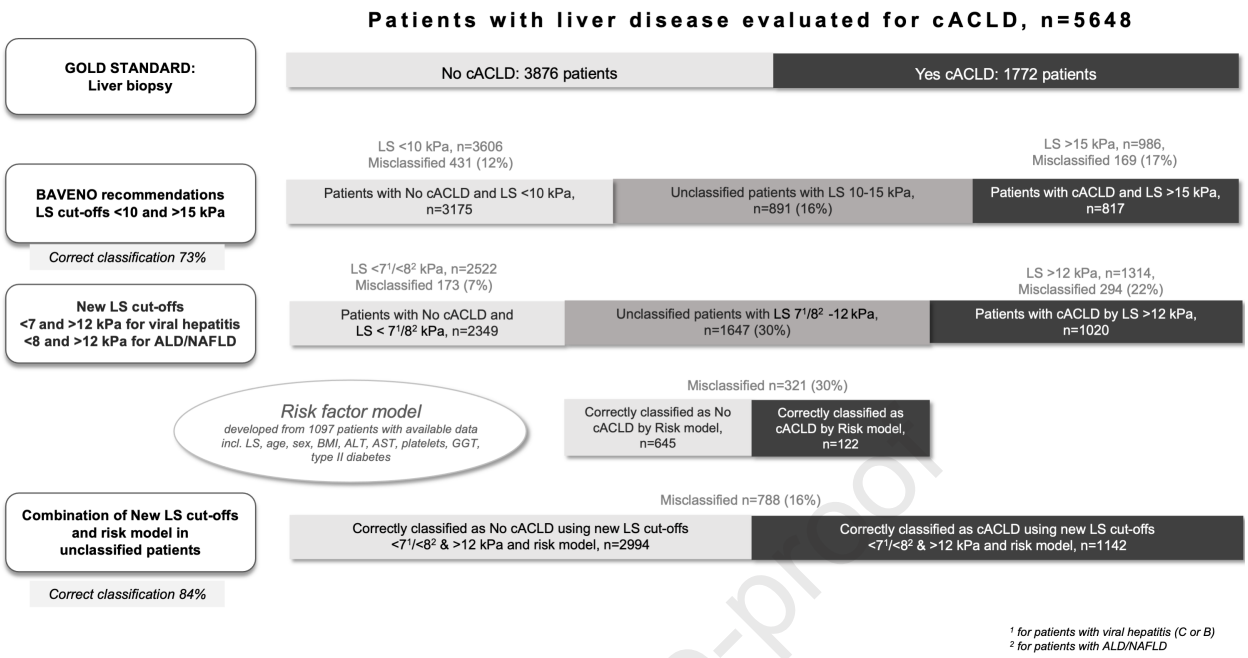


Fig 3



HIGHLIGHTS

- cACLD is the spectrum of advanced fibrosis and cirrhosis in asymptomatic patients.
- The Baveno VI consensus suggested a dual LS cut-off to diagnose/rule out cACLD.
- Proposed LS cut-offs <10/>15 kPa had 75%/96% Se/Sp to rule out/in cACLD.
- We showed that LS cut-offs <7/>12 kPa are more optimal (Se/Sp 91%/92%).
- In ALD and NAFLD, a cut-off <8 kPa to rule out cACLD can be used (Se=93%).