Hepatology Publish Ahead of Print
DOI:10.1097/HEP.0000000000001658

Head-to-head comparison between vibration-controlled transient elastography and histology in predicting liver-related events due to metabolic dysfunction-associated steatotic liver disease

Yangyue Zhang[1,2], Hye Won Lee[3], Huapeng Lin[4], Emmanuel Tsochatzis[5], Salvatore Petta[6], Elisabetta Bugianesi[7], Masato Yoneda[8], Ming-Hua Zheng[9], Hannes Hagström[10,11], Jérôme Boursier[12,13], José Luis Calleja[14], George Boon-Bee Goh[15], Wah-Kheong Chan[16], Rocio Gallego-Durán[17], Arun J. Sanyal[18], Victor de Lédinghen[19], Philip N Newsome[20], Jian-Gao Fan[21], Michelle Lai[22], Céline Fournier-Poizat[19], Grace Lai-Hung Wong[1,2], Grazia Pennisi[6], Angelo Armandi[7], Atsushi Nakajima[8], Wen-Yue Liu[23], Ying Shang[10], Marc de Saint-Loup[12], Elba Llop[14], Kevin Kim Jun Teh[15], Carmen Lara-Romero[17], Amon Asgharpour[18], Sara Mahgoub[24], Mandy Sau-Wai Chan[19], Clemence M Canivet[12,13], Manuel Romero-Gomez[17], Seung Up Kim[3], Vincent Wai-Sun Wong[1,2], Laurent Castéra[25], Terry Cheuk-Fung Yip[1,2]

Affiliations:

1. Department of Medicine and Therapeutics, Medical Data Analytics Centre, The Chinese University of Hong Kong, Hong Kong, China
2. State Key Laboratory of Digestive Disease, Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong, China
3. Department of Internal Medicine, Institute of Gastroenterology, Yonsei University College of Medicine, Seoul, Republic of Korea

4. Department of Gastroenterology and Hepatology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
5. University College London Institute for Liver and Digestive Health, Royal Free Hospital and UCL, London, United Kingdom
6. Section of Gastroenterology, PROMISE, University of Palermo, Italy
7. Department of Medical Sciences, Division of Gastroenterology, A.O.U. Città della Salute e della Scienza di Torino, University of Turin, Turin, Italy
8. Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, Yokohama, Japan
9. MAFLD Research Center, Department of Hepatology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
10. Department of Medicine, Huddinge, Karolinska Institutet, Sweden
11. Division of Hepatology, Department of Upper GI Diseases, Karolinska University Hospital, Huddinge, Stockholm, Sweden
12. Hepato-Gastroenterology and Digestive Oncology Department, Angers University Hospital, Angers, France
13. HIFIH Laboratory, SFR ICAT 4208, Angers University, Angers, France
14. Department of Gastroenterology and Hepatology, Hospital Universitario Puerta de Hierro, Universidad Autonoma de Madrid, Spain
15. Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore
16. Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
17. UCM Digestive Diseases, Virgen del Rocio University Hospital, SeLiver Group, Institute of Biomedicine of Seville, Ciberehd, Department of Medicine, University of Seville, Seville, Spain
18. Stravitz-Sanyal Institute for Liver Disease and Metabolic Health, Virginia Commonwealth University School of Medicine, Richmond, VA, USA
19. Echosens, Paris, France
20. Roger Williams Institute of Liver Studies, School of Immunology & Microbial Sciences, Faculty of Life Sciences and Medicine, King's College London, Foundation for Liver Research and King's College Hospital, London

21. Department of Gastroenterology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai Key Lab of Pediatric Gastroenterology and Nutrition, Shanghai, China
22. Division of Gastroenterology & Hepatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
23. Department of Endocrinology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
24. National Institute for Health Research, Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, Birmingham, United Kingdom
25. Université Paris Cité, UMR1149 (CRI), INSERM, Paris, France; Service d'Hépatologie, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris (AP-HP), Clichy, France

Author Contributions:

VW-SW designed the study. TC-FY, HL, HWL, ET, SP, EB, MY, M-HZ, HH, JB, JLC, GB-BG, W-KC, RG-D, AJS, VdL, PNN, J-GF, GL-HW, GP, AA, AN, W-YL, YS, MdS-L, EL, KKJT, CL-R, AA, SM, CMC, MR-G, SUK and VW-SW collected data in this study. ET, SP, EB, M-HZ, HH, JB, JLC, GB-BG, W-KC, AJS, VdL, PNN, MR-G, SUK and VW-SW supervised the project. YZ, TC-FY, and VW-SW were responsible for data analysis and data interpretation, and drafted the manuscript. YZ and TC-FY prepared the figures. All authors provided review and editing of the manuscript and approved the final version of the manuscript.

Address for Correspondence:

Prof. Laurent Castéra

Department of Hepatology, Hopital Beaujon, Université Paris Cité, Clichy, France

Email: laurent.castera@bjn.aphp.fr

Dr. Terry Cheuk-Fung Yip

Department of Medicine and Therapeutics, Prince of Wales Hospital, Shatin, Hong Kong, China

Tel: (852) 3505 3125

Email: tcfyip@cuhk.edu.hk

Financial support and Sponsorship:

The study was supported in part by the Noncommunicable Chronic Diseases-National Science and Technology Major Project of China (2023ZD0508700), General Research Fund by the Hong Kong SAR Government (14106923), and Direct Grant of The Chinese University of Hong Kong (2024.065).

Conflicts of interest:

Emmanuel Tsochatzis consults for and is on the speakers' bureau for Novo Nordisk and Boehringer Ingelheim. He consults for Pfizer, Madrigal, MSD, and Siemens. He is on the speakers' bureau for Gilead, Echosens, AbbVie, Dr. Falk, and AstraZeneca. Salvatore Petta advises or is on the speakers' bureau for AbbVie, Echosens, MSD, Novo Nordisk, Pfizer, and Resalis. Elisabetta Bugianesi consults for, advises, and is on the speakers' bureau for MSD and Novo Nordisk. She consults for and advises Boehringer Ingelheim and Lilly. She consults for and is on the speakers' bureau for Madrigal. She consults for Pfizer. She received grants from Gilead. Masato Yoneda is on the speakers' bureau for Kowa. He received grants from Gilead. Hannes Hagström consults for AstraZeneca. He received grants paid to the institution from AstraZeneca, Echosens, Gilead, Intercept, MSD, and Pfizer. Jérôme Boursier consults for, advises, is on the speakers' bureau for, and received grants paid to the institution from Intercept. He consults for, is on the speakers' bureau for, and received grants paid to the institution from Siemens. He consults for and received grants paid to the institution from Echosens. He consults for AstraZeneca. He is on the speakers' bureau for AbbVie and Gilead. He advises BMS, Pfizer, MSD, and Novo Nordisk. He received grants paid to the institution from Diafir and Inventiva. José Luis Calleja consults for and is on the speakers' bureau for Echosens, Gilead, and AbbVie. George Boon-Bee Goh consults for Roche and Ionis. He is on the speakers' bureau for Echosens, Viatris, Abbott, and Novo Nordisk. Wah-Kheong Chan consults for, advises, is on the speakers' bureau for, and received grants from Abbott, AbbVie, and Roche. He consults, advises, and is on the speakers' bureau for Novo Nordisk. He consults for and advises Zuellig and Boehringer Ingelheim. He advises and is on the speakers' bureau for Viatris. He is on the speakers' bureau for Echosens and Hisky. Arun J. Sanyal consults for and received grants paid to the institution from AstraZeneca, BMS, Gilead, Pfizer, Merck, Genfit, Intercept, Novo Nordisk, Novartis, Salix,

Allergan, Echosens, Zydus, and Lilly. He advises or is on the DSMB for Bard Peripheral Vascular, NGM, and Sequana. He received grants from and owns stock in Durect and Tiziana. He consults for Abbott, ProMED, Satellite Bio, Corcept, Arrowhead, Boston Pharmaceuticals, Variant, Inventiva, Cascade, 89bio, Alnylam, Amgen, Regeneron, Boehringer Ingelheim, Genentech, Janssen, Lipocine, Madrigal, GSK, Akero, PathAI, HistoIndex, Poxel, Myovant, Median Technologies, Siemens, Sun Pharma, Surrozen, Takeda, Terns, and Valeant. He received grants paid to the institution from Albireo, Mallinckrodt, Ocelot, Perspectum, and Inversago. He owns stock in Exalenz and Genfit. He owns intellectual property rights with Elsevier and UpToDate. Victor de Lédinghen is employed by Echosens. Philip N. Newsome consults for, is on the speakers' bureau for, and received grants from Novo Nordisk. He consults for Boehringer Ingelheim, Gilead, Intercept, Poxel, Pfizer, BMS, Lilly, Madrigal, and GSK. He is on the speakers' bureau for AiCME. Michelle Lai consults for and received grants from Madrigal. She advises Altimmune. She received grants from Inventiva. Céline Fournier-Poizat is employed by Echosens. Grace Lai-Hung Wong consults for, advises, is on the speakers' bureau for, and received grants from Gilead. She consults for, advises, and is on the speakers' bureau for GSK and Janssen. She consults for and advises AstraZeneca. She is on the speakers' bureau for Abbott, AbbVie, Ascletis, BMS, Echosens, Furui, and Roche. Atsushi Nakajima is on the speakers' bureau for and received grants from Mochida, Biofermin, and EA Pharma. He received grants from Astellas and ASKA. He is on the speakers' bureau for Kowa, MSD, Boehringer Ingelheim, Novo Nordisk, and GSK. Amon Asgharpour consults for Galectin. He advises Madrigal, Intercept, and Novo Nordisk. He is on the speakers' bureau for CLDF and Medscape. Mandy Sau-Wai Chan is employed by Echosens. Manuel Romero-Gómez consults for, is on the speakers' bureau for, and received grants from Siemens. He is on the speakers' bureau for and received grants from Echosens. He consults for Ipsen, Gilead, and UCB Pharma. He received grants from Novo Nordisk and Theratechnologies. Vincent Wai-Sun Wong consults for, advises, is on the speakers' bureau for, and received grants from Gilead. He consults for, advises, and is on the speakers' bureau for AbbVie and Novo Nordisk. He consults for and advises Boehringer Ingelheim, Echosens, Intercept, Inventiva, Pfizer, Sagimet, Visirna, and Target PharmaSolutions. He is on the speakers' bureau for Abbott and Unilab. He is a co-founder of and owns stock in Illuminatio Medical Technology. Laurent Castéra consults for and is on the speakers' bureau for Boehringer Ingelheim, Echosens, Madrigal, Gilead, and Novo Nordisk. He consults for Boston

Pharmaceuticals, GSK, MSD, Novo Nordisk, Sagimet, Siemens, and Pfizer. He is on the speakers' bureau for AstraZeneca and Inventiva. Terry Cheuk-Fung Yip consults for, advises, and received grants from Gilead. The remaining authors have no conflicts to report.

Keywords:
Hepatocellular carcinoma; MASLD; VCTE; hepatic decompensation; fibrosis staging

List of Abbreviations:

ALT = Alanine Transaminase
AUROC = Area Under the Receiver Operating Characteristic Curve
AUPRC = Area Under the Precision-Recall Curve
BMI = Body Mass Index
CI = Confidence Interval
HCC = Hepatocellular Carcinoma
IDI = Integrated Discrimination Improvement
LRE = Liver-Related Event
LSM = Liver Stiffness Measurement
MASLD = Metabolic Dysfunction-Associated Steatotic Liver Disease
MASH = Metabolic Dysfunction-Associated Steatohepatitis
NLR = Negative Likelihood Ratio
NPV = Negative Predictive Value
NRI = Net Reclassification Improvement
PLR = Positive Likelihood Ratio
PPV = Positive Predictive Value
VCTE = Vibration-Controlled Transient Elastography

Graphical Abstract

GA1

ABSTRACT

Background & Aims: Liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE) correlates well with liver-related events (LREs), but previous head-to-head comparisons with liver histology were limited by small sample size. This study aimed to compare the prognostic performance of LSM and histology for predicting LREs in patients with metabolic dysfunction-associated steatotic liver disease (MASLD).

Approach & Results: We analyzed data from 3,532 MASLD patients (mean age 51.9 years, 57.3% male) in the multicenter VCTE-Prognosis cohort, all having undergone both LSM and liver biopsy. The primary outcome was LREs, defined as hepatic decompensation, liver transplantation, or liver-related death. Secondary outcomes included HCC and decompensation analyzed separately. Median baseline LSM was 8.8 kPa; 33.5% had F3–F4 fibrosis. Over a median follow-up of 56.6 months, 126 patients (3.6%) developed LREs (123 decompensation). LSM and histology demonstrated similar prognostic performance for LREs, with comparable 5-year area under the receiver operating characteristic curve (AUROC) values (0.870 vs. 0.869), integrated AUROCs (0.878 vs. 0.852), and integrated precision-recall curves (0.137 vs. 0.0.68). The 5-year integrated Brier scores were also similar (1.389% vs. 1.391%), and the integrated discrimination improvement index showed no significant difference. Similar results were found across all the outcomes, time-points, and sensitivity analyses.

Conclusions: In this large MASLD cohort, LSM by VCTE showed comparable prognostic accuracy to histology. As a non-invasive tool, LSM may serve as an alternative surrogate endpoint in clinical trials.

INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD) is the most prevalent chronic liver disease, impacting 30% of the population globally.[1] It is also one of the important causes of cirrhosis and hepatocellular carcinoma (HCC).[1] The severity of liver fibrosis, a key predictor of clinical outcomes, is traditionally assessed by liver biopsy.[2] However, biopsy has been criticized for being invasive, costly, setting-limited, and occasionally associated with morbidity and mortality.[2] Therefore, having a safe, affordable and accessible tool to measure liver fibrosis and predict liver-related clinical outcomes becomes crucial.

Non-invasive liver stiffness measurement (LSM) using vibration-controlled transient elastography (VCTE) has been shown to be appropriate in this context. Comparing to other non-invasive methods, VCTE demonstrates superior performance in detecting fibrosis stage. An individual data meta-analysis reported an area under the receiver operating characteristic curve (AUROC) of 0.85 for detecting advanced fibrosis and 0.90 for cirrhosis, outperforming other serum-based tests.[3] Additionally, a prospective validation study confirmed the high diagnostic accuracy of VCTE, showing it to have the highest AUROC for identifying advanced fibrosis among single biomarkers.[4]

Beyond detecting fibrosis stage, the prognostic performance of VCTE in predicting liver-related clinical outcomes has also been demonstrated. In the VCTE-Prognosis study, we found that the cumulative incidence rate of liver-related events (LREs) significantly increased over 5 years in patients with LSM in a dose-dependent manner, and changes in LSM further refined the prediction.[5] Similar findings were reported by Gawrieh et al., who found that the risk of developing LREs increased when LSM exceeded 10 kPa, with a 1.9-fold higher risk in patients with values between 10 to 14.9 kPa, and an 8.3-fold higher risk in those with values greater than 15 kPa.[6] The Baveno VII consensus also suggested the use of LSM in detecting portal hypertension and high-risk varices,[7] which both manifest disease progression in advanced liver disease.

Meanwhile, an individual participant data meta-analysis by the LITMUS group validated the use of VCTE in predicting clinical outcomes in MASLD, by comparing it to liver histology. This research included 2518 participants, of whom 5.8% (n=145) experienced liver-related clinical outcomes. Similarly, LSM by VCTE was associated with increased risk of LRE in a stepwise manner.[8] Additionally, the AUROC for LSM by VCTE and liver histology showed no significant differences, indicating comparable performance in predicting liver-related clinical outcomes.[8] However, the generalizability of the findings is geographically restricted to Europe and Asia, as data from other areas are lacking.[8] Meanwhile, previous studies were also limited by small sample sizes and a limited number of events.

On the other hand, the current endpoint for metabolic dysfunction-associated steatohepatitis (MASH) drug development, accepted by the regulators, is histological response, including MASH resolution and fibrosis improvement.[9] However, the invasive nature of liver histology significantly hinders participants' recruitment and complicates research work. Additionally, liver biopsy has been challenged for intra- and inter-observer variability,[10] further limiting its efficacy as research tool. In response, the US Food and Drug Administration has recently outlined regulatory considerations regarding non-invasive tests, like LSM by VCTE, as a reasonably likely surrogate endpoint to predict clinical benefits for MASH therapies.[11] Altogether, this highlights the need for head-to-head comparisons between the non-invasive and histological assessments to evaluate their efficacy, and further support the adoption of non-invasive tests as surrogate endpoints.

In this context, we performed an analysis of the multicenter VCTE-Prognosis cohort, aiming to compare the prognostic performance of LSM by VCTE and fibrosis stage by histology for predicting liver-related clinical outcomes in patients with MASLD.

METHODS

Study design and participants

The VCTE-Prognosis study included patients with MASLD who underwent VCTE examination at 16 tertiary centers across the US, Europe, and Asia, with data prospectively collected from 14 of these centers. The study design was described in detail in previous research.[5] The specific

analysis plan and combined evaluation presented in this study were not pre-specified at the time of data collection and are therefore considered post hoc analyses. We included adult patients aged 18 or older who underwent both VCTE and liver biopsy for the assessment of MASLD. Patients with other liver diseases were excluded, including those with chronic viral hepatitis, HIV infection, excessive alcohol consumption (greater than 30 g/day for males and 20 g/day for females), secondary causes of hepatic steatosis, or a history of HCC, hepatic decompensation, liver resection, liver transplant, or other malignancies.

The study protocol received approval from the institutional review boards of the participating sites and was conducted following the principles of the Declaration of Helsinki. While patients provided written informed consent for participation in the prospective programs at the local sites, consent for the secondary analysis was waived.

Assessments

During each clinical visit, the medical history was documented. Body mass index (BMI) was calculated by dividing weight (kg) by height (m) squared. A venous blood sample was collected after the patients fasted for at least 8 hours to assess kidney function, liver biochemistry, complete blood count, glucose, and lipids. Liver stiffness was assessed using the VCTE device (FibroScan, Echosens, Paris, France) by trained operators, with a requirement of at least 10 valid acquisitions per patient.[12] Liver biopsy was performed to determine the histological fibrosis stage, based on the Nonalcoholic Steatohepatitis Clinical Research Network scoring system.

Outcomes

The primary outcome consisted of a composite of LREs including hepatic decompensation (ascites, spontaneous bacterial peritonitis, variceal hemorrhage, hepatic encephalopathy or hepatorenal syndrome), liver transplant, and liver-related death. Secondary outcomes were HCC and hepatic decompensation, analyzed separately. Event diagnoses were based on prospective follow-up, medical record review, or validated registries with the outcomes of interest correctly classified in over 90% of cases. Events were recorded based on their first occurrence and were not censored thereafter.

Statistical analysis

To mitigate immortal time bias, the baseline date was defined as the latter of the first VCTE assessment date and the blood test date, if they were not performed on the same day, as previously described.[5] Continuous variables are presented as either the mean (standard deviation) or median (25^{th} to 75^{th} percentile [P25-P75]), depending on the distribution, while categorical variables are reported as frequencies (percentages). The risk categories of LSM were stratified based on previously employed cut-offs,[5,8] with less than 10 kPa as low risk, 10 to 20 kPa as intermediate risk, and greater than 20 kPa as high risk. Liver histology was categorized into three groups as F0 to F2, F3 and F4 based on the Nonalcoholic Steatohepatitis Clinical Research Network scoring system, as previously mentioned.[5] To account for competing events, the cumulative incidence of primary and secondary outcomes was estimated and compared across different risk categories using Aalen-Johansen method and Gray's test, respectively. Meanwhile, the cumulative incidence rate for LREs was also assessed by combining LSM and liver histology. Patients were first stratified by histological risk group and then further divided according to LSM risk categories. Concordance or discordance between LSM and histology was determined by cross-referencing their respective risk classifications. The individual concordance between LSM and liver histology was illustrated through violin plots, showing the distribution of LSM by fibrosis stage, and analyzed through Kendall's Tau and Spearman's rank correlation. To further explore discordant classifications, we performed logistic regression with relevant clinical characteristics, including age, sex, center (Europe/USA vs. Asia), diabetes, hypertension, BMI and alanine transaminase (ALT). Discordant classification was defined according to overestimation and underestimation models (Supplementary Methods 1, http://links.lww.com/HEP/K340). The competing event for both primary and secondary outcomes included non-liver-related mortality, while HCC was additionally considered as a competing event for hepatic decompensation, as previously outlined.[5]

To evaluate the prognostic performance, the time-dependent sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR) and negative likelihood ratio (NLR) at 2, 3 and 5 years were assessed. Integrated Brier score was employed to evaluate and compare the overall accuracy. Both the time-dependent area under the receiver-operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC), accounting for competing events, were assessed and compared between LSM by

VCTE and histological fibrosis stage at the 2, 3, and 5-year time points to evaluate discriminatory performance.[13,14] We also calculated the integrated time-dependent AUROC and AUPRC over a 5-year period to summarize overall performance. Categorical net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were assessed with reference to histological fibrosis stage to further evaluate and compare the predictive performance of LSM for clinical outcomes.[15,16] We performed calibration plots for both LSM and liver histology at the 2-, 3-, and 5-year time points to visualize the difference between the predicted and observed risks. To better inform future clinical practice, the estimated likelihood ratio and the post-test probability for different LSM values by VCTE were calculated. To test the robustness of our findings, we conducted a series of sensitivity analyses for the primary outcome using four approaches which were specifically outlined in Supplementary Methods 2, http://links.lww.com/HEP/K341 . All data analysis was conducted using R (version 4.3.1, R Core Team, 2023) and the R packages used were listed in Supplementary Methods 3, http://links.lww.com/HEP/K342 . Statistical tests were all conducted two-sided, with a significance level set at $p<0.05$.

RESULTS

Participants

Between February 2004 and January 2023, a total of 17,949 patients who had undergone at least one VCTE examination were identified. After applying the inclusion and exclusion criteria, 14,417 patients were excluded, leaving 3,532 patients with both liver histology and VCTE results for analysis (Figure 1). The patients had a mean age of 51.9 (13.3) years, and 42.7% (n=1,509) were female (Table 1). The median time interval between VCTE and liver biopsy was 0.93 months (P25-P75: 0-7.13). Diabetes was present in 1,471 patients (41.7%), while 1,639 (41.7%) had hypertension. The median LSM value was 8.8 kPa (P25-P75: 6.1-13.3), and 1,182 (33.5%) patients had a histological stage of F3-4. In terms of geographic distribution, 45.2% (n=1,597) of the patients were from Asia, while the remaining 54.8% (n=1,935) were from the US or Europe (Supplementary Table 1, http://links.lww.com/HEP/K344).

Liver-Related Events

Over a median follow-up of 56.6 months (P25-P75: 30.5-91.2), 126 (3.6%) patients developed LREs, including 123 (3.5%) incidences of hepatic decompensation (Supplementary Table 2, http://links.lww.com/HEP/K344). For patients who developed LREs, the median follow-up was 42.8 months (P25-P75: 18.2–76.4), whereas for those who did not develop LREs, the median follow-up was 57.0 months (P25-P75: 31.1–91.7). The most frequently observed hepatic decompensating event was ascites (n=90), followed by variceal bleeding (n=34), hepatic encephalopathy (n=32), spontaneous bacterial peritonitis (n=3) and hepatorenal syndrome (n=3). The 2-, 3-, and 5-year cumulative incidence (95% CI) of LREs were 1.3% (0.9%-1.7%), 1.8% (1.4%-2.3%), and 3.0% (2.4-3.7%), respectively. The corresponding cumulative incidences were 0.3% (0.2%-0.6%), 0.7% (0.5%-1.1%), and 1.1% (0.8%-1.6%) for HCC; and 1.3% (0.9%-1.7%), 1.8% (1.4%-2.3%), and 3.0% (2.3%-3.7%) for hepatic decompensation.

Risk Stratification by Liver Histology and LSM by VCTE

When classified by LSM, patients with LSM above 20 kPa had the highest 5-year cumulative incidence at 14.4% (10.8%-18.4%), compared to 3.3% (2.2%-4.8%) in those with LSM between 10-20 kPa, and 0.3% (0.1%-0.7%) in those with LSM below 10 kPa (Gray's test, $p<0.001$) (Figure 2A). A similar pattern was observed when trichotomized by liver histology stage, the 5-year cumulative incidence of LREs was highest in patients with F4 at 16.0% (12.3%-20.2%), followed by 3.0% (1.8%-4.7%) in those with F3 and 0.4% (0.2%-0.8%) in those with F0–F2 (Gray's test, $p<0.001$) (Figure 2B). These differences remained significant at the 2- and 3-year time points for both tests when comparing trichotomized groups, as well as in the secondary analyses for HCC and hepatic decompensation (Supplementary Figures 1, http://links.lww.com/HEP/K343 and 2, http://links.lww.com/HEP/K343 , all $p<0.001$ by Gray's test).

In the combined analysis of patients with F0–F2, the 5-year cumulative incidence of LREs increased with LSM cutoffs, showing 0.1% (0.0%-0.4%) in those with LSM below 10 kPa (concordant low-risk), 1.1% (0.4%-2.7%) for LSM between 10 and 20 kPa (discordant), and 1.8% (0.1%-8.6%) for LSM above 20 kPa (discordant) ($p=0.005$) (Supplementary Figure 3A, http://links.lww.com/HEP/K343). Among patients with F3, those with LSM between 10 and 20 kPa (concordant) had a 5-year incidence of 2.9% (1.3%-5.5%), while no events were recorded in

those with LSM below 10 kPa. In contrast, patients with LSM above 20 kPa (discordant) had a significantly higher incidence of 7.8% (3.8%-13.8%) (p<0.001) (Supplementary Figure 3B, http://links.lww.com/HEP/K343). For patients with F4, those with LSM above 20 kPa (concordant high-risk) had a 5-year cumulative incidence of 20.7% (15.1%-26.9%), compared to 10.8% (6.1%-17.0%) and 10.7% (2.6%-25.3%) in those with LSM between 10 and 20 kPa and below 10 kPa, respectively (p=0.028) (Supplementary Figure 3C, http://links.lww.com/HEP/K343). Stratification by fibrosis stage within LSM categories yielded identical incidence values, as the patient combinations remained to be same but were regrouped. However, significant difference across groups was still observed (p<0.001), with patients with a stage of F4 showing the highest incidence rate in developing LREs irrespective of LSM values (Supplementary Figures 4A-C, http://links.lww.com/HEP/K343).

When examining the concordance between LSM and histology at the individual level, median LSM values showed a dose-response relationship with fibrosis stage, while the ranges overlapped (Supplementary Figure 5, http://links.lww.com/HEP/K343). The correlation was moderately positive, with Kendall's Tau ($\tau = 0.462$, $p < 0.001$) and Spearman's rank correlation ($\rho = 0.591$, $p < 0.001$). In the adjusted underestimation model, F4 fibrosis and recruitment in Europe or the USA were associated with higher odds of underestimation (adjusted OR 1.748, 95% CI 1.300-2.350, $p < 0.001$; adjusted OR 1.500, 95% CI 1.093–2.058, $p = 0.012$, respectively) (Supplementary Table 3, http://links.lww.com/HEP/K344). In contrast, older age, higher BMI, elevated ALT, and the presence of diabetes were associated with lower odds of underestimation. In the adjusted overestimation model, F2 fibrosis, older age, diabetes, higher BMI, and elevated ALT were associated with increased odds of overestimation, whereas recruitment in Europe or the USA was associated with decreased odds (adjusted OR 0.620, 95% CI 0.499-0.771, $p < 0.001$).

Discriminatory Performance of Liver Histology and LSM by VCTE

Five-year estimates showed that liver histology and LSM demonstrated comparable performance in predicting LREs, with AUROCs of 0.869 (95% CI 0.839-0.895) and 0.870 (95% CI 0.834-0.905), respectively (Table 2). Similarly, AUPRC values were 0.141 (95% CI 0.103-0.181) for

histology and 0.210 (95% CI 0.143-0.312) for LSM. Comparable results were observed in secondary analyses of hepatic decompensation and HCC. For hepatic decompensation, 5-year AUROCs were 0.868 (95% CI 0.839-0.894) for histology and 0.862 (95% CI 0.825-0.899) for LSM, with 5-year AUPRCs of 0.140 (95% CI 0.104-0.182) and 0.196 (95% CI 0.134-0.297), respectively. In predicting HCC, histology had a 5-year AUROC of 0.842 (95% CI 0.790-0.884) and a 5-year AUPRC of 0.044 (95% CI 0.025-0.069), while LSM had a 5-year AUROC of 0.826 (95% CI 0.760-0.884) and a 5-year AUPRC of 0.057 (95% CI 0.031-0.146). The patterns remained consistent at the 2- and 3-year time points across all outcomes.

The integrated time-dependent AUROC for LSM in predicting LREs was 0.878 (95% CI 0.838-0.915), compared to 0.852 (95% CI 0.808-0.890) for histological fibrosis stage, demonstrating similar predictive ability (p=0.159) (Figure 3A). Secondary analysis of hepatic decompensation showed a comparable result, with integrated AUROCs of 0.864 (95% CI 0.816-0.905) for LSM and 0.856 (95% CI 0.814-0.892) for histology (p=0.613) (Figure 3B). For HCC, the corresponding values were 0.821 (95% CI 0.707-0.902) for LSM and 0.813 (95% CI 0.735-0.879) for histology (p=0.810) (Figure 3C), further supporting the similar performance. The integrated time-dependent AUPRC for LSM in predicting LREs was 0.137 (95% CI 0.089-0.233), compared to 0.068 (95% CI 0.050-0.090) for liver histology (p=0.050) (Figure 4A). In predicting hepatic decompensation, the integrated AUPRC of LSM was 0.116 (95% CI 0.079-0.194), while the corresponding value for histology was 0.068 (95% CI 0.050–0.091) (p=0.087) (Figure 4B). For HCC, the integrated AUPRCs for LSM and histology were 0.037 (95% CI 0.020-0.106) and 0.020 (95% CI 0.012-0.034), respectively (p=0.447) (Figure 4C).

Prognostic Performance of Liver Histology and LSM by VCTE at Specific Cutoffs

In predicting LRE development in 5 years, histological F3-4 and F4 demonstrated sensitivities of 91.9% and 70.9%, specificities of 68.0% and 87.0%, PPVs of 8.3% and 14.7%, and PLR of 2.9 and 5.5, respectively (Table 3). For LSM thresholds of ≥10 kPa and ≥20 kPa, the corresponding sensitivities were 92.9% and 60.5%, specificities were 60.8% and 89.2%, PPVs were 6.9% and 14.9%, and PLRs were 2.4 and 5.6, respectively. NPVs remained consistently high across all tests and cutoffs, ranging from 98.6 to 99.6 while NLRs remained between 0.1 to 0.4. The overall accuracy was higher for both tests when using the higher cutoff, namely histology F4 or

LSM ≥ 20 kPa, with 86.4% and 88.0%, respectively. The prognostic performance of liver histology and LSM for LREs at 2 and 3 years was summarized in Table 3, while secondary analyses for HCC and hepatic decompensation were presented in Supplementary Tables 4, http://links.lww.com/HEP/K344 and 5, http://links.lww.com/HEP/K344 , respectively.

Calibration of Liver Histology and LSM by VCTE

When comparing predicted and actual risks, both LSM and histology demonstrated good calibration in LREs prediction (Supplementary Figure 6, http://links.lww.com/HEP/K343). This pattern was consistent across all time points and outcomes (Supplementary Figures 7, http://links.lww.com/HEP/K343 and 8, http://links.lww.com/HEP/K343). The 5-year integrated Brier score for LSM in predicting LREs was 1.389%, compared to 1.391% for liver histology, indicating highly similar performance (Supplementary Figure 9A, http://links.lww.com/HEP/K343). Throughout the follow-up period, the Brier scores for both measures remained closely aligned, a similar pattern that was also observed in secondary analyses of hepatic decompensation and HCC (Supplementary Figures 9A-C, http://links.lww.com/HEP/K343).

In the paired comparison of the two tools for predicting LREs, LSM again demonstrated similar performance relative to histologic fibrosis stage, with a 5-year event NRI of 0.018 (95% CI -0.060 - 0.094), an overall NRI of -0.068 (95% CI -0.148 - 0.009), and the IDI index of 0.010 (95% CI -0.033 - 0.065) (Table 4). A similar pattern was observed in other investigated time points and the secondary analyses of hepatic decompensation and HCC, where the majority of NRI values and all IDI values were comparable between the two methods.

Likelihood ratio and post-test probability of LSM by VCTE

The estimated likelihood ratio indicated a reduced risk of LREs when LSM was below 10 kPa, with values of 0.14 at 8 kPa and 0.37 at 10 kPa (Supplementary Figure 10, http://links.lww.com/HEP/K343). The likelihood ratio reached a neutral reference point at 12.7 kPa, where the risk of developing LREs was balanced. At 15 kPa, the likelihood of LRE occurrence increased by 72%, and the risk increased 2.83-fold when LSM exceeded 20 kPa. Over the 5-year follow-up, the post-test probability of LREs remained below 0.02 when LSM

was <10 kPa for individuals with a pre-test probability of 0.01 or 0.05, indicating the target threshold for low-risk profile (Supplementary Figure 11, http://links.lww.com/HEP/K343). However, a marked increase in post-test probability occurred between 10 and 20 kPa, reaching 0.13 at LSM of 20 kPa for a pre-test probability of 0.05. Beyond 20 kPa, the probability continued to increase but at a slower rate, exceeding 0.21 at LSM of 30 kPa with the same pre-test probability.

Sensitivity analysis

Of the cohort of 3,532 patients, 3,086 (87.4%) were collected prospectively, among whom 116 LREs were observed, 114 of which were hepatic decompensation (Supplementary Table 6, http://links.lww.com/HEP/K344). The 5-year cumulative incidence (95% CI) of LREs was 15.9% (11.8%-20.4%) in patients with LSM above 20 kPa, compared with 4.1% (2.7%-5.9%) in those with LSM 10 to 20 kPa and 0.4% (0.1%-0.9%) in those with LSM less than 10 kPa ($p<0.001$) (Supplementary Figure 12A, http://links.lww.com/HEP/K343). A similar trend was seen with histological stage with 18.6% (14.2%-23.5%) in F4, 3.5% (2.1%-5.4%) in F3, and 0.4% (0.2%-0.9%) in F0 to 2 ($p<0.001$) (Supplementary Figure 12B). At year 5, the time-dependent AUROC for predicting LREs was comparable between histology (0.873, 95% CI 0.839-0.901) and LSM (0.872, 95% CI 0.835-0.907) (Supplementary Table 7, http://links.lww.com/HEP/K344). The integrated AUROC also did not differ significantly between LSM (0.879, 95% CI 0.833-0.915) and histology (0.852, 95% CI 0.806-0.893) ($p=0.162$) (Supplementary Figure 13, http://links.lww.com/HEP/K343).

When stratified by region, patients in Asia showed a stepwise increase in cumulative incidence rates adopted the LSM cut-offs, showing 0.1% (0.0%-0.6%) for LSM less than 10 kPa, 2.5% (1.3%-4.4%) for LSM between 10 to 20 kPa, and 10.2% (5.8%-16.0%) for LSM above 20 kPa (Supplementary Figure 14A, http://links.lww.com/HEP/K343). A similar pattern was observed when stratifying incidence rates by histological stage, with 0.4% (0.1%-1.1%) for F0-2, 2.6% (1.2%-5.1%) for F3, and 11.7% (6.8%-18.1%) for F4 (Supplementary Figure 14B, http://links.lww.com/HEP/K343). In patients from Europe and the USA, a progressive increase in cumulative incidence rates was also seen when applying both LSM cut-offs and fibrosis stages. The 5-year cumulative incidence rates were 0.5% (0.2%-1.3%), 4.3% (2.5%-6.9%), and 17.2%

(12.2%-22.9%) according to LSM cut-offs, and 0.3% (0.1%-0.8%), 3.4% (1.7%-6.0%), and 18.5% (13.6%-24.0%) according to histological stages (Supplementary Figure 15A and B, http://links.lww.com/HEP/K343). At year 5, the time-dependent AUROC for LSM in predicting LREs was 0.859 (95% CI 0.795-0.909) in Asia, compared with 0.850 (95% CI 0.788-0.904) for histology. The integrated AUROC also showed no significant difference, with 0.839 (95% CI 0.744-0.912) for LSM and 0.839 (95% CI 0.742-0.920) for histology (p=0.990) (Supplementary Figure 16A, http://links.lww.com/HEP/K343). In Europe and the USA, the time-dependent AUROC for predicting 5-year LREs was 0.874 (95% CI 0.820-0.918) for LSM and 0.875 (95% CI 0.841-0.905) for histology. The integrated AUROCs were also comparable between LSM and fibrosis stage, at 0.890 (95% CI 0.833-0.929) and 0.850 (95% CI 0.795-0.894), respectively (p=0.107) (Supplementary Figure 16B, http://links.lww.com/HEP/K343).

The median date of VCTE measurement in the cohort was May 4, 2016. Among patients who underwent VCTE before this date, the cumulative incidence of LREs increased progressively with higher LSM cut-offs, with 0.1% (0.0%-0.5%) for LSM below 10 kPa, 3.2% (1.8%-5.1%) for LSM 10 to 20 kPa, and 15.6% (10.9%-21.0%) for LSM above 20 kPa (p<0.001) (Supplementary Figure 17A, http://links.lww.com/HEP/K343). Similarly, the incidence rate was highest in patients with F4 fibrosis at 15.9% (11.4%-21.1%), followed by 2.0% (0.8%-4.2%) in F3, and 0.4% (0.2%-1.0%) in F0-2 (p<0.001) (Supplementary Figure 17B, http://links.lww.com/HEP/K343). These patterns remained consistent among patients who underwent VCTE after the median date, with cumulative incidence rates similar to those observed before (Supplementary Figures 18A and B, http://links.lww.com/HEP/K343). The time-dependent AUROC for predicting 5-year LREs was 0.901 (95% CI 0.860-0.938) for LSM and 0.879 (95% CI 0.836-0.914) for histology among patients who underwent VCTE before the median date. The corresponding values were also comparable for those who underwent VCTE after the median date, at 0.818 (95% CI 0.712-0.914) for LSM and 0.878 (95% CI 0.829-0.918) for histology. The integrated AUROC for LSM was 0.903 (95% CI 0.831-0.953) and 0.840 (95% CI 0.766-0.906) for patients who had VCTE before and after the median date, respectively (Supplementary Figures 19A and B, http://links.lww.com/HEP/K343). Compared with liver histology, the integrated AUROC values were 0.863 (95% CI 0.791-0.916, p=0.106) for those before and 0.844 (95% CI 0.783-0.899, p=0.885) for those after.

When comparing the first and second LSM values within 18 months, a positive correlation was observed (r=0.788, 95% CI 0.765-0.808, p<0.001) (Supplementary Figure 20, http://links.lww.com/HEP/K343). The mean difference between the two measurements was 0.58 kPa (95% CI 0.26-0.90, p<0.001), with wide limits of agreement ranging from -10.3 kPa to 11.4 kPa (Supplementary Figure 21, http://links.lww.com/HEP/K343). The average and lower LSM values between first and second VCTE measurements within an 18-month interval showed a strong correlation (r=0.966, 95% CI 0.962-0.970, p<0.001) (Supplementary Figure 22, http://links.lww.com/HEP/K343). On average, the difference between them was 1.68 kPa (95% CI 1.56-1.81, p<0.001), with upper limit of agreement to 6.03 kPa (Supplementary Figure 23, http://links.lww.com/HEP/K343). When categorized by average LSM, the cumulative incidence rates increased stepwise across cut-offs, with 0.6% (0.2%-1.8%) for less than 10 kPa, 3.1% (1.4%-5.9%) for between 10 and 20 kPa, and 11.3% (6.0%-19.3%) for above 20 kPa (Supplementary Figure 24A, http://links.lww.com/HEP/K343). Similar patterns were observed when stratified by lower LSM, with 0.9% (0.3%-2.0%) for below 10 kPa, 3.6% (1.5%-7.1%) for between 10 and 20 kPa, and 14.7% (7.3%-24.5%) for greater than 20 kPa (Supplementary Figure 24B, http://links.lww.com/HEP/K343). The 5-year time-dependent AUROCs for predicting LREs were 0.796 (95% CI 0.654-0.923) for average LSM and 0.777 (95% CI 0.613-0.918) for lower LSM, compared with 0.875 (95% CI 0.806-0.931) for histology. The integrated AUROCs were 0.838 (95% CI 0.701-0.925) for average LSM and 0.823 (95% CI 0.692-0.917) for lower LSM, with p-values of 0.700 and 0.516, respectively, compared with histology (Supplementary Figures 25A and B, http://links.lww.com/HEP/K343).

DISCUSSION

In this analysis of a large multicenter longitudinal cohort of patients with MASLD, LSM by VCTE demonstrated prognostic performance comparable to fibrosis stage by liver histology in predicting LREs, including hepatic decompensation, HCC, liver transplant, and liver-related mortality. The two methods achieved a comparable discriminatory performance and risk stratification, and their similar predictive accuracy was further supported by the integrated Brier score and IDI. These findings remained consistent across both primary and secondary outcomes, as well as all investigated time points and sensitivity analysis. Our results add to the growing

evidence supporting non-invasive tools like LSM by VCTE as a feasible, affordable, and accessible alternative to histology for risk prediction of liver-related clinical outcomes, with potential implications for clinical practice and drug development.

We found that LSM and histology demonstrated nearly identical discriminatory performance, as reflected in AUROC, AUPRC, and the integrated Brier score, for predicting LREs, as well as in separate analyses for hepatic decompensation and HCC. These results align with existing literature, validating LSM by VCTE as a predictor of clinical outcomes in chronic liver disease,[8,17,18] and supporting the EASL-EASD-EASO Clinical Practice Guidelines' recommendation to use non-invasive tools for prognostic monitoring in MASLD.[19] Additionally, sensitivity analyses suggest that this approach could be applied in both Asian and Western populations, and that the validity of LSM is unlikely to be affected by intra-individual variability or by technical advancements.

The Baveno VII consensus has recommended a rule of 5 (10-15-20-25 kPa), using VCTE, to progressively identify the risk of clinically significant portal hypertension and decompensation.[7] Specifically, the recommendation included a LSM value < 10 kPa indicating an exclusion from compensated advanced chronic liver disease, while a value ≥ 20 kPa necessitating an endoscopic screening for varices, a major complication from portal hypertension. In compliance with the Baveno VII consensus, we applied the cut-offs of 10 and 20 kPa to determine their prognostic utility. We found that risk stratification using these thresholds simulated histologic staging (F0–F2, F3, F4), with 5-year LRE incidences of 0.3%, 3.3%, and 14.4% for LSM versus 0.4%, 3.0%, and 16% for histology. Traditionally, LSM was considered inaccurate when it suggested a different degree of fibrosis than histology. However, in our stratified analysis by histological fibrosis stage, LSM at the cutoffs of 10 and 20 kPa continued to identify patients with varying risks of LREs (Supplementary Figure 3, http://links.lww.com/HEP/K343). In patients with F0-F2 fibrosis, the 5-year cumulative incidence of LREs was below 2% regardless of LSM, suggesting that the high LSM was still largely false positives. However, in patients with F3 or F4 fibrosis, the 5-year cumulative incidence of LREs could differ by more than 2-fold when LSM was below 10 kPa versus over 20 kPa. This may suggest that LSM could reflect physical properties beyond fibrosis stage alone and provide additional prognostic value. Meanwhile,

variables like BMI and diabetes status should be taken in consideration when there is misclassification. Furthermore, the prognostic performance metrics (e.g., sensitivity, specificity) were highly comparable between LSM and histology across all outcomes and timepoints. The likelihood ratio and the post-test probability also indicated 10 kPa as the target threshold for reduced risk for LREs and identified a significantly increased risk when LSM reaches 20 kPa. Specifically, a LSM of 10 kPa had a likelihood ratio of 0.37, corresponding to a 59.5% reduction in the 5-year post-test probability of LREs for LSM <10 kPa, assuming a pre-test probability of 0.2. Conversely, a LSM of 20 kPa yielded a likelihood ratio of 2.83, indicating at least a 2.32-fold increase in the post-test probability of developing LREs within 5 years for LSM ≥20 kPa even with a pre-test probability of 0.01. These two cut-offs helped to identify patients who could be excluded from intensive intervention and those who need to be prioritized for advanced care. Similar findings were reported in an individual data meta-analysis comparing LSM and histology adopting the same cut-offs.[8] Collectively, the cut-offs, 10 kPa and 20 kPa, are not only suggested to rule-in advance fibrosis and cirrhosis, respectively,[3] but we also validated in the prognostic settings, further helping to reduce the need of liver biopsy in assessing disease progression and regression.

Fibrosis severity, assessed by liver biopsy, has been well established as an independent predictor of prognosis in chronic liver disease,[20] and being the only accepted method to determine MASH resolution and fibrosis improvement in the drug development for MASH.[21] However, our study demonstrated strong concordance between LSM and histology in LRE prediction, in terms of the discriminatory and prognostic performance and calibration and reclassification metrics, suggesting LSM could potentially serve as a non-invasive surrogate endpoint in clinical trials. Additionally, previous studies showed that the dynamic changes in LSM values correlated with likelihood for LREs, reflecting its responsiveness to disease severity, a critical feature for evaluating treatment efficacy. Although our study did not assess longitudinal LSM changes, prior research found that a 30% or greater reduction in LSM, regardless of baseline value, was associated with reduced LREs risk.[5] Furthermore, longitudinal change in LSM has also been identified as independently associated with hepatic decompensation, HCC, overall mortality, and liver-related mortality.[22] Altogether, LSM by VCTE not only mirrored histological staging and

reflected dynamic changes but also translated these findings into clinical benefits, demonstrating significant potential as a surrogate for treatment response.

The main strength of this study was the direct comparison between LSM by VCTE and liver histology for predicting LREs within the same group of participants, conducted in a large, multicenter cohort spanning North America, Europe, and Asia, with a relatively high number of observed LREs. However, the study also has limitations. Although the median follow-up duration was 55.8 months, this may still be considered relatively short to fully capture the long natural history of chronic liver disease progression to cirrhosis and its complications.[23] Nonetheless, current guidelines typically recommend evaluation of MASLD at intervals of 1 to 3 years, the timeframe evaluated in this study was clinically relevant.[19,24] Moreover, as this was a natural history cohort, the results do not reflect the performance of liver histology and LSM in patients receiving pharmacological treatment for MASH. Another limitation is that the data were collected from tertiary referral centers, which may not be representative of the general population. Therefore, future studies should aim to evaluate the prognostic performance of LSM in broader populations. Additionally, although current evidence supports an association between longitudinal changes in LSM and LRE risk, future research should focus on direct comparisons between changes in LSM and histologic fibrosis stage over time to determine whether non-invasive measurements mirror histological changes, and whether such changes translate into meaningful clinical outcomes.

CONCLUSION

In this large multicenter cohort, we demonstrated that LSM by VCTE had prognostic performance very similar to that of liver histology in predicting liver-related clinical outcomes, including in analyses of hepatic decompensation and HCC among patients with MASLD. In line with current recommendations, our findings further support the use of non-invasive tools to assess LRE risk. Moreover, the nearly identical performance suggests that LSM by VCTE could serve as a more feasible and safer surrogate than liver biopsy, potentially facilitating the conduct of clinical trials for MASH drug development.

ACKNOWLEDGMENTS

Assistance with the study: None

Presentation: None

REFERENCES

1. Wong VW, Ekstedt M, Wong GL, Hagstrom H. Changing epidemiology, global trends and implications for outcomes of NAFLD. *J Hepatol*. Sep 2023;79(3):842-852. doi:10.1016/j.jhep.2023.04.036
2. Sanyal AJ, Castera L, Wong VW. Noninvasive Assessment of Liver Fibrosis in NAFLD. *Clin Gastroenterol Hepatol*. Jul 2023;21(8):2026-2039. doi:10.1016/j.cgh.2023.03.042
3. Mozes FE, Lee JA, Selvaraj EA, Jayaswal ANA, Trauner M, Boursier J, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. *Gut*. May 2022;71(5):1006-1019. doi:10.1136/gutjnl-2021-324243
4. Vali Y, Lee J, Boursier J, Petta S, Wonders K, Tiniakos D, et al. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. *The Lancet Gastroenterology & Hepatology*. 2023;8(8):714-725. doi:10.1016/S2468-1253(23)00017-1
5. Lin H, Lee HW, Yip TC, Tsochatzis E, Petta S, Bugianesi E, et al. Vibration-Controlled Transient Elastography Scores to Predict Liver-Related Events in Steatotic Liver Disease. *JAMA*. Apr 16 2024;331(15):1287-1297. doi:10.1001/jama.2024.1447
6. Gawrieh S, Vilar-Gomez E, Wilson LA, Pike F, Kleiner DE, Neuschwander-Tetri BA, et al. Increases and decreases in liver stiffness measurement are independently associated with the risk of liver-related events in NAFLD. *J Hepatol*. Oct 2024;81(4):600-608. doi:10.1016/j.jhep.2024.05.008
7. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno VIIF. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol*. Apr 2022;76(4):959-974. doi:10.1016/j.jhep.2021.12.022
8. Mozes FE, Lee JA, Vali Y, Alzoubi O, Staufer K, Trauner M, et al. Performance of non-invasive tests and histology for the prediction of clinical outcomes in patients with non-alcoholic fatty liver disease: an individual participant data meta-analysis. *Lancet Gastroenterol Hepatol*. Aug 2023;8(8):704-713. doi:10.1016/S2468-1253(23)00141-3
9. Wong VW, Chitturi S, Wong GL, Yu J, Chan HL, Farrell GC. Pathogenesis and novel treatment options for non-alcoholic steatohepatitis. *Lancet Gastroenterol Hepatol*. Sep 2016;1(1):56-67. doi:10.1016/S2468-1253(16)30011-5

10. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. Jun 2005;128(7):1898-906. doi:10.1053/j.gastro.2005.03.084

11. Anania FA, Hager R, Higgins K, Makar GA, Siegel J, Tran TT. Non-Invasive Tests: Establishing efficacy for metabolic dysfunction associated steatohepatitis beyond the biopsy-current perspectives from the division of hepatology and nutrition, US Food and Drug Administration. *Hepatology*. Aug 26 2025;doi:10.1097/HEP.0000000000001509

12. Wong VW, Irles M, Wong GL, Shili S, Chan AW, Merrouche W, et al. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. *Gut*. Nov 2019;68(11):2057-2064. doi:10.1136/gutjnl-2018-317334

13. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. Dec 30 2013;32(30):5381-97. doi:10.1002/sim.5958

14. Yuan Y, Zhou QM, Li B, Cai H, Chow EJ, Armstrong GT. A threshold-free summary index of prediction accuracy for censored time to event data. *Stat Med*. May 10 2018;37(10):1671-1681. doi:10.1002/sim.7606

15. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. Jan 15 2011;30(1):11-21. doi:10.1002/sim.4085

16. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med*. Jun 30 2013;32(14):2430-42. doi:10.1002/sim.5647

17. Pang JX, Zimmer S, Niu S, Crotty P, Tracey J, Pradhan F, et al. Liver stiffness by transient elastography predicts liver-related complications and mortality in patients with chronic liver disease. *PLoS One*. 2014;9(4):e95776. doi:10.1371/journal.pone.0095776

18. 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*. Sep 2016;65(3):570-8. doi:10.1016/j.jhep.2016.04.023

19. European Association for the Study of the L, European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines on

the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J Hepatol*. Sep 2024;81(3):492-542. doi:10.1016/j.jhep.2024.04.031

20. Ekstedt M, Hagstrom H, Nasr P, Fredrikson M, Stal P, Kechagias S, Hultcrantz R. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology*. May 2015;61(5):1547-54. doi:10.1002/hep.27368

21. Dufour JF, Wong VW. Non-invasive assessment of MASH resolution. *Gut*. Jul 11 2024;73(8):1227-1228. doi:10.1136/gutjnl-2024-332060

22. Petta S, Sebastiani G, Vigano M, Ampuero J, Wai-Sun Wong V, Boursier J, et al. Monitoring Occurrence of Liver-Related Events and Survival by Transient Elastography in Patients With Nonalcoholic Fatty Liver Disease and Compensated Advanced Chronic Liver Disease. *Clin Gastroenterol Hepatol*. Apr 2021;19(4):806-815 e5. doi:10.1016/j.cgh.2020.06.045

23. Zhang X, Yip TC, Tse YK, Hui VW, Li G, Lin H, et al. Duration of type 2 diabetes and liver-related events in nonalcoholic fatty liver disease: A landmark analysis. *Hepatology*. Dec 1 2023;78(6):1816-1827. doi:10.1097/HEP.0000000000000432

24. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology*. May 1 2023;77(5):1797-1835. doi:10.1097/HEP.0000000000000323

Figure 1. Patient flowchart

Abbreviations: HCC, hepatocellular carcinoma. MASLD, metabolic dysfunction-associated steatotic liver disease. LSM, liver stiffness measurement. VCTE, vibration-controlled transient elastography.

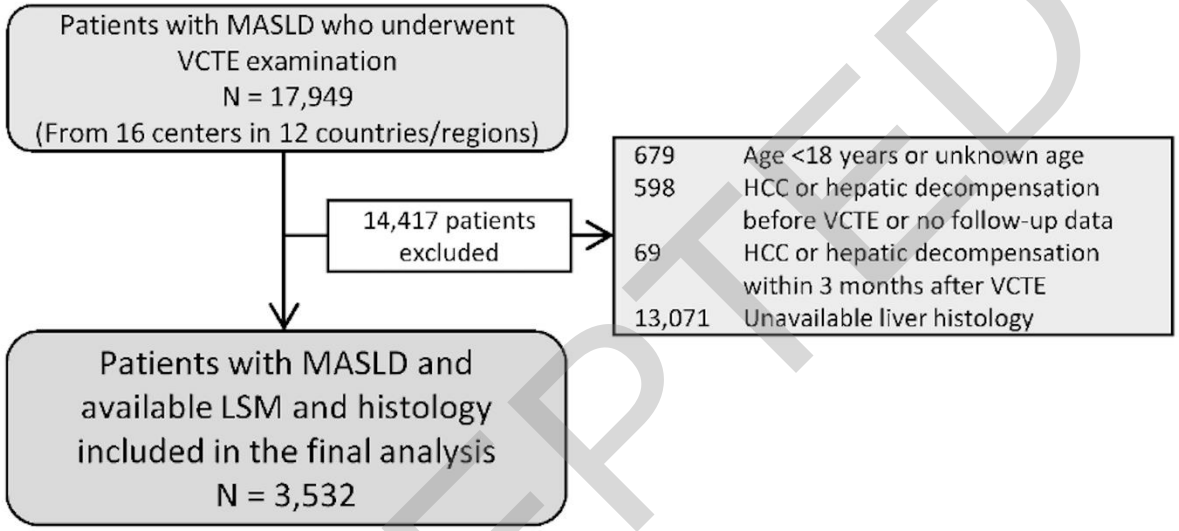
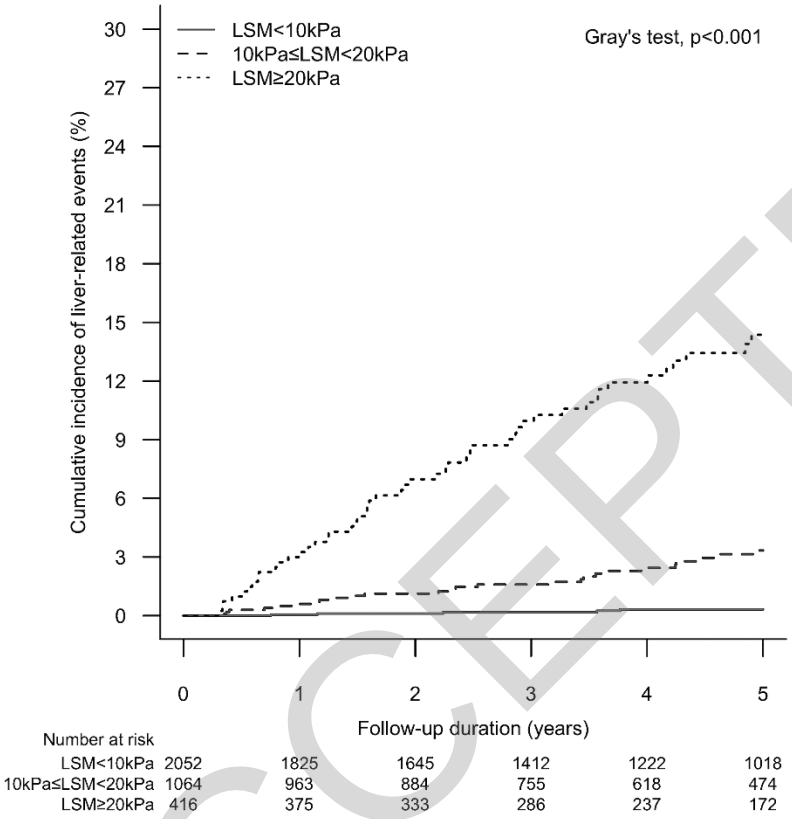


Figure 2. Cumulative incidence of liver-related events in A. patients with LSM <10 kPa, LSM 10-19.9 kPa, and LSM ≥20 kPa; and B. patients with F0-2, F3, and F4 by histology.



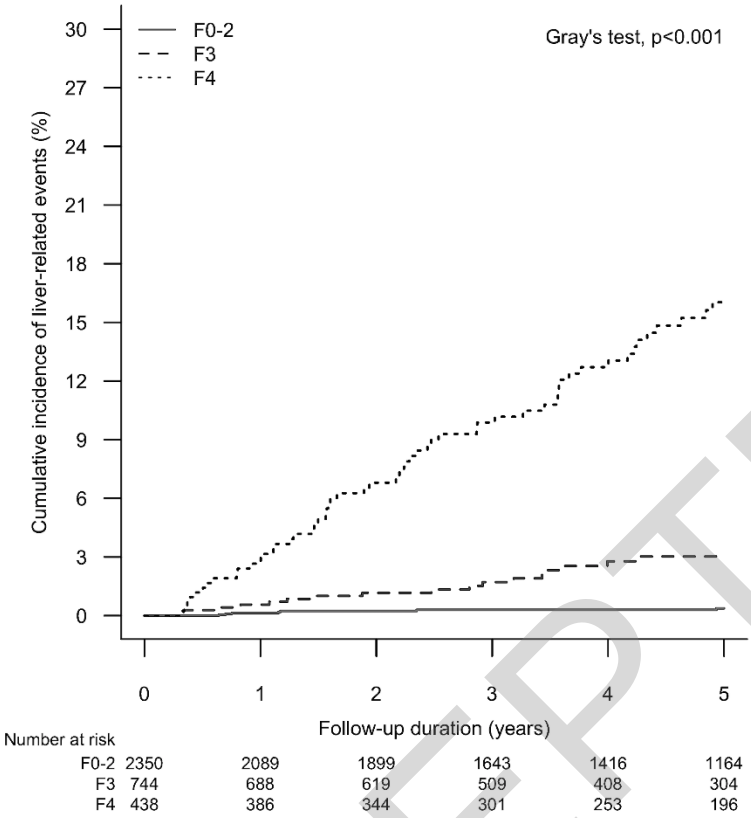
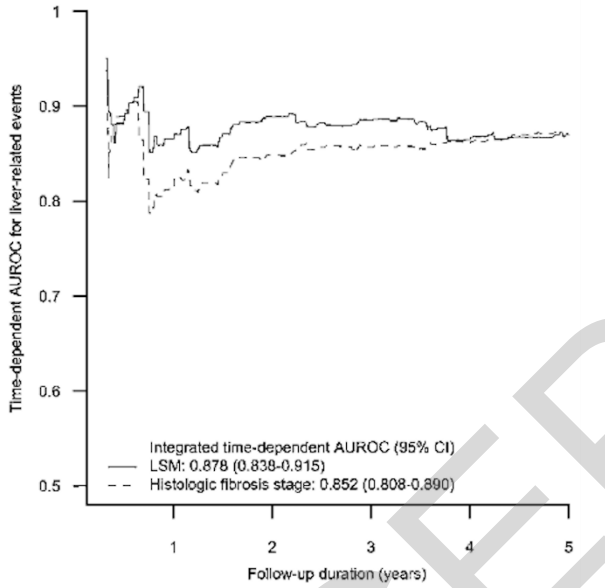
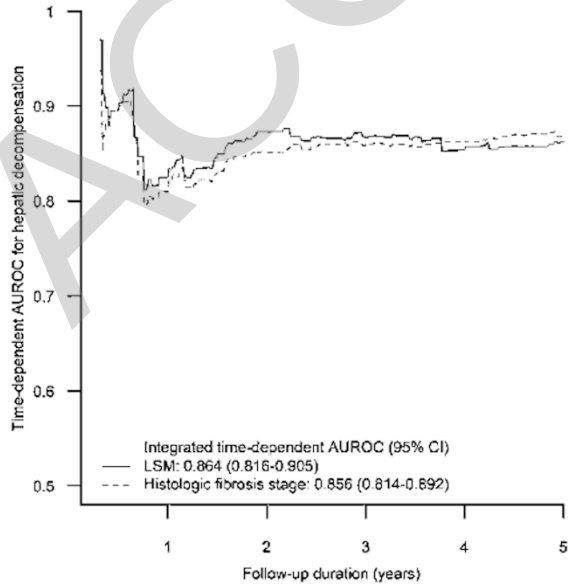


Figure 3. Integrated time-dependent area under the receiver operating characteristic curve (AUROC) for liver stiffness measurement (LSM) and histological fibrosis stage for predicting A. liver-related events, B. hepatic decompensation, and C. hepatocellular carcinoma (HCC) over 5 years.





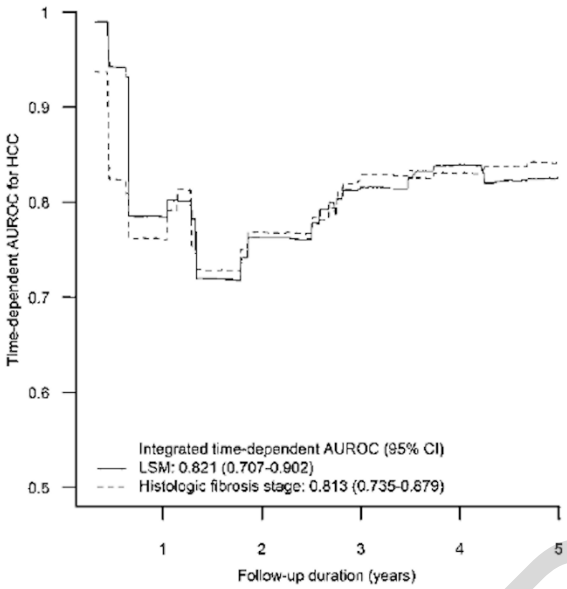
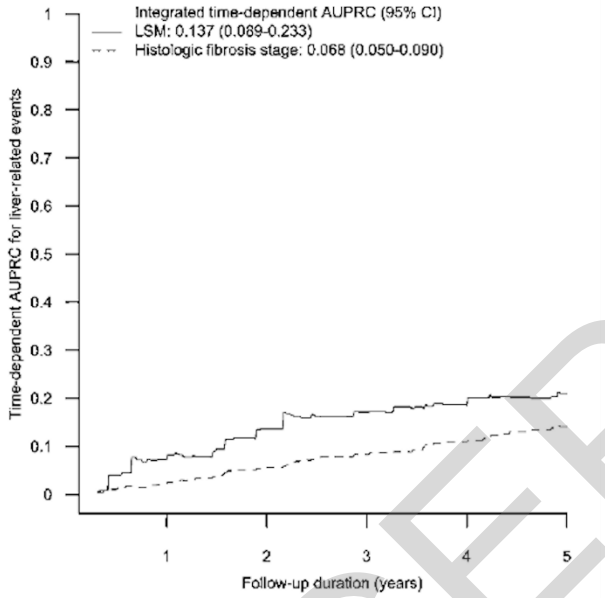
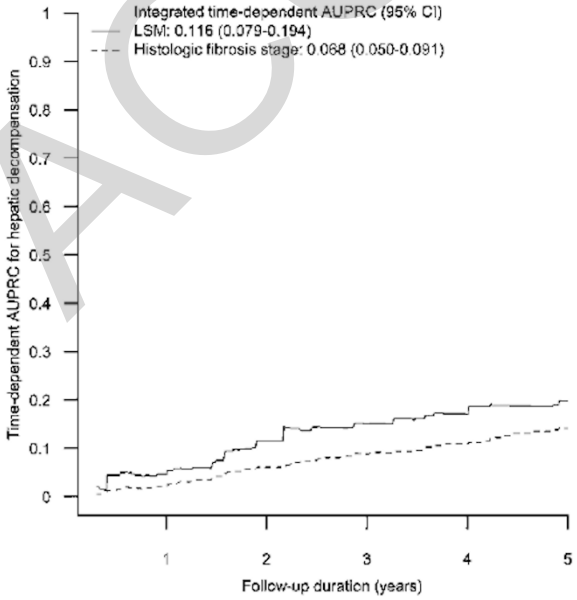


Figure 4. Integrated time-dependent area under the precision-recall curve (AUPRC) for liver stiffness measurement (LSM) and histological fibrosis stage for predicting A. liver-related events, B. hepatic decompensation, and C. hepatocellular carcinoma (HCC) over 5 years.





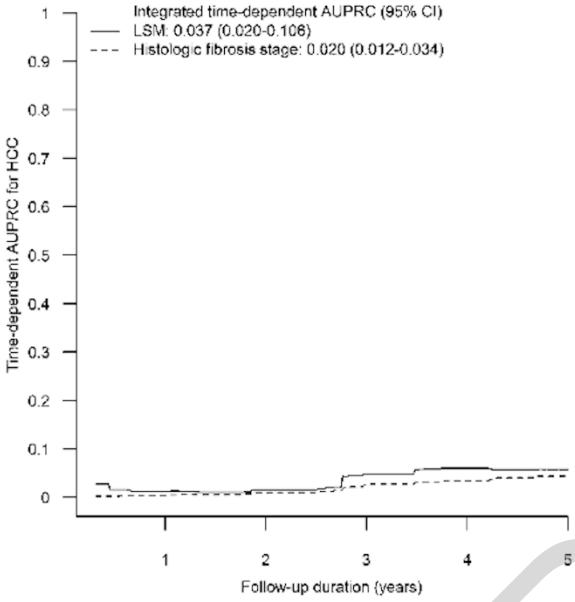


Table 1. Clinical characteristics of the cohort of patients with available liver stiffness measurement (LSM) and liver histology.

Characteristics	All patients N = 3,532
Age (years)	51.9 (13.3)
Female sex, n (%)	1509 (42.7)
Patients collected prospectively, n (%)	3086 (87.4)
BMI (kg/m^2)	29.7 (26.5 to 33.5)
Diabetes, n (%)	1,471 (41.7)
Hypertension, n (%)	1,639 (46.4)
ALT (IU/L)	54 (35 to 85)
AST (IU/L)	41 (29 to 61)
GGT (IU/L)	60 (37 to 110)
Albumin (g/L)	44.1 (4.7)
Total bilirubin (µmol/L)	10.9 (8.0 to 15.3)
Platelets (×10^9/L)	231 (186 to 276)
Creatinine (µmol/L)	71 (60 to 82)
FibroScan	
Liver stiffness measurement (kPa)	8.8 (6.1 to 13.3)
Controlled attenuation parameter (dB/m)	320 (287 to 352)
Non-invasive tests	
Fibrosis-4 index	1.27 (0.80 to 2.06)
Agile 3+	0.33 (0.09 to 0.74)
Agile 4	0.04 (0.01 to 0.18)
FibroScan-AST	0.53 (0.31 to 0.71)
Fibrosis stage[a]	
0	576 (16.3)
1	1,189 (33.7)

2	585 (16.6)
3	744 (21.1)
4	438 (12.4)
Follow-up duration (months)	56.6 (30.5 to 91.2)
Follow-up duration for patients with events (months)	57.0 (31.1 to 91.7)
Follow-up duration for patients without events (months)	42.8 (18.2 to 76.4)
Time interval between VCTE and liver biopsy (months)	0.93 (0 to 7.13)

Data were presented as n (%), mean (standard deviation), or median (25^{th}-75^{th} percentile).

[a] Fibrosis stage (0-4) according to the NASH CRN system. Stage 0, no fibrosis; Stage 1, centrilobular pericellular fibrosis; Stage 2: centrilobular and periportal fibrosis; Stage 3: bridging fibrosis; Stage 4, cirrhosis.

Liver stiffness measurement by vibration-controlled transient elastography is a non-invasive method, which helps in assessing the extent of fibrosis; Controlled attenuation parameter quantifies liver steatosis non-invasively, by measuring the attenuation of ultrasound waves through the liver, which helps in assessing liver steatosis.

Abbreviations: AST, aspartate aminotransferase. ALT, alanine aminotransferase. BMI, body-mass index. GGT, gamma-glutamyl transpeptidase.

Table 2. Paired comparisons of the liver stiffness measurement (LSM) versus histological fibrosis stage on the discriminatory performance for predicting 2-year, 3-year and 5-year liver-related events (N=3,532).

Liver-related events						
Years of follow-up	**AUROC (95% CI) of histological fibrosis stage**	**AUROC (95% CI) of LSM**	**p value**	**AUPRC (95% CI) of histological fibrosis stage**	**AUPRC (95% CI) of LSM**	**p value**
2 years	0.848 (0.798-0.894)	0.889 (0.838-0.934)	0.028	0.056 (0.034-0.079)	0.136 (0.075-0.246)	0.047
3 years	0.856 (0.815-0.893)	0.885 (0.848-0.923)	0.096	0.083 (0.056-0.114)	0.171 (0.103-0.279)	0.034
5 years	0.869 (0.839-0.895)	0.870 (0.834-0.905)	0.915	0.141 (0.103-0.181)	0.210 (0.143-0.312)	0.068
Hepatic decompensation						
Years of follow-up	**AUROC (95% CI) of histological fibrosis stage**	**AUROC (95% CI) of LSM**	**p value**	**AUPRC (95% CI) of histological fibrosis stage**	**AUPRC (95% CI) of LSM**	**p value**
2 years	0.851 (0.805-0.895)	0.873 (0.818-0.923)	0.241	0.060 (0.039-0.086)	0.115 (0.068-0.206)	0.079
3 years	0.859 (0.821-	0.867 (0.825-0.909)	0.619	0.088 (0.061-0.120)	0.150 (0.094-	0.081

	0.895)				0.250)	
5 years	0.868 (0.839-0.894)	0.862 (0.825-0.899)	0.732	0.140 (0.104-0.182)	0.196 (0.134-0.297)	0.112
Hepatocellular carcinoma						
Years of follow-up	**AUROC (95% CI) of histological fibrosis stage**	**AUROC (95% CI) of LSM**	**p value**	**AUPRC (95% CI) of histological fibrosis stage**	**AUPRC (95% CI) of LSM**	**p value**
2 years	0.768 (0.631-0.879)	0.762 (0.569-0.913)	0.922	0.010 (0.003-0.020)	0.015 (0.005-0.041)	0.500
3 years	0.829 (0.756-0.888)	0.816 (0.720-0.894)	0.674	0.028 (0.014-0.047)	0.048 (0.020-0.168)	0.564
5 years	0.842 (0.790-0.884)	0.826 (0.760-0.884)	0.548	0.044 (0.025-0.069)	0.057 (0.031-0.146)	0.637

Abbreviations: AUROC, area under the receiver operating characteristic curve. AUPRC, area under the precision-recall curve. LSM, liver stiffness measurement.

Table 3. Prognostic performance of LSM and histology for the development of liver-related events at 2, 3, and 5 years

Category	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	LR + (95% CI)	LR – (95% CI)	Overall accuracy (%) (95% CI)
2 years							
Histology							
F3-4 (vs F0-2)	87.8(77.3-97.3)	65.9 (64.2-67.5)	3.2 (2.1-4.3)	99.8 (99.5-100.0)	2.6 (2.2-2.9)	0.2 (0.0-0.3)	66.2 (64.5-67.8)
F4 (vs F0-3)	67.9 (53.3-81.6)	87.7(86.5-88.9)	6.6 (4.3-8.9)	99.5 (99.3-99.8)	5.5(4.3-6.8)	0.4(0.2-0.5)	87.4(86.2-88.6)
LSM							
≥10 kPa (*vs* <10kPa)	92.7 (83.6-100.0)	58.0 (56.3-59.7)	2.8 (1.9-3.6)	99.8 (99.6-100.0)	2.2 (2.0-2.4)	0.1 (0.0-0.3)	58.6 (86.9-89.2)
≥20 kPa (*vs* <20kPa)	68.0 (53.1-82.8)	88.4 (87.2-89.5)	7.0 (4.6-9.6)	99.5 (99.3-99.8)	5.9 (4.4-7.4)	0.4 (0.2-0.5)	88.1 (86.9-89.2)
3 years							
Histology							
F3-4 (vs F0-2)	89.6 (81.4-96.7)	66.1 (64.4-67.9)	4.7 (3.4-6.1)	99.7 (99.5-99.9)	2.6 (2.3-2.9)	0.2 (0.1-0.3)	66.7 (64.9-68.4)
F4 (vs F0-3)	69.5 (57.2-81.6)	87.3 (86.0-88.6)	9.3 (6.5-12.4)	99.4 (99.1-99.7)	5.5 (4.4-6.6)	0.3 (0.2-0.5)	86.9 (85.6-88.2)
LSM							

	≥10 kPa (*vs* <10kPa)	93.0 (85.4-98.6)	57.9 (56.2-59.8)	4.0 (2.9-5.0)	99.8 (99.5-99.9)	2.2 (2.0-2.4)	0.1 (0.0-0.3)	58.8 (57.1-60.6)
	≥20 kPa (*vs* <20kPa)	68.0 (55.9-79.7)	88.3 (86.9-89.6)	9.8 (6.8-12.8)	99.3 (99.0-99.6)	5.8 (4.6-7.1)	0.4 (0.2-0.5)	87.8 (86.4-89.1)
5 years								
Histology								
	F3-4 (vs F0-2)	91.9 (85.6-96.9)	68.0 (65.9-69.9)	8.3 (6.4-10.2)	99.6 (99.3-99.9)	2.9 (2.6-3.1)	0.1 (0.0-0.2)	68.8 (66.8-70.7)
	F4 (vs F0-3)	70.9 (61.5-80.6)	87.0 (85.6-88.5)	14.7 (11.1-18.5)	99.0 (98.6-99.3)	5.5 (4.5-6.5)	0.3 (0.2-0.4)	86.4 (84.9-87.9)
LSM								
	≥10 kPa (*vs* <10kPa)	92.9 (86.9-98.2)	60.8 (58.7-62.9)	6.9 (5.4-8.5)	99.6 (99.3-99.9)	2.4 (2.2-2.6)	0.1 (0.0-0.2)	62.2 (60.1-64.2)
	≥20 kPa (*vs* <20kPa)	60.5 (50.1-70.8)	89.2 (87.9-90.5)	14.9 (10.8-19.3)	98.6 (98.2-99.1)	5.6 (4.5-6.9)	0.4 (0.3-0.6)	88.0 (86.7-89.3)

Abbreviations: LSM, liver stiffness measurement. NLR, Negative Likelihood Ratio. NPV, Negative Predictive Value. PLR, Positive Likelihood Ratio. PPV, Positive Predictive Value

Table 4. Paired comparisons of the liver stiffness measurement (LSM) versus histological fibrosis stage for predicting 2-, 3-, and 5-year liver-related events, hepatic decompensation and hepatocellular carcinoma by the categorical net reclassification improvement (NRI) and integrated discrimination improvement index (IDI).

2-year liver-related events				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.073 (0.000 – 0.162)	-0.085 (-1.004 - -0.068)	-0.011 (-0.085 – 0.077)	0.035 (-0.002 – 0.096)
3-year liver-related events				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.050 (-0.024 – 0.134)	-0.085 (-1.01 - -0.069)	-0.035 (-0.111 – 0.047)	0.034 (-0.010 – 0.093)
5-year liver-related events				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.018 (-0.060 - 0.094)	-0.086 (-0.103 - -0.071)	-0.068 (-0.148 – 0.009)	0.010 (-0.033 – 0.065)

2-year hepatic decompensation				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.043 (-0.137 - 0.214)	-0.075 (-0.093 - -0.057)	-0.031 (-0.213 - 0.141)	0.014 (-0.014 - 0.055)
3-year hepatic decompensation				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	-0.009 (-0.167 - 0.156)	-0.076 (-0.094 - -0.058)	-0.085 (-0.241 - 0.082)	0.012 (-0.023 - 0.060)
5-year hepatic decompensation				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	-0.072 (-0.212 - 0.060)	-0.079 (-0.098 - -0.060)	-0.151 (-0.294 - -0.018)	-0.005 (-0.043 - 0.041)
2-year hepatocellular carcinoma				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.256	-0.074	0.182	-0.003

	(-0.124 - 0.650)	(-0.092 - -0.056)	(-0.192 - 0.578)	(-0.010 - 0.001)
3-year hepatocellular carcinoma				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	0.060 (-0.214 - 0.333)	-0.074 (-0.093 - -0.056)	-0.014 (-0.290 - 0.254)	0.005 (-0.015 - 0.059)
5-year hepatocellular carcinoma				
Model	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**IDI (95% CI)**
Histological fibrosis stage	Reference	Reference	Reference	Reference
LSM	-0.027 (-0.272 - 0.239)	-0.075 (-0.095 - -0.058)	-0.102 (-0.346 - 0.167)	-0.002 (-0.026 - 0.046)