Received: 8 July 2022 | Accepted: 8 July 2023

DOI: 10.1097/HEP.0000000000000542

ORIGINAL ARTICLE

OPEN

# Serum identification of at-risk MASH: The metabolomics-advanced steatohepatitis fibrosis score (MASEF)

**Mazen Noureddin[1,2] | Emily Truong[3] | Rebeca Mayo[4] | Ibon Martínez-Arranz[4] | Itziar Mincholé[4] | Jesus M. Banales[5,6] | Marco Arrese[7] | Kenneth Cusi[8] | María Teresa Arias-Loste[9] | Radan Bruha[10] | Manuel Romero-Gómez[11] | Paula Iruzubieta[9] | Rocio Aller[12] | Javier Ampuero[13] | José Luis Calleja[14] | Luis Ibañez-Samaniego[15] | Patricia Aspichueta[16,17,18] | Antonio Martín-Duce[19] | Tatyana Kushner[20] | Pablo Ortiz[4] | Stephen A. Harrison[21] | Quentin M. Anstee[22] | Javier Crespo[9] | José M. Mato[18,23] | Arun J. Sanyal[24]**

[1]Houston Methodist Hospital, Houston Research Institute Houston, Texas, USA

[2]Houston Research Institute, Houston, Texas, USA

[3]Karsh Division of Gastroenterology and Hepatology, Cedars-Sinai Medical Center, Los Angeles, CA

[4]OWL Metabolomics, Derio, Spain

[5]Biodonostia Research Institute, Donostia University Hospital, University of the Basque Country (UPV-EHU), CIBERehd, IKERBASQUE, Donostia, Spain

[6]Department of Biochemistry and Genetics, School of Sciences, University of Navarra, Pamplona, Spain

[7]Department of Gastroenterology, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[8]University of Florida, Gainesville, Florida, USA

[9]Marqués de Valdecilla University Hospital, Cantabria University, IDIVAL, Santander, Spain

[10]General University Hospital and the First Faculty of Medicine, Charles University, Prague, Czech Republic

[11]Valme University Hospital, CIBERehd, Seville, Spain

[12]Clinic University Hospital, University of Valladolid, Valladolid, Spain

[13]Virgen del Rocío University Hospital, Sevilla, Spain

[14]Puerta del Hierro University Hospital, Madrid, Spain

[15]Gregorio Marañón University Hospital, Madrid, Spain

[16]Department of Physiology, Faculty of Medicine and Nursing, University of the Basque Country UPV/EHU, Leioa, Spain

[17]Biocruces Bizkaia Health Research Institute, Barakaldo, Spain

[18]National Institute for the Study of Liver and Gastrointestinal Diseases (CIBERehd, Instituto de Salud Carlos III), Madrid, Spain

[19]Príncipe de Asturias University Hospital, Alcalá University, Madrid, Spain

[20]Division of Liver Diseases, Icahn School of Medicine at Mount Sinai, New York City, New York, USA

[21]Pinnacle Clinical Research, San Antonio, Texas, USA

[22]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[23]CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain

[24]Virginia Commonwealth University Medical Center, Richmond, Virginia, USA

**Abbreviations:** ≥F2, significant fibrosis; ≥F3, advanced fibrosis; ≥F4, fibrosis stage 4 or cirrhosis; AASLD, American Association for the Study of Liver Diseases; AGA, American Gastroenterology Association; ALT, alanine aminotransferase; AST, aspartate aminotransferase; at-risk MASH, MASH + NAS ≥ 4 + significant fibrosis (≥F2); AUC, area under the ROC curve; BMI, body mass index; EASL, European Association for the Study of Liver Diseases; F0, fibrosis stage 0; F1, fibrosis stage 1; F2, fibrosis stage 2; F3, fibrosis stage 3; FAST, FibroScan-AST; FIB-4, fibrosis-4 index; GGT, gamma-glutamyl transferase; HbA1c, glycosylated hemoglobin; HOMA, homeostatic model assessment; LSM, liver stiffness measurement; MASEF, Metabolomics-Advanced StEatohepatitis Fibrosis Score; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver-operating characteristic; VCTE, vibration-controlled transient elastography.

Mazen Noureddin, Emily Truong, and Rebeca Mayo contributed equally.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepjournal.com.

**Correspondence**
Mazen Noureddin, Houston Methodist Hospital, Director Houston Research Institute, 1155 Dairy Ashford, Houston, Texas 77079, USA.
Email: noureddinmd@houstonresearchinstiute.com

Arun Sanyal, Stravitz-Sanyal Institute of Liver Disease and Metabolic Health, Virginia Commonwealth University, and Division of Gastroenterology, Hepatology, and Nutrition, VCU School of Medicine, Richmond VA, Box 980341, Virginia Commonwealth University, Richmond, VA, 804-828-4060, USA.
Email: Arun.sanyal@vcuhealth.org

## Abstract

**Background:** Early identification of those with NAFLD activity score ≥ 4 and significant fibrosis (≥F2) or at-risk metabolic dysfunction-associated steatohepatitis (MASH) is a priority as these patients are at increased risk for disease progression and may benefit from therapies. We developed and validated a highly specific metabolomics-driven score to identify at-risk MASH.

**Methods:** We included derivation (n = 790) and validation (n = 565) cohorts from international tertiary centers. Patients underwent laboratory assessment and liver biopsy for metabolic dysfunction-associated steatotic liver disease. Based on 12 lipids, body mass index, aspartate aminotransferase, and alanine aminotransferase, the MASEF score was developed to identify at-risk MASH and compared to the FibroScan-AST (FAST) score. We further compared the performance of a FIB-4 + MASEF algorithm to that of FIB-4 + liver stiffness measurements (LSM) by vibration-controlled transient elastography (VCTE).

**Results:** The diagnostic performance of the MASEF score showed an area under the receiver-operating characteristic curve, sensitivity, specificity, and positive and negative predictive values of 0.76 (95% CI 0.72–0.79), 0.69, 0.74, 0.53, and 0.85 in the derivation cohort, and 0.79 (95% CI 0.75–0.83), 0.78, 0.65, 0.48, and 0.88 in the validation cohort, while FibroScan-AST performance in the validation cohort was 0.74 (95% CI 0.68–0.79; $p$ = 0.064), 0.58, 0.79, 0.67, and 0.73, respectively. FIB-4 + MASEF showed similar overall performance compared with FIB-4 + LSM by VCTE ($p$ = 0.69) to identify at-risk MASH.

**Conclusion:** MASEF is a promising diagnostic tool for the assessment of at-risk MASH. It could be used alternatively to LSM by VCTE in the algorithm that is currently recommended by several guidance publications.

## INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD),[1] a name recently adopted by consensus for the disease previously known as NAFLD, is the leading cause of chronic liver disease worldwide that affects 25% of the general population and is the leading cause of cirrhosis and liver transplantation in women.[2,3] The spectrum of MASLD includes isolated steatosis that may progress to metabolic dysfunction-associated steatohepatitis (MASH),[1] previously known as NASH, fibrosis, and eventually cirrhosis or HCC.[4] Patients with significant (fibrosis stage 2 or higher, ≥F2) or advanced fibrosis (fibrosis stage 3 or higher, ≥F3) have been shown to have a higher risk of morbidity and mortality.[5,6] Thus, clinical trials, especially phase 3 registry trials, have focused on MASH with ≥F2, also known as "at-risk MASH."[7]

Liver biopsy remains the gold standard to identify those with at-risk MASH.[4] Given the challenges and risks associated with liver biopsies, recent noninvasive tests, including the serum-based NIS4, FibroScan-AST (FAST), and MRI-AST (MAST) scores, have been developed to identify patients with at-risk MASH.[8–10] The NIS4 score is a serum-based test to identify patients at-risk MASH.[8] The FAST score uses a combination of liver stiffness measurement (LSM) and controlled attenuation parameters by vibration-controlled transient elastography (VCTE) with aspartate aminotransferase (AST),[9] whereas the MAST score uses LSM by magnetic resonance elastography and MRI proton density fat fraction (MRI-PDFF) combined with AST.[10] Both the FAST and MAST scores rely on the availability of VCTE or advanced liver imaging and are more widely used in clinical trials, while the NIS4 remains not as widespread due to its recent approval

and lack of validation outside of its original publication. As the enhanced liver fibrosis test is a serum test identifying those mainly at high risk for progression to cirrhosis or the development of liver-related events, enhanced liver fibrosis test lacks the ability to identify steatohepatitis, the primary disease driver.[11]

Recently, the American Gastroenterology Association (AGA) and the American Association for the Study of Liver Diseases (AASLD) issued a call advocating for MASLD screening in high-risk populations.[12–14] The AGA, AASLD, and the European Association for the Study of Liver Diseases (EASL) have recommended starting with a Fibrosis-4 index score (FIB-4) followed by LSM on VCTE to identify at-risk MASH.[12–14] These recommendations for sequential testing are based on how (1) FIB-4 has at least 30% of the patients failing to be classified and therefore falling into an indeterminate zone for diagnosis and (2) sequential testing has been shown to narrow this indeterminate zone.[15] However, these testing recommendations have inherent limitations. FIB-4 was developed specifically to identify patients with advanced fibrosis ($\geq$F3) and so does not target MASH.[16] Furthermore, the core assumption when applying FIB-4 to a high-risk population is that the pretest probability of MASH is high, but this assumption does not always hold true. In addition, a proportion of individuals in the FIB-4 indeterminate zone are subsequently required to be classified by LSM by VCTE, and another proportion of those patients who fall into the indeterminate zone (8–12 kPa) on LSM by VCTE require further testing. VCTE also remains machine-dependent and limited in accessibility, particularly outside of gastroenterology and hepatology practices.[16] All of these reasons provide a rationale for the continued development of biomarker panels that can be used as a point of care to identify those with at-risk MASH. Metabolomics are serum/plasma-based testing that measure lipids, carbohydrates, amino acids, and nucleic acids, among other metabolites. Reflecting the underlying pathophysiologic processes in MASLD and MASH, metabolites are attractive noninvasive tests for the identification of those with at-risk MASH. Previous studies from our group have shown that a metabolomics-driven score accurately differentiates normal liver from NAFLD (AUC $0.88 \pm 0.05$, sensitivity 0.94, and specificity 0.57) and NAFL from NASH (AUC $0.79 \pm 0.04$, sensitivity 0.70, and specificity 0.81).[17–19] We hypothesize that the *M*etabolomics-*A*dvanced *StE*atohepatitis *F*ibrosis Score (MASEF score) is superior or comparable to VCTE-driven scores in identifying patients with at-risk MASH. We aimed to develop and validate a highly specific metabolomics-driven score to accurately identify patients with at-risk MASH. We compared this score to the other noninvasive tests, including FAST and LSM by VCTE. We further compared FIB-4 + LSM by VCTE as recommended by the AGA, AASLD, and EASL with FIB-4 + MASEF to determine if MASEF can decrease the number of patients in the indeterminate zone who may subsequently need a referral to hepatology and increase the number of patients correctly classified.

# METHODS

## Study design

This is an international cross-sectional multicenter study, with data collected between 2015 and 2021, consisting of derivation (n = 790) and validation (n = 565) patient cohorts. The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines were followed to report the development and validation of the prediction model for the diagnosis of at-risk MASH patients (see Supplementary Table 1, http://links.lww.com/HEP/H911).[20,21]

## Cohorts

The derivation cohort included data from the ARREST Cohort[22] (with patients from Israel, United States, and Mexico) and 7 separate tertiary care centers, including the University of Florida, United States; Donostia University Hospital, Spain; Pontificia Universidad Católica de Chile, Chile; Fondazione IRCCS Ca' Granda—Ospedale Maggiore Policlinico, Italy; Marqués de Valdecilla University Hospital, Spain; General University Hospital and the First Faculty of Medicine, Charles University, Czech Republic; and Virgen de Valme University Hospital, Spain. The validation cohort included data from Spain's NASH Registry (from 6 hospitals: Marqués de Valdecilla University Hospital; Clinic University Hospital, University of Valladolid; Virgen del Rocío University Hospital; Gregorio Marañón University Hospital; Ramón y Cajal University Hospital; and Puerta del Hierro University Hospital) and 6 separate tertiary care centers including Marqués de Valdecilla University Hospital, Spain; Cruces University Hospital, Spain; Cedars-Sinai Medical Center, United States; Virginia Commonwealth University, United States; Príncipe de Asturias University Hospital, Alcalá University, Spain; and Mount Sinai Health System, United States.

The derivation cohort data were first collected between 2015 and 2018, and the validation cohort was subsequently collected between 2019 and 2021. Both the derivation and validation cohorts included those aged 18 years or older who underwent laboratory testing and liver biopsy for suspicion of MASLD within a 6-month period. Exclusion criteria for all cohorts included history of liver diseases such as $\alpha$1-antitrypsin deficiency, autoimmune hepatitis, chronic hepatitis B or C, hemochromatosis, primary biliary cholangitis, primary sclerosing cholangitis, Wilson's disease, medications or supplements that may cause hepatic

steatosis, history of HIV, and excessive alcohol use, defined as ≥14 or ≥7 standard drinks of alcohol/week for men or women, respectively. For algorithm development, only MASLD patients without cirrhosis were included, whereas MASLD patients with fibrosis stage 4 (F4) were also evaluated in the validation cohort.

Patients underwent in-depth medical history taking, physical examination, and laboratory studies before being selected for liver biopsy based on LSM >7 kPa on VCTE or other methods that suggested that these patients may have significant or advanced fibrosis. This approach has become a standard in clinical practice and clinical trials, as performing liver biopsy without prior stratification could be unethical.

All procedures were conducted in accordance with both the Declarations of Helsinki and Istanbul and with the appropriate ethics and/or institutional review committee of each contributing institution. All patients gave written informed consent to participate in the study.

## Liver histology

Expert liver pathologists blinded to both clinical and imaging data completed histological assessments for their own center or study cohorts independently. Patients were diagnosed with MASH if they met Brunt criteria[23] and were assessed for NAFLD activity score (NAS) using Kleiner criteria.[24] For the NAS score ranging from 0 to 8, liver histology was scored and summed for steatosis from 0 to 3, hepatocellular ballooning from 0 to 2, and lobular inflammation from 0 to 3. Staged from 0 to 4, fibrosis was defined as fibrosis stage 0: no fibrosis; fibrosis stage 1: either mild-moderate perisinusoidal or periportal fibrosis; fibrosis stage 2 (F2): both perisinusoidal and portal/periportal fibrosis; fibrosis stage 3 (F3): bridging or advanced fibrosis; and fibrosis stage 4 (F4): cirrhosis.

## MASEF score

The test was developed initially to identify MASH with at least 1 point in each category of steatosis, lobular inflammation, and ballooning, as well as ≥F2. The MASEF score provides the logistic probability score that allows for obtaining a predicted probability score (ranging from 0 to 1) of at-risk MASH using a multivariable logistic regression algorithm. The final MASEF score included 12 lipids, body mass index (BMI), AST, and alanine aminotransferase (ALT). Importantly, we used machine learning models to develop the final score.

## Outcomes

The primary outcome of this study was the diagnosis of at-risk MASH, defined as MASH with NAS ≥ 4 and significant fibrosis (≥F2), with NAS scoring at least 1 point in each category of steatosis, lobular inflammation, and ballooning. This group of patients is targeted for pharmacological therapy in MASH, as this group carries an increased risk of morbidity and mortality. Our secondary outcomes were (1) to compare the MASEF score to the FAST score and (2) to compare the current recommended care pathway starting with FIB-4 followed by LSM by VCTE to an alternative care pathway starting with FIB-4 followed by the MASEF score.

## Statistical analysis

The MASEF score was developed on the 790 patients in the derivation cohort. The selection of parameters was based on the combination of lipidomic features and clinical variables related to MASH and fibrosis. Parameters were combined into a multivariable logistic regression algorithm. From a feature set including ~240 lipids, BMI, AST, ALT, platelets, and glycosylated hemoglobin (HbA1c), a first selection process was applied using a variance threshold and a feature importance analysis.[25] Among the different approaches for model selection, the MASEF score algorithm was selected using Training and Test Cohorts (derivation cohort recursively divided into training and test sets).[26] Varying model complexity (adding or deleting parameters), we measured the model error to select the optimal complexity. To evaluate the real performance of the model and avoid overfitting, the estimator was internally validated using a 10-fold cross-validation procedure splitting all the samples into 10 groups, where the estimator learned using 9 groups and was evaluated in the last remaining group.[26] Statistical analyses were performed using Python with scikit-learn v1.0.2 and pandas v1.1.4.[27,28]

Performance was achieved in terms of area under the receiver-operating characteristic (AUC) curve using R software v4.0.3 (packages: 'ROCR' v1.0-11; 'pROC' v1.17.0.1). The AUC is reported along with a 95% CI calculated with 2000 stratified bootstrap replicates. The 1-score cutoff was obtained as the optimal Youden's J statistic (or J point), where sensitivity + specificity −1 is maximum, as shown in Figure 1. Multiple score cutoffs corresponding to high sensitivity (cutoff A) and high specificity (cutoff B) were calculated in the derivation cohort and then applied to the validation cohort. The best 2 cutoffs corresponding to 80% sensitivity and 90% specificity were chosen and are comparable to the serum-based NIS4 test.[8] Values between these 2 cutoffs (A–B) are usually known as indeterminate or "grey zone" values. In addition, as mentioned, we have also reported various cutoffs' performances in Supplementary Table 2, http://links.lww.com/HEP/H912. For each cutoff value, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were estimated using package 'caret' v6.0-88. CIs were

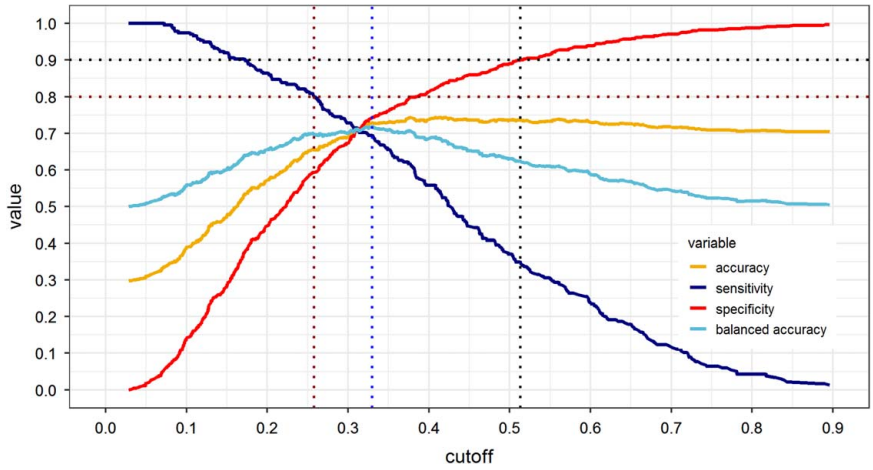


**FIGURE 1** Cumulative distribution analysis for MASEF score in the derivation cohort. This plot displays sensitivity, specificity, accuracy, and balanced accuracy versus all possible MASEF score cutoffs. The values for (sensitivity + specificity − 1) are also represented to illustrate the evaluation of Youden's J statistic (or J point), the cutoff for maximum sensitivity + specificity − 1. Cutoff for 80% sensitivity = 0.258. Cutoff for 90% specificity = 0.513. Cutoff for the optimal J point = 0.330. Abbreviations: MASEF, metabolomics-advanced steatohepatitis fibrosis score.

calculated as exact binomial confidence limits using package 'epiR' v2.0.4190.

To assess differences in clinical variables among groups, statistical significance was determined using the chi-square test for categorical variables and the Kruskal-Wallis *H*-test for continuous variables.

# RESULTS

## Baseline characteristics

The derivation cohort of 790 subjects was used to develop the MASEF score, which was then validated in a separate cohort of 565 subjects (see Supplementary Figure 1, http://links.lww.com/HEP/H913 Supplementary Figure 2, http://links.lww.com/HEP/H913). Table 1 shows the derivation and validation cohorts' demographic, metabolic, serologic, and histologic characteristics. Given that the MASEF score classifies populations into at-risk MASH or not at-risk MASH, Table 1 has been split into these 2 categories. In terms of demographic, metabolic, and serologic characteristics, the derivation/validation cohort respectively had a mean age of 51.4 (± 11.6)/54.0 (± 10.6) years with 51%/51% females and 28%/51% with type 2 diabetes. In the derivation/validation cohorts, respectively, mean AST was 39.8 (± 29.4)/40.1 (± 26.6) IU/L, mean ALT was 56.5 (± 45.2)/52.4 (± 39.4) IU/L, mean triglyceride was 170.3 (± 109.0)/163.7 (± 116.7) mg/dL, and mean HbA1c was 6.5% (± 1.2%) / 6.2% (± 1.9%). In addition, a comparison of demographic, metabolic, serologic, and histologic characteristics in patients with and without at-risk MASH between derivation and validation cohorts is shown in Supplementary Table 3, http://links.lww.com/HEP/H914.

Regarding histologic characteristics, the prevalence of MASLD with ≥F2 and ≥F3 was respectively 31.5% and 12.2% in the derivation cohort, whereas the prevalence of MASLD with ≥F2, ≥F3, and F4 was 32.6%, 20.0%, and 6.4%, respectively in the validation cohort. Patients with F4 were excluded from the derivation cohort to avoid bias in the algorithm development that could artificially increase the performance of the test, but F4 patients were included in the validation cohort. Furthermore, at-risk MASH was reported in 235 (29.7%) of 790 patients in the derivation cohort and 165 (29.2%) of 565 patients in the validation cohort (Table 1).

## Risk factors of at-risk MASH

In both derivation and validation cohorts, at-risk MASH was significantly associated with increased age, glucose, HbA1c, AST, and ALT (Table 1). In addition, at-risk MASH in the derivation cohort was significantly associated with increased homeostatic model assessment (HOMA) and gamma-glutamyl transferase (GGT) (HOMA and GGT were not available in the validation cohort), whereas at-risk MASH in the validation cohort was significantly associated with traditional risk factors such as increased triglyceride, and total cholesterol (Table 1).

## The MASEF score and its performance

Models combining metabolic features and clinical variables, including BMI, ALT, AST, platelets, and HbA1c, were compared (see statistical analysis). The final algorithm, comprising 12 lipids, BMI, ALT, and AST, was found to be the best model with the highest AUC in the derivation cohort for identifying at-risk MASH patients, resulting in the MASEF score. The panel of 12 lipids

**TABLE 1** Characteristics of the Derivation and Validation Cohorts: (a) Characteristics of the Derivation Cohort. (b) Characteristics of the Validation Cohort

| | At-risk MASH | Not at-risk MASH | *p* | Overall |
|---|---|---|---|---|
| (A) Characteristics of the derivation cohort | | | | |
| Demographics | | | | |
| n, (%) | 235 (29.7) | 555 (70.3) | — | 790 |
| Age (y) | 52.7 ± 11.0 (235) | 50.8 ± 11.8 (555) | 0.0389 | 51.4 ± 11.6 (790) |
| BMI (kg/m$^2$) | 35.6 ± 7.4 (235) | 34.8 ± 7.7 (555) | 0.0518 | 35.1 ± 7.6 (790) |
| Sex, n (%) | Female 132 (56) | Female: 271 (49) | 0.0704 | Female: 403 (51) |
| | Male: 103 (44) | Male: 284 (51) | — | Male 387 (49) |
| Metabolic | | | | |
| Diabetes (type 2), n (%) | 70 (30) | 149 (27) | 0.106 | 219 (28) |
| Abdominal perimeter (cm) | 148.8 ± 12.4 (4) | 133.4 ± 18.3 (30) | 0.092 | 135.2 ± 18.3 (34) |
| Blood | | | | |
| Glucose (mg/dL) | 127.4 ± 40.4 (195) | 119.2 ± 35.0 (464) | 8.16e-3 | 121.6 ± 36.8 (659) |
| HbA1c (%) | 6.8 ± 1.2 (189) | 6.4 ± 1.2 (449) | < 0.001 | 6.5 ± 1.2 (638) |
| HOMA | 8.1 ± 6.6 (45) | 4.6 ± 3.8 (141) | < 0.001 | 5.4 ± 4.8 (186) |
| AST (IU/L) | 52.8 ± 37.4 (235) | 34.3 ± 23.2 (555) | < 0.001 | 39.8 ± 29.4 (790) |
| ALT (IU/L) | 70.2 ± 48.4 (235) | 50.6 ± 42.4 (555) | < 0.001 | 56.5 ± 45.2 (790) |
| GGT (IU/L) | 72.0 ± 99.2 (181) | 51.9 ± 59.9 (377) | < 0.001 | 58.5 ± 75.4 (558) |
| Triglyceride (mg/dL) | 179.3 ± 137.4 (197) | 166.7 ± 95.1 (492) | 0.644 | 170.3 ± 109.0 (689) |
| Total cholesterol (mg/dL) | 191.9 ± 46.8 (231) | 190.2 ± 44.8 (547) | 0.651 | 190.7 ± 45.4 (778) |
| LDL cholesterol (mg/dL) | 114.7 ± 39.0 (191) | 116.4 ± 39.8 (487) | 0.968 | 115.9 ± 39.5 (678) |
| HDL cholesterol (mg/dL) | 45.0 ± 13.5 (197) | 45.3 ± 12.6 (490) | 0.637 | 45.2 ± 12.8 (687) |
| Histology | | | | |
| Steatosis in NAS, n (%) | 0: 0 (0) | 0: 0 (0) | < 0.001 | 0: 0 (0) |
| | 1: 81 (34) | 1: 299 (54) | — | 1: 380 (48) |
| | 2: 94 (40) | 2: 155 (28) | — | 2: 249 (32) |
| | 3: 60 (26) | 3: 101 (18) | — | 3: 161 (20) |
| Ballooning in NAS, n (%) | 0: 0 (0) | 0: 265 (48) | < 0.001 | 0: 265 (34) |
| | 1: 127 (54) | 1: 231 (42) | — | 1: 358 (45) |
| | 2: 108 (46) | 2: 59 (11) | — | 2: 167 (21) |
| Inflammation in NAS, n (%) | 0: 0 (0) | 0: 65 (12) | < 0.001 | 0: 65 (8) |
| | 1: 36 (15) | 1: 294 (53) | — | 1: 330 (42) |
| | 2: 156 (66) | 2: 171 (31) | — | 2: 327 (41) |
| | 3: 43 (18) | 3: 25 (5) | — | 3: 68 (9) |
| Fibrosis, n (%) | F0: 0 (0) | F0: 216 (39) | < 0.001 | F0: 216 (27) |
| | F1: 0 (0) | F1: 325 (59) | — | F1: 325 (41) |
| | F2: 142 (60) | F2: 11 (2) | — | F2: 153 (19) |
| | F3: 93 (40) | F3: 3 (1) | — | F3: 96 (12) |
| | F4: 0 (0) | F4: 0 (0) | — | F4: 0 (0) |
| (B) Characteristics of the validation cohort | | | | |
| Demographics | | | | |
| n (%) | 165 (29.2) | 400 (70.8) | — | 565 |
| Age (y) | 57.7 ± 9.1 (165) | 52.4 ± 10.8 (400) | < 0.001 | 54.0 ± 10.6 (565) |
| BMI (kg/m$^2$) | 34.5 ± 6.6 (165) | 38.4 ± 7.7 (400) | < 0.001 | 37.2 ± 7.6 (565) |
| Sex, n (%) | Female: 83 (50) | Female: 205 (51) | 0.714 | Female: 288 (51) |
| | Male: 72 (44%) | Male: 164 (41%) | — | Male: 236 (42%) |
| Metabolic | | | | |
| Diabetes (type 2), n (%) | 115 (70) | 174 (44) | < 0.001 | 289 (51) |

**TABLE 1.** (*continued*)

| | At-risk MASH | Not at-risk MASH | *p* | Overall |
|---|---|---|---|---|
| Abdominal perimeter (cm) | N/A ± N/A (N/A) | N/A ± N/A (N/A) | N/A | N/A ± N/A (N/A) |
| Blood | | | | |
| Glucose (mg/dL) | 131.7 ± 46.3 (116) | 105.9 ± 37.7 (329) | < 0.001 | 112.6 ± 41.6 (445) |
| HbA1c (%) | 6.6 ± 1.0 (106) | 6.1 ± 2.1 (289) | < 0.001 | 6.2 ± 1.9 (395) |
| HOMA | NaN ± NA (0) | NaN ± NA (0) | NA | NaN ± NA (0) |
| AST (IU/L) | 56.7 ± 33.8 (165) | 33.3 ± 19.2 (400) | < 0.001 | 40.1 ± 26.6 (565) |
| ALT (IU/L) | 71.3 ± 46.6 (165) | 44.6 ± 33.2 (400) | < 0.001 | 52.4 ± 39.4 (565) |
| GGT (IU/L) | NaN ± NA (0) | NaN ± NA (0) | NA | NaN ± NA (0) |
| Triglyceride (mg/dL) | 181.1 ± 145.8 (133) | 157.3 ± 103.6 (365) | 0.0219 | 163.7 ± 116.7 (498) |
| Total cholesterol (mg/dL) | 187.3 ± 38.3 (131) | 176.2 ± 38.3 (374) | 4.25e-3 | 179.0 ± 38.6 (505) |
| LDL cholesterol (mg/dL) | 107.1 ± 36.4 (119) | 104.5 ± 34.3 (330) | 0.469 | 105.2 ± 34.9 (449) |
| HDL cholesterol (mg/dL) | 47.6 ± 10.8 (125) | 43.4 ± 13.9 (338) | < 0.001 | 44.5 ± 13.3 (463) |
| Histology | | | | |
| Steatosis in NAS, n (%) | 0: 0 (0) | 0: 50 (12) | < 0.001 | 0: 50 (9) |
| | 1: 44 (27) | 1: 187 (47) | — | 1: 231 (41) |
| | 2: 85 (52) | 2: 108 (27) | — | 2: 193 (34) |
| | 3: 36 (22) | 3: 55 (14) | — | 3: 91 (16) |
| Ballooning in NAS | 0: 0 (0) | 0: 130 (32) | < 0.001 | 0: 130 (23) |
| | 1: 64 (39) | 1: 192 (48) | — | 1: 256 (45) |
| | 2: 99 (60) | 2: 77 (19) | — | 2: 176 (31) |
| Inflammation in NAS, n (%) | 0: 0 (0) | 0: 111 (28) | < 0.001 | 0: 111 (20) |
| | 1: 96 (58) | 1: 196 (49) | — | 1: 292 (52) |
| | 2: 57 (35) | 2: 74 (18) | — | 2: 131 (23) |
| | 3: 12 (7) | 3: 19 (5) | — | 3: 31 (5) |
| Fibrosis, n (%) | F0: 0 (0) | F0: 197 (49) | < 0.001 | F0: 197 (35) |
| | F1: 0 (0) | F1: 184 (46) | — | F1: 184 (33) |
| | F2: 68 (41) | F2: 3 (1) | — | F2: 71 (13) |
| | F3: 70 (42) | F3: 7 (2) | — | F3: 77 (14) |
| | F4: 27 (16) | F4: 9 (2) | — | F4: 36 (6) |

Note: ≥F2, fibrosis stage 2 or higher; At-risk MASH, MASH + NAS ≥ 4 + significant fibrosis (≥F2).
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; F0, fibrosis stage 0; F1, fibrosis stage 1; F2, fibrosis stage 2; F3, fibrosis stage 3; F4, fibrosis stage 4 or cirrhosis; GGT, gamma-glutamyl transferase; HbA1c, glycosylated hemoglobin; HOMA, homeostatic model assessment; MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score.

included 2 triglycerides, 5 glycerophosphocholines, 1 cholesteryl ester, 1 ceramide, and 3 sphingomyelins.

Figure 1 demonstrates the MASEF score's performance versus the cutoff in the derivation cohort. To evaluate the contribution of the lipidomic versus clinical variables to the final model, the MASEF score performance was compared with 2 different models, including (1) only clinical variables (BMI, AST, and ALT) or (2) only lipidomic features and BMI (without ALT and AST). The MASEF score's overall performance was better than those of the 2 aforementioned models (see Supplementary Table 4, http://links.lww.com/HEP/H915).

Table 2 shows the performances of MASEF and FAST scores in identifying at-risk MASH in terms of AUC, accuracy, sensitivity, specificity, PPV, and NPV. The performances were assessed using a 1-score cutoff that is the optimal Youden's J statistic (or J point). The 1-score cutoff for MASEF was obtained in the derivation cohort, as shown in Figure 1, and then was applied to the validation cohort. FAST J point was calculated in the validation cohort, as it was not available in the derivation cohort. MASEF score sensitivity and NPV were statistically higher ($p < 0.001$) than FAST in the validation cohort, however, the specificity and PPV were statistically lower than FAST. ROC plots of sensitivity versus 1-specificity for all possible 1-score cutoffs for the MASEF and FAST scores are shown in Figure 2.

Table 3 shows the performances of the MASEF and FAST scores with 2-score cutoffs corresponding to 90% specificity and 80% sensitivity. MASEF cutoffs were calculated in the derivation cohort and applied in the

**TABLE 2** Performances of the MASEF and FAST Scores in Identifying At-Risk MASH

| Cohort | MASEF | | FAST |
|---|---|---|---|
| | Derivation | Validation | Validation |
| AUC (*p*) | 0.756 | 0.789 (0.89) | 0.736 (0.064) |
| (95% CI) | (0.718–0.792) | (0.750–0.827) | (0.683, 0.788) |
| Cutoff | 0.33 | 0.33 | 0.637 |
| Accuracy (*p*), % | 72.9 | 69.0 (0.134) | 70.4 (0.73) |
| Sensitivity (*p*), % | 69.4 | 78.2 (0.66) | 58.5 ($p < 0.001$) |
| Specificity (*p*), % | 74.4 | 65.2 (0.003) | 79.0 (0.001) |
| PPV (*p*), % | 53.4 | 48.1 (0.24) | 66.7 (0.001) |
| NPV (*p*), % | 85.2 | 87.9 (0.33) | 72.6 ($p < 0.001$) |
| Total patients (n) | 790 | 565 | 311 |

Notes: The performances of the MASEF and FAST scores in identifying at-risk MASH are assessed using a 1-score cutoff that is the optimal Youden's J statistic (or J point). The 1-score cutoff for MASEF was obtained in the derivation cohort and applied to the validation cohort. The 1-score cutoff for FAST was calculated in the validation cohort, as FAST was not available in derivation cohort. There is no grey zone as there is only 1 cutoff.
≥F2, fibrosis stage 2 or higher; At-risk MASH, MASH + NAS ≥4 + significant fibrosis (≥F2).
Abbreviations: AUC, area under the receiver-operating characteristic curve; FAST, FibroScan-aspartate aminotransferase; MASEF, metabolomics-advanced steatohepatitis fibrosis score; MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value.

validation cohort. FAST's previously published cutoffs of 0.35 and 0.67 were applied in the validation cohort[9] as FAST was not available in the derivation cohort. Other various specificity and sensitivity threshold performances of the MASEF score are shown in Supplementary Table 2, http://links.lww.com/HEP/H912.

Compared with FAST, MASEF exhibited numerically higher AUC ($p < 0.064$) in the validation cohort (Table 3). In addition, MASEF demonstrated similar PPV and sensitivity, accuracy, and overall higher specificity ($p < 0.042$) and NPV ($p < 0.004$) at the 80% sensitivity cutoff. MASEF also exhibited similar sensitivity, specificity, accuracy, and PPV and overall higher NPV ($p < 0.023$) at the 90% specificity cutoff compared to FAST's validation cohort (Table 3, Figure 3).

## FIB-4 + MASEF in comparison with FIB-4 + LSM by VCTE

We further compared the performance of a FIB-4 + MASEF algorithm to that of FIB-4 + LSM by VCTE, the current recommendation of many societies' guidance, in the same subpopulation of 310/565 patients with available LSM by VCTE data in the validation cohort.[4,29] In all, 133 (43%) of 310 patients had FIB-4 < 1.30 and were classified as low risk of having at-risk MASH, 37 (12%) of 310 patients had FIB-4 > 2.67 and were classified as high risk of having at-risk MASH, and 140 (45%) of 310 were classified into the indeterminate or grey zone and then were further analyzed by MASEF score or LSM by VCTE (Table 4).

When using MASEF as the second test after FIB-4, 19 (14%) of 140 patients had MASEF < 0.258 and were classified as not at-risk MASH, 57 (41%) of 140 had MASEF > 0.513 and were classified as at-risk MASH, and 64 (45%) fell into the indeterminate zone. Among patients with MASEF < 0.258, 15 (79%) were correctly classified (NAS ≤ 3 with any fibrosis or fibrosis ≤ F1 with any NAS), and 4 (21%) were misclassified (NAS ≥ 4 with ≥F2). Among patients with MASEF > 0.513, 37 (65%) were correctly classified, and 20 (35%) were misclassified. Among the 64 (46%) of 140 patients classified into the grey zone, 28 (44%) patients were at-risk MASH.

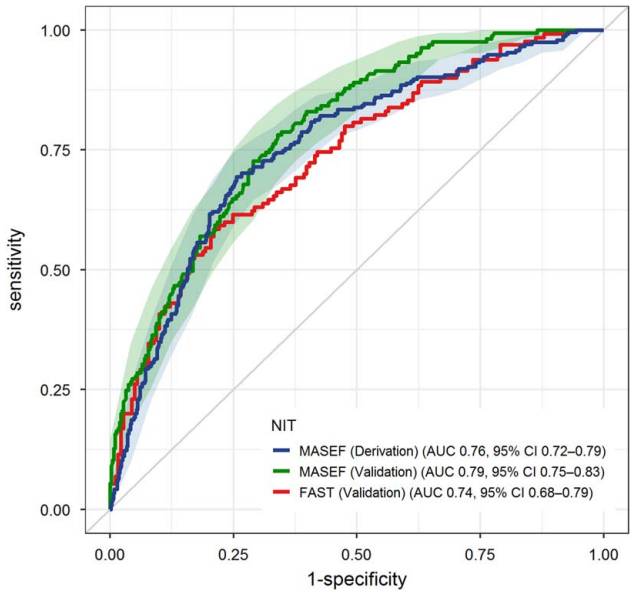


**FIGURE 2** Diagnostic Performances of the MASEF and FAST Scores in Terms of AUC for the Diagnosis of At-Risk MASH in the Derivation and Validation Cohorts ≥F2 = fibrosis stage 2 or higher. At-risk MASH = MASH + NAS ≥4 + significant fibrosis (≥F2). Abbreviations: AUC, area under the receiver-operating characteristic curve; FAST, FibroScan-aspartate aminotransferase; MASEF, metabolomics-advanced steatohepatitis fibrosis score NAS, NAFLD activity score; MASH, metabolic dysfunction-associated steatohepatitis; NIT, noninvasive test.

**TABLE 3** Performances of the MASEF and FAST Scores with 2-score Cutoffs in Identifying At-risk MASH

| | MASEF | | FAST |
|---|---|---|---|
| **Cohort** | **Derivation, %** | **Validation, %** | **Validation, %** |
| AUC (*p*) | 0.756 | 0.789 (0.89) | 0.736 (0.064) |
| (95% CI) | (0.718–0.792) | (0.750–0.827) | (0.683, 0.788) |
| Cutoff A | 0.258 | 0.258 | 0.35 |
| n | 377 | 215 | 91 |
| Sensitivity (*p*) | 80.0 | 89.1 (0.022) | 85.4 (0.44) |
| Specificity (*p*) | 59.5 | 49.2 (0.002) | 39.8 (0.042) |
| PPV (*p*) | 45.5 | 42.0 (0.37) | 50.5 (0.059) |
| NPV (*p*) | 87.5 | 91.6 (0.163) | 79.1 (0.004) |
| Cutoff B | 0.513 | 0.513 | 0.67 |
| n | 137 | 121 | 98 |
| Sensitivity (*p*) | 34.9 | 44.2 (0.074) | 52.3 (0.21) |
| Specificity (*p*) | 90.1 | 88.0 (0.36) | 83.4 (0.172) |
| PPV (*p*) | 59.9 | 60.3 (1.0) | 69.4 (0.21) |
| NPV (*p*) | 76.6 | 79.3 (0.33) | 70.9 (0.023) |
| Grey zone A-B, n (%) | 276, (35.0) | 229, (40.5) | 122, (39.2) |
| Total patients (n) | 790 | 565 | 311 |

Notes: For MASEF, the performances were assessed with the cutoffs determined in the derivation cohort and applied in the validation cohort. The lower cutoff is a rule-out threshold corresponding to 80% sensitivity in the derivation cohort, whereas the higher cutoff comprises a rule-in threshold corresponding to 90% specificity in the derivation cohort. For FAST, only available in the validation cohort, the previously published cutoffs were used. Participants with a score in-between these cutoffs are in the grey zone.

≥F2, fibrosis stage 2 or higher; At-risk MASH, MASH + NAS ≥4 + significant fibrosis (≥F2).

Abbreviations: AUC, area under the receiver-operating characteristic curve; FAST, FibroScan-aspartate aminotransferase; MASEF, metabolomics-advanced steatohepatitis fibrosis score MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value.

When using LSM by VCTE as the second test after FIB-4, 36 (25%) of 140 patients had LSM < 8 kPa and were classified as not at-risk MASH, 53 (38%) of 140 had LSM > 12 kPa and were classified as at-risk MASH, and 51 (36%) fell into the indeterminate zone. Among patients with LSM < 8 kPa, 24 (67%) were correctly classified (NAS ≤ 3 with any fibrosis or fibrosis ≤ F1 with any NAS), and 12 (33%) were misclassified (NAS ≥ 4 with ≥F2). Among patients with LSM > 12 kPa, 32 (60%) were correctly classified, and 21 (39%) were misclassified. Among the 51 (36%) of 140 patients classified into the grey zone, 25 (49%) patients were at-risk MASH.

The performances of FIB-4 + LSM by VCTE and FIB-4 + MASEF using a 1-score cutoff are shown in Table 5. Individual scores (FIB-4, LSM, and MASEF) were also compared with the combination algorithms to evaluate the performance of each method. The overall performance of FIB-4 + MASEF, though higher, was not statistically different from that of FIB-4 + LSM by VCTE ($p = 0.69$). The MASEF score alone obtained higher sensitivity ($p < 0.001$) than the FIB-4 + LSM by

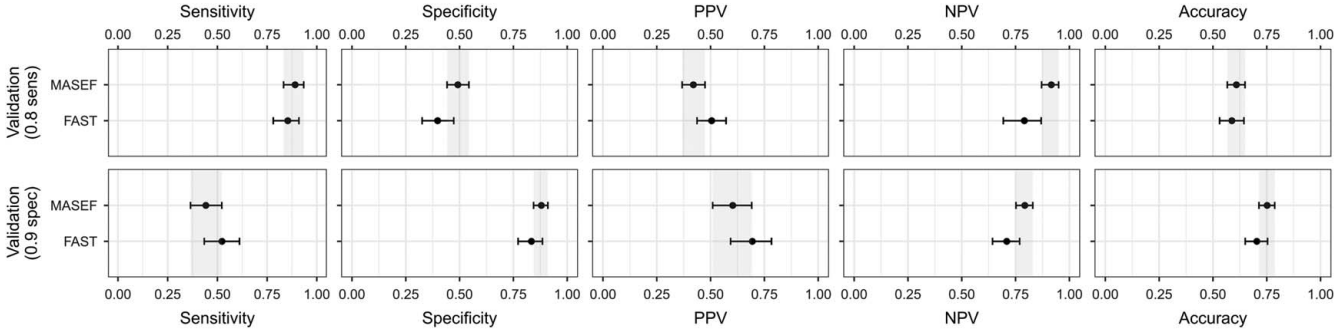


**FIGURE 3** Diagnostic performances of the MASEF and FAST scores with 2-score cutoffs for the diagnosis of at-risk MASH. This plot displays sensitivity, specificity, NPV, PPV, and accuracy in the validation cohort for the cutoffs corresponding to 80% sensitivity (rule-out threshold) and 90% specificity (rule-in threshold). In each plot, the CI of the MASEF score for each characteristic is represented with a grey band. ≥F2 = fibrosis stage 2 or higher. At-risk MASH = MASH + NAS ≥4 + significant fibrosis (≥F2). Abbreviations: FAST, FibroScan-aspartate aminotransferase; MASEF, metabolomics-advanced steatohepatitis fibrosis score; MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value.

**TABLE 4** Patient classification using FIB-4 + MASEF in comparison with FIB-4 + LSM by VCTE. 310 patients were first classified using FIB-4

| ALGORITHM | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FIB-4 | | | | | | | | | | | |
| FIB-4 <1.3, Low Risk (N= 133) | | 1.3 ≥ FIB-4 ≤ 2.67, Indeterminate Risk (N= 140) | | | | | | FIB-4 > 2.67, High Risk (N=37) | | | |
| Not At-risk MASH | At-risk MASH | Not At-risk MASH | | At-risk MASH | | | | Not At-risk MASH | At-risk MASH | | |
| | | 71 | | 69 | | | | | | | |
| | | MASEF | | | | | | | | | |
| | | MASEF < 0.258, Low Risk (N=19) | | 0.258 ≥ MASEF ≤ 0.513, Indeterminate Risk (N=64) | | MASEF > 0.513, High Risk (N=57) | | | | | |
| | | Not At-risk MASH | At-risk MASH | Not At-risk MASH | At-risk MASH | Not At-risk MASH | At-risk MASH | | | | |
| 97 | 36 | 15 | 4 | 36 | 28 | 20 | 37 | 11 | 26 | | |
| | | LSM (VCTE) | | | | | | | | | |
| | | LSM (VCTE) < 8 kPa, Low Risk (N= 36) | | 8 kPa ≥ LSM (VCTE) ≤ 12 kPa, Indeterminate Risk (N= 51) | | LSM (VCTE) > 12 kPa, High Risk (N= 53) | | | | | |
| | | Not At-risk MASH | At-risk MASH | Not At-risk MASH | At-risk MASH | Not At-risk MASH | At-risk MASH | | | | |
| | | 24 | 12 | 26 | 25 | 21 | 32 | | | | |

Notes: Those patients that FIB-4 classified as intermediate risk (140) were subsequently analyzed with a second test, MASEF score or LSM by VCTE.
≥F2, fibrosis stage 2 or higher; At-risk MASH, MASH + NAS ≥4 + significant fibrosis (≥F2).
Abbreviations: FIB-4, fibrosis-4 index; LSM, liver stiffness measurement; MASEF, metabolomics-advanced steatohepatitis fibrosis score; MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value; VCTE, vibration-controlled transient elastography.

the VCTE algorithm. All the individual scores provided an indeterminate zone that was statistically higher ($p < 0.001$) than the combination of FIB-4 + LSM by VCTE or FIB-4 + MASEF (Table 5, Figure 4).

## DISCUSSION

In this study, we developed and validated the novel, blood-based Metabolomics-Advanced StEatohepatitis Fibrosis Score (MASEF score) that noninvasively identifies patients with at-risk MASH who may benefit from pharmacotherapies and intervention.

Given that no significant differences were found between FIB-4 + MASEF and FIB-4 + LSM by VCTE, MASEF is a promising candidate to be used alternatively to LSM by VCTE in the FIB-4 + LSM by VCTE algorithm that is currently recommended by the AGA, AASLD, and EASL. The current FIB-4 + LSM by VCTE algorithm invites ambiguity regarding the next steps for patients who fall into the indeterminate zone on LSM by VCTE (8-12 kPa) and who may warrant subsequent referral to a hepatologist. In contrast, the FIB-4 + MASEF algorithm's numerically higher accuracy, sensitivity, NPV, and PPV improve the identification of previously indeterminate patients on FIB-4 + LSM by VCTE and may thus decrease the number of subsequent hepatology referrals. Furthermore, whereas VCTE's use remains limited by personnel training, machine expense, and operating costs and is therefore available mainly to gastroenterologists, MASEF is a blood-based test that, once available, can be routinely requested in the primary care clinic setting. Replacement of LSM by VCTE with MASEF in the FIB-4 + LSM by VCTE algorithm currently recommended by the AGA, AASLD, and EASL, therefore, expands the identification of patients to include at-risk MASH, which is associated with a higher degree of disease progression to fibrosis and cirrhosis and remains a target for novel pharmacotherapies and clinical trials.

**TABLE 5** Performance of FIB-4 + LSM by VCTE in Comparison with FIB-4 + MASEF

| Algorithm | Accuracy (*p*) % | Sensitivity (*p*) % | Specificity (*p*) % | PPV (*p*) % | NPV (*p*) % | Indeterminate zone (*p*) % |
|---|---|---|---|---|---|---|
| FIB-4 + LSM by VCTE | 69.1 | 54.7 | 79.1 | 64.4 | 71.6 | 16 |
| FIB-4 + MASEF | 71.1 (0.69) | 61.2 (0.42) | 78.3 (0.99) | 67.0 (0.83) | 73.7 (0.77) | 21 (0.22) |
| FIB-4 | 72.4 (0.54) | 41.9 (0.150) | 89.8 (0.033) | 70.3 (0.67) | 72.9 (0.9) | 45 ($p < 0.001$) |
| LSM | 65.6 (0.5) | 79.3 ($p < 0.001$) | 55.8 ($p < 0.001$) | 56.5 (0.31) | 78.8 (0.3) | 37 ($p < 0.001$) |
| MASEF | 71.6 (0.66) | 84.7 ($p < 0.001$) | 61.9 (0.005) | 62.2 (0.87) | 84.5 (0.051) | 45 ($p < 0.001$) |
| FAST | 73.7 (0.35) | 77.9 (0.001) | 70.0 (0.135) | 69.1 (0.61) | 78.7 (0.28) | 40 ($p < 0.001$) |

Notes: The performances of FIB-4 + LSM by VCTE and FIB-4 + MASEF, in the subpopulation where LSM by VCTE data were available (310 patients), are compared using a 2-score cutoff that is the optimal Youden's J statistic (or J point). Individual scores, including FAST, were also compared with the combination algorithms. No significant differences were found between the combination algorithms of FIB-4 + LSM by VCTE and FIB-4 + MASEF.
Abbreviations: FAST, FibroScan-aspartate aminotransferase; FIB-4, fibrosis-4 index; LSM, liver stiffness measurement; MASEF, metabolomics-advanced steatohepatitis fibrosis score; NPV, negative predictive value; PPV, positive predictive value; VCTE, vibration-controlled transient elastography.

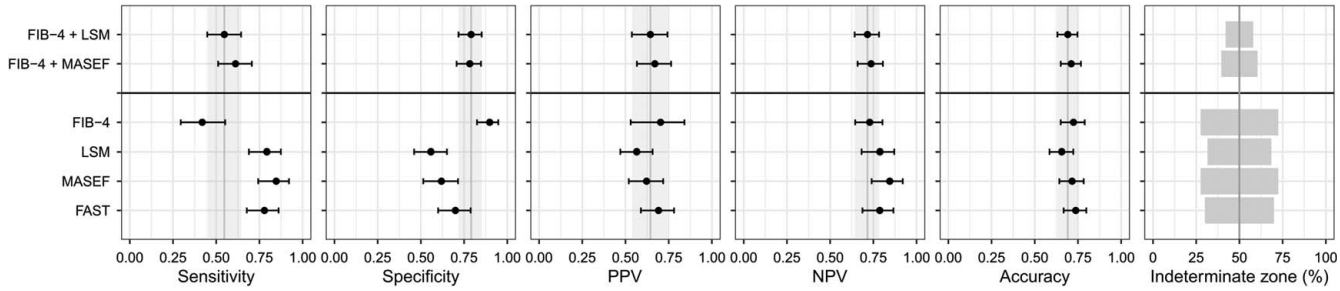


**FIGURE 4** Diagnostic performance of FIB-4 + LSM by VCTE in comparison with FIB-4 + MASEF, FIB-4, LSM, and MASEF scores for the diagnosis of at-risk MASH. This plot displays sensitivity, specificity, NPV, PPV, accuracy, and indeterminate zone in the validation cohort. In each plot, the CI of the FIB-4 + LSM by VCTE for each characteristic is represented with a grey band. At-risk MASH = MASH + NAS ≥4 + significant fibrosis (≥F2). FIB-4 = fibrosis-4 index. Abbreviations: LSM, liver stiffness measurement; MASEF, metabolomics-advanced steatohepatitis fibrosis score; MASH, metabolic dysfunction-associated steatohepatitis; NAS, NAFLD activity score; NPV, negative predictive value; PPV, positive predictive value; VCTE, vibration-controlled transient elastography.

This study has limitations. First, as this was a multicenter international study, liver biopsies were not read centrally. However, histological readings were performed by experienced hepatopathologists. Second, the FAST score and LSM by VCTE data were available in only a subset of patients, yet this data set was large enough to make meaningful comparisons. Third, FIB-4 was developed specifically to identify patients with advanced fibrosis (≥F3), and there is still a paucity of data for using the test in primary care settings, where the prevalence of advanced fibrosis is much lower than in secondary care settings. Furthermore, the MASEF score was developed using cohorts with a higher prevalence of significant and advanced fibrosis (≥F2). For those reasons, further studies should be conducted to assess the performance of FIB-4 + MASEF in primary care settings. Finally, some patients underwent some sort of staging before liver biopsy. However, this approach has become a standard in both clinical practice and clinical trials, as conducting liver biopsy without prior stratification could be unethical.

In addition, this study has major strengths. First, study data were collected from numerous international centers spanning multiple years and are thus very clinically applicable in the real-world setting. Second, the MASEF score performed similar to FAST. Third, MASEF is a blood-based test with clinical variables that are widely available and often routinely obtained in primary care or ambulatory care clinics at low cost, unlike other noninvasive staging tests that may require VCTE or MRI that are much more expensive and limited to gastroenterologists. Of note, the addition of lipids to the clinical parameters and vice versa has improved the score's performance, which is plausible as metabolomics and lipid changes are essential to the pathogenesis of MASH. Fourth, MASEF's cohorts had large sample sizes empowering robust derivation and validation. Finally, to our knowledge, this is the first study that uses a blood-based score as a replacement for LSM by VCTE in the AGA's currently recommended FIB-4 + LSM by VCTE algorithm.

Utilizing the MASEF score as a replacement for LSM by VCTE in the AGA's currently recommended FIB-4 + LSM by VCTE algorithm is feasible and practical, given MASEF's similar performance, high accuracy, and increased accessibility due to the fact that MASEF is a blood-based test. Accurate identification of at-risk MASH will allow patients at high risk of severe liver disease progression to be targeted for novel pharmacotherapies and clinical trials.

## AUTHOR CONTRIBUTIONS

Mazen Noureddin: data interpretation, acquisition of human samples and data, manuscript draft, critical revision of the manuscript for important intellectual content, and final approval of the version to be published. Emily Truong: data interpretation, manuscript draft, critical revision of the manuscript for important intellectual content, and final approval of the version to be published. Rebeca Mayo: data acquisition, data analysis, data interpretation, manuscript draft, critical revision of the manuscript for important intellectual content, and final approval of the version to be published. Ibon Martínez-Arranz: data analysis, data interpretation, and critical revision of the manuscript for important intellectual content. Itziar Mincholé: data analysis, statistical and informatics analysis, data interpretation, manuscript draft, and critical revision of the manuscript for important intellectual content. Jesus M. Banales, Marco Arres, Kenneth Cusi, María Teresa Arias-Loste, Radan Bruha, Manuel Romero-Gómez, Paula Iruzubieta, Rocio Aller, Javier Ampuero, Luis Ibañez-Samaniego, Patricia Aspichueta, and Antonio Martín-Duce: acquisition of human samples and data. José Luis Calleja, Javier Crespo, Tatyana Kushner, and Arun J. Sanyal: acquisition of human samples and data, critical revision of the manuscript for important intellectual content. Pablo Ortiz: study concept and design, study supervision, data interpretation, critical revision of the manuscript for important intellectual content. Stephen A. Harrison, Quentin M. Anstee, and José M.

Mato: critical revision of the manuscript for important intellectual content.

## FUNDING INFORMATION

Marco Arrese received funding from Fondo Nacional de Ciencia y Tecnología de Chile (FONDECYT) under grant agreement No. 1191145. Quentin M. Anstee is supported by the Liver Investigation: Testing Marker Utility in Steatohepatitis (LITMUS) consortium, which received funding from the Innovative Medicines Initiative 2 Joint Undertaking, under grant agreement No. 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. He is a Newcastle NIHR Biomedical Research Centre investigator.

## CONFLICTS OF INTEREST

Mazen Noureddin consults and received grants from Gilead, GlaxoSmithKline, Madrigal, Novo Nordisk, Takeda, and Terns. He consults, advises, and owns stock in Cytodyn. He consults and advises 89BIO, Altimmune, Boehringer Ingelheim, Merck, and Perspecturm. He has received grants and owns stock in Viking. He advises Blade, cohBar, EchoSens, Fractyl, Intercept, NorthSea, Pfizer, Siemens, and Roche Diagnostics. He has received grants from Allergan, Bristol Myers Squibb, Conatus, Corcept, Enanta, Galmed, Galectin, Genfit, Madrigal, Novartis, Pfizer, Shire, and Zydus. He owns stock in Anaetos, ChronWell, Cima, and Rivus Pharma. Rebeca Mayo is employed by OWL Metabolomics. Ibon Martínez-Arranz is employed by OWL Metabolomics. Itziar Mincholé is employed by OWL Metabolomics. Jesus M. Banales consults and advises OWL Metabolomics. He consults and received grants from Albireo. He received grants and is on the speakers' bureau for Incyte. He consults for CymaBay, Ikan Biotech, and QED Therapeutics. He is on the speakers' bureau for Intercept. Kenneth Cusi consults and received grants from Novo Nordisk. He consults for Aligos, Allergan, Altimmune, Arrowhead, AstraZeneca, Bristol Myers Squibb, Boehringer Ingelheim, Coherus, Covance, Eli Lilly, Fractyl, Genentech, Gilead, Hanmi, Intercept, Janssen, Madrigal, Pfizer, Prosciento, Sagimet, and Siemens. He received grants from Cirius, Echosens, Inventiva, LabCorp, National Institute of Health, Nordic Bioscience, Novartis, Poxel, and Zydus. Manuel Romero-Gómez advises and received grants from Gilead, Intercept, and Siemens. He consults for Galmed. He advises AbbVie, Alfa-Sigma, Madrigal, Novo Nordisk, and Pfizer. He received grants from Thera Therapeutics. Tatyana Kushner advises and received grants from Gilead. She consults for AbbVie. She advises Bausch and GlaxoSmithKline. Pablo Ortiz is employed by and owns stock in OWL Metabolomics. He is employed by Biosfer Teslab and Health in Code. Stephen A. Harrison consults, advises, is involved with trials, received grants, and owns stock in Akero, Galectin, GENFIT, Hepion, and NGM Bio. He consults, advises, is involved with trials, and received grants from Axcella, Gilead, Intercept, Madrigal, and Poxel. He consults, advises, received grants, and owns stock in NorthSea Therapeutics. He consults, advises, and is involved with trials for Terns. He consults, advises, and received grants from HighTide, Novartis, Novo Nordisk, and Sagimet. He consults, advises, and owns stock in HistoIndex, Metacrine, and Sonic Incytes. He consults, received grants, and owns stock in Cirius. He consults, is involved with trials, and received grants from ENYO and Viking. He is involved with trials and received grants from Genentech. He consults and is involved with trials for Ionis. He consults and received grants from CiVi, CymaBay, Galmed, and Pfizer. He consults and owns stock in Hepta Bio. He consults and advises for Altimmune, Echosens North America, Foresite Labs, and Medpace. He advises and owns stock in ChronWell. He consults for AgomAb, Alentis, Aligos Therapeutics, Alimentiv, Blade, Bluejay, Boston Pharmaceuticals, Boxer Capital, Can-Fite BioPharma, the Chronic Liver Disease Foundation (CLDF), CohBar, Corcept, Fibronostics, Fortress Biotech, Galecto, Gelesis, Glaxo Smith Kline, GNS Healthcare, GRI Bio, Hepagene, Indalo, Inipharm, Innovate Biopharmaceuticals, Kowa Research Institute, Merck, MGGM, NeuroBo, Nutrasource, Perspectum, Piper Sandler, Prometic (now Liminal BioSciences), Ridgeline Therapeutics, Silverback, and Zafgen (now Larimar). He advises Arrowhead BVF Partners, Humana, and Pathai. He received grants from Bristol Myers Squibb, Conatus, Immuron, and Second Genome. Quentin M. Anstee, on behalf of Newcastle University, consults and advises for Alimentiv, Akero, AstraZeneca, Axcella, 89Bio, Boehringer Ingelheim, Bristol Myers Squibb, Galmed, GENFIT, Genentech, Gilead, GlaxoSmithKline, Hanmi, HistoIndex, Intercept, Inventiva, Ionis, IQVIA, Janssen, Madrigal, Medpace, Merck, NGM Bio, Novartis, Novo Nordisk, PathAI, Pfizer, Prosciento, Poxel, Resolution Therapeutics, Ridgeline Therapeutics, Roche, RTI, Shionogi, and Terns. He is on the speakers' bureau for Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal, Medscape, and Springer Healthcare. He received grants from AstraZeneca, Boehringer Ingelheim, and Intercept. He holds intellectual property rights with Elsevier, Ltd. He serves as coordinator of the IMI2 LITMUS consortium. Arun J. Sanyal consults, advises, and received grants from AstraZeneca, Boehringer Ingelhiem, Bristol Myers Squibb, Eli Lilly, Fractyl, Gilead, Inventiva, Madrigal, Mallinckrodt, Merck, Novartis, Novo Nordisk, and Pfizer. He consults and advises 89 Bio, Albrio, Alnylam, Amgen, Covance, Genetech, GenFit, Hemoshear, HistoIndex, Intercept, Jannsen, NGM Bio, Path AI, Poxel, Prosciento, Regeneron, Roche, Salix, Sequana, Terns, and Tiziana. He advises Akero, Axcella, Blade, Eisai, Hanmi, LG, Pharmanest, Pliant, Siemens,

Takeda, Thera Technologies, Valeant, and Zealand Pharma. He received grants from Echosence-Sandhill, Galmed, Madrigal, Merck, Viking, and Zydus. He owns stock in Durect, Exhalenz, Galmed, GENFIT, Hemoshear, Rivis, Siemens, and Tiziana. He has other interests with Elsevier and UpToDate. The remaining authors (Emily Truong, Marco Arrese, María Teresa Arias-Loste, Radan Bruha, Paula Iruzubieta, Rocio Aller, Javier Ampuero, José Luis Calleja, Luis Ibañez-Samaniego, Patricia Aspichueta, Antonio Martín-Duce, Javier Crespo, and José M. Mato) have no conflicts to report.

## ORCID

*Mazen Noureddin* https://orcid.org/0000–0003–2127–2040
*Emily Truong* https://orcid.org/0000–0002–9167–3299
*Rebeca Mayo* https://orcid.org/0000–0001–9733–2121
*Ibon Martínez-Arranz* https://orcid.org/0000–0001–9483–8426
*Itziar Mincholé* https://orcid.org/0000–0001–5707–4359
*Jesus M. Banales* https://orcid.org/0000–0002–5224–2373
*Paula Iruzubieta* https://orcid.org/0000–0001–9476–1801
*José Luis Calleja* https://orcid.org/0000–0002–2265–6591
*Antonio Martín-Duce* https://orcid.org/0000–0001–8751–9224
*Pablo Ortiz* https://orcid.org/0000–0002–2698–8470
*Quentin M. Anstee* https://orcid.org/0000–0002–9518–0088
*Javier Crespo* https://orcid.org/0000–0001–8248–0172
*José M. Mato* https://orcid.org/0000–0003–1264–3153
*Arun J. Sanyal* https://orcid.org/0000–0001–8682–5748

## REFERENCES

1. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. NAFLD Nomenclature consensus group. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology. 2023;78:1966–86.
2. Noureddin M, Vipani A, Bresee C, Todo T, Kim IK, Alkhouri N, et al. NASH leading cause of liver transplant in women: updated analysis of indications for liver transplant and ethnic and gender variances. Am J Gastroenterol. 2018;113:1649–59.
3. Setiawan VW, Stram DO, Porcel J, Lu SC, Noureddin M, Angeles L, et al. Prevalence of chronic liver disease and cirrhosis by underlying cause in understudied ethnic groups: the Multiethnic Cohort. Hepatology. 2017;64:1969–77.
4. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67:328–57.
5. Sanyal AJ, Van Natta ML, Clark J, Neuschwander-Tetri BA, Diehl A, Dasarathy S, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med. 2021;385:1559–69.
6. Hagström H, Nasr P, Ekstedt M, Hammar U, Stål P, Hultcrantz R, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol. 2017;67:1265–73.
7. Noureddin M, Muthiah MD, Sanyal AJ. Drug discovery and treatment paradigms in nonalcoholic steatohepatitis. Endocrinol Diabetes Metab. 2020;3:e00105.
8. Harrison SA, Ratziu V, Boursier J, Francque S, Bedossa P, Majd Z, et al. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020;5:970–85.
9. Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan WK, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020;5:362–73.
10. Noureddin M, Truong E, Gornbein JA, Saouaf R, Guindi M, Todo T, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. J Hepatol. 2022;76:781–7.
11. Vali Y, Lee J, Boursier J, Spijker R, Löffler J, Verheij J, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis. J Hepatol. 2020;73:252–62.
12. Kanwal F, Shubrook JH, Adams LA, Pfotenhauer K, Wai-Sun Wong V, Wright E, et al. Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. Gastroenterology. 2021;161:1657–69.
13. Kanwal F, Shubrook JH, Younossi Z, Natarajan Y, Bugianesi E, Rinella ME, et al. Preparing for the NASH Epidemic: A Call to Action. Diabetes Care. 2021;44:2162–72.
14. 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. 2023;77:1797–835.
15. Anstee QM, Lawitz EJ, Alkhouri N, Wong VW-S, Romero-Gomez M, Okanoue T, et al. Noninvasive tests accurately identify advanced fibrosis due to NASH: Baseline Data From the STELLAR Trials. Hepatology. 2019;70:1521–30.
16. Younossi ZM, Noureddin M, Bernstein D, Kwo P, Russo M, Shiffman ML, et al. Role of noninvasive tests in clinical gastroenterology practices to identify patients with nonalcoholic steatohepatitis at high risk of adverse outcomes: Expert Panel Recommendations. Am J Gastroenterol. 2021;116:254–62.
17. Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. Nat Rev Gastroenterol Hepatol. 2021;18:835–56.
18. Alonso C, Noureddin M, Lu SC, Mato JM. Biomarkers and subtypes of deranged lipid metabolism in nonalcoholic fatty liver disease. World J Gastroenterol. 2019;25:3009–20.
19. Mayo R, Crespo J, Martínez-Arranz I, Banales JM, Arias M, Mincholé I, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: Results from discovery and validation cohorts. Hepatol Commun. 2018;2:807–20.
20. Moons KGM, Altman DG, Reitsma JB, Ioannidis JA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:1–73.

21. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. Eur Urol. 2015;67:1142–51.
22. Ratziu V, de Guevara L, Safadi R, Poordad F, Fuster F, Flores-Figueroa J, et al. Aramchol in patients with nonalcoholic steatohepatitis: a randomized, double-blind, placebo-controlled phase 2b trial. Nat Med. 2021;27:1825–35.
23. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol. 1999; 94:2467–74.
24. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
25. He Z, Yu W. Stable feature selection for biomarker discovery. Comput Biol Chem. 2010;34:215–25.
26. Trevor H, Robert T, Jerome F. The Elements of Statistical Learning, 2nd ed. Springer, New York, NY. 2009.
27. Pedregosa F, Grisel O, Weiss R, Passos A, Brucher M, Varoquax G, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–30.
28. McKinney W Data Structures for Statistical Computing in Python. Proc. 9th Python Sci. Conf. 2010;1:56–61.
29. Chang D, Truong E, Mena EA, Pacheco F, Wong M, Guindi M, et al. Machine learning models are superior to noninvasive tests in identifying clinically significant stages of NAFLD and NAFLD-related cirrhosis. Hepatology. 2022;77:546–57.

**How to cite this article:** Noureddin M, Truong E, Mayo R, Martínez-Arranz I, Mincholé I, Banales JM, et al. Serum identification of at-risk MASH: The metabolomics-advanced steatohepatitis fibrosis score (MASEF). Hepatology. 2024;79:135–148. https://doi.org/10.1097/HEP.0000000000000542