Hepatology Publish Ahead of Print
DOI:10.1097/HEP.0000000000000351

Noninvasive Assessment of Liver Disease Severity in Patients with Nonalcoholic Fatty Liver Disease (NAFLD) and Type 2 Diabetes

AUTHORS: Grazia Pennisi[1*], Marco Enea[1*], Vincenzo Falco[1], Guruprasad P. Aithal[2], Naaventhan Palaniyappan[2], Yusuf Yilmaz[3], Jerome Boursier[4], Christophe Cassinotto[5], Victor de Lédinghen[6], Wah Kheong Chan[7], Sanjiv Mahadeva[7], Peter Eddowes[2], Philip Newsome[2], Thomas Karlas[8], Johannes Wiegand[8], Vincent Wai-Sun Wong[9], Jörn M. Schattenberg[10], Christian Labenz[10], Won Kim[11], Myoung Seok Lee[11], Monica Lupsor-Platon[12], Jeremy F. L. Cobbold[13], Jian-Gao Fan[14], Feng Shen[14], Katharina Staufer[15,16], Michael Trauner[15], Rudolf Stauber[17], Atsushi Nakajima[18], Masato Yoneda[18], Elisabetta Bugianesi[19], Ramy Younes[19], Silvia Gaia[19], Ming-Hua Zheng[20], Calogero Cammà[1], Quentin M. Anstee[21,22], Ferenc Emil Mózes[23], Michael Pavlides[13, 23], Salvatore Petta[1].

* These authors equally contributed to the study

INSTITUTIONS:

1Sezione di Gastroenterologia, PROMISE, University of Palermo, Italy; [2] Dipartimento di Biomedicina, Neuroscienze e Diagnostica avanzata (BIND), University of Palermo, Palermo, Italy.
2 NIHR Nottingham Biomedical Research Centre (BRC), Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK; Nottingham Digestive Diseases Centre, Translational Medical Sciences, School of Medicine, University of Nottingham, Nottingham, UK.
3 Department of Gastroenterology, School of Medicine, Recep Tayyip Erdogan University, Rize, Turkey .
4 Hepato-Gastroenterology Department, Angers University Hospital, Angers, France; HIFIH Laboratory, UPRES EA3859, Angers University, Angers, France.
5 Department of Diagnostic and Interventional Radiology, Saint-Eloi Hospital, University Hospital of Montpellier, Montpellier, France.
6 Hepatology Unit, University Hospital Bordeaux and INSERM U-1053, Bordeaux University, Pessac, France.
7 Department of Medicine, Faculty of Medicine, University of Malaya, Malaysia.
8 Department of Oncology, Gastroenterology, Hepatology, Pulmonology and Infectious Diseases, University Hospital Leipzig, Leipzig, Germany.
9 Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong.
10 Department of Internal Medicine I, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany; Cirrhosis Center Mainz (CCM), University Medical Center of the Johannes Gutenberg-University, Mainz, Germany.

11 Department of Internal Medicine, Division of Gastroenterology and Hepatology, Seoul National University College of Medicine. Seoul Metropolitan Government Boramae Medical Center, Seoul, Korea.

12 Department of Medical Imaging, Iuliu Hatieganu, University of Medicine and Pharmacy, Regional Institute of Gastroenterology and Hepatology "Prof. Dr. Octavian Fodor", Cluj-Napoca, Romania.
13 Translational Gastroenterology Unit, University of Oxford, Oxford, UK.
14 Department of Gastroenterology, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China.
15 Department of Internal Medicine III, Division of Gastroenterology & Hepatology, Medical University of Vienna, Austria.
16 Department of General Surgery, Division of Transplantation, Medical University of Vienna, Austria.
17 Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Austria.
18 Department of Gastroenterology and Hepatology, Yokohama City University School of Medicine.
19 Division of Gastroenterology and Hepatology, Department of Medical Sciences, University of Turin, Italy.
20 MAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University; Key Laboratory of Diagnosis and Treatment for The Development of Chronic Liver Disease in Zhejiang Province, Wenzhou, China.
21 Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, Tyne and Wear, UK.
22 Newcastle NIHR Biomedical Research Center, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, Tyne and Wear, UK
23 Cardiovascular Medicine, Radcliffe Department of Medical Sciences, University of Oxford, Oxford, UK.

CORRESPONDING AUTHOR: Prof. Salvatore Petta, Section of Gastroenterology and

Hepatology, PROMISE, Policlinico Universitario Paolo Giaccone, Piazza delle Cliniche, 2, 90127

Palermo, Italy. Phone: +39 091 6552274. Fax +39 091 655 2156. E-mail: salvatore.petta@unipa.it.

NUMBER OF FIGURES/SUPPLEMENTAL FIGURES: 3/6

NUMBER OF TABLES/SUPPLEMENTAL TABLES: 5/14

LIST OF ABBREVIATIONS: AUROC: Area Under ROC Curve; LSM: Liver Stiffness

Measurement; NAFLD: Nonalcoholic fatty liver disease; DCA: Decision Curve Analysis.

CONFLICT OF INTEREST:

Guruprasad P. Aithal consults and advises through Nottingham University Consultants Team.

Yusuf Yilmaz consults for Echosens.

Jerome Boursier consults for Echosens.

Victor de Ledinghen is on the speakers' bureau for Echosens.

Wah Kheong Chan consults for or advises Abbvie, Boehringer Ingelheim, Novo Nordisk and Roche; and is on the speakers' bureau for Viatris and Hisky Medical.

Thomas Karlas advises, is on the speakers' bureau for, and received grants from Echosens.

Johannes Wiegand received grants from Echosens.

Vincent Wai-Sun Wong consults and received grants from Gilead. He consults for AbbVie, Boehringer Ingelheim, Echosens, Intercept, Inventiva, Novo Nordisk, Pfizer, Sagimet Biosciences, and TARGET PharmaSolutions.

Vincent Wong consults for or advises Boehringer Ingelheim, Echosens, Intercept, Inventiva, Pfizer, Sagimet Biosciences, and TARGET PharmaSolutions. He consults for or advises, and is on the speakers' bureau for AbbVie and NovoNordisk. He is on the speakers' bureau for Abbott. He consults for, advises, and received grants from Gilead. He is a co-founder of Illuminatio Medical Technology Limited.

Jorn Schattenberg consults for, is on the speakers' bureau for and received grants from Boehringer Ingelheim and Histoindex. He consults for and is on the speakers' bureau for Novo Nordisk, Madrigal, and Echosens. He consults for and received grants from Gilead. He consults for Apollo Endosurgery, Albireo, Bayer, BMS, GSK, Intercept Pharmaceuticals, Ipsen, Inventiva Pharma, MSD, Northsea Therapeutics, Novartis, Pfizer, Roche, Sanofi, and Siemens Healthineers. He is on the speakers' bureau for MedPublico GmbH. He received grants from Siemens Healthcare GmbH.

Won Kim consults for and is on the speakers' bureau for Boehringer-Ingelheim, Novonordisk, HK Inoen, Standigm, PharmaKing, KOBIOLABS, Ildong, Olix Pharma, Samil, TSD Life Sciences, Daewoong Pharmaceutical and Eisai. He received grants from Gilead, Novartis, Pfizer, Roche, Ildong, Galmed, Dicerna, Celgene and Enyo. He owns stocks in KOBIOLABS and Lepidyne.

Katharina Staufer is an employee of Versantis AG.

Michael Trauner advises, is on the speakers' bureau for and received grants from Abbvie, Albireo, BIOMx, BI, Falk, Gilead, Genfit, Hightide, Intercept, Janssen, MSD, Novartis, Phenex, Pliant, Regulus, Siemens, Shire, and BMS. He received grants from Alnylam, Cymabay, Takeda, and UltraGenyx He is listed as a co-inventor on a patent filed by The Medical Universities of Graz and Vienna on medical use of norUDCA.

Elisabetta Bugianesi consults for and advises Novo Nordisk. She consults for and is on the speakers' bureau for MSD. She is on the speakers' bureau for and received grants from Gilead. She consults for Boehringer Ingelheim, Lilly and Novo Nordisk.

Ramy Younes is employed by Boehringer Ingelheim.

Ming-Hua Zheng is on the speakers' bureau for Hisky Medical.

Calogero Camma advises Eisai, Ipsen, Roche and AstraZeneca.

Quentin M. Anstee consults for, on behalf of Newcastle University, 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, NGMBio, Novartis, Novo Nordisk, PathAI, Pfizer, Prosciento, Poxel, Resolution Therapeutics, Roche, Ridgeline Therapeutics, RTI, Shionogi, and Terns. He is on the speaker' bureau for Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal, Medscape, and Springer Healthcare. He received grants from AstraZeneca, Boehringer Ingelheim, and Intercept. He serves on the DSMB, on behalf of Newcastle University, for Medpace (North Sea Therapeutics). Michael Pavlides owns stock in Perspectum Ltd.

Salvatore Petta advises and is on the speakers' bureau for AbbVie, Echosens, Gilead, Intercept, MSD, Novonordisk, and Pfizer**.**

Author contributions: FEM: data curation, MP: data curation and individual patient data, JFLC: individual patient data. All authors contributed significant intellectual content and approved the final manuscript.

Acknowledgements: This individual patient data meta-analysis is being conducted as part of the imaging study in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) project. The LITMUS project is funded by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under Grant Agreement 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the view of the authors and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

QMA is supported by the Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne, UK.

MP and JFLC acknowledge support from the Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK

SP acknowledge support from Ministero della Salute Italiana project PNRR-MAD-2022-12375656

Graphical Abstract

GA1

Abstract

Background: We evaluated the diagnostic accuracy of simple non-invasive tests(NITs) in NAFLD patients with type 2 diabetes(T2D).

Methods: This was an individual patient data meta-analysis of 1780 patients with biopsy-proven NAFLD and T2D. The index tests of interest were FIB-4, NAFLD Fibrosis Score(NFS), APRI, liver stiffness measurement(LSM) by vibration-controlled transient elastography(VCTE) and AGILE 3+. The target conditions were advanced fibrosis, nonalcoholic steatohepatitis(NASH) and fibrotic NASH(NASH plus F2-F4 fibrosis). The diagnostic performance of NITs individually or in sequential combination was assessed by area under receiver operating characteristic curve(AUROC) and by decision curve analysis(DCA). Comparison with 2278 NAFLD patients without T2D was also made.

Results: In NAFLD with T2D LSM and AGILE 3+ outperformed both NFS and FIB-4 for advanced fibrosis(AUROC:LSM 0.82,AGILE 3+ 0.82,NFS 0.72,FIB-4 0.75,APRI 0.68;p<0.001 of LSM-based vs simple serum tests), with an uncertainty area of 12%-20%.The combination of serum-based with LSM-based tests for advanced fibrosis led to a reduction of 40% to 60% in necessary LSM tests. DCA showed that all scores had modest net benefit for ruling-out advanced fibrosis at the risk threshold of 5%-10% of missing advanced fibrosis. LSM and AGILE 3+ outperformed both NFS and FIB-4 for fibrotic NASH(AUROC LSM 0.79,AGILE 3+ 0.77,NFS 0.71,FIB-4 0.71;p<0.001 of LSM-based vs simple serum tests). All noninvasive scores were sub-optimal for diagnosing NASH.

Conclusions:LSM and AGILE 3+ individually or in low availability setting in sequential combination after FIB-4 or NFS have a similar good diagnostic accuracy for advanced fibrosis and an acceptable diagnostic accuracy for fibrotic NASH in NAFLD patients with T2D.

Introduction

Nonalcoholic fatty liver disease (NAFLD), affecting roughly 25% of the general adult population [1], is a leading cause of chronic liver disease [2]. NAFLD complications [3,4], and the severity of liver fibrosis are the main drivers of prognosis in NAFLD, with more severe liver fibrosis incurring higher risk of developing liver-related events (LRE; hepatocellular carcinoma and liver decompensation) and extrahepatic events (mostly cardiovascular events and extrahepatic cancer) [5,6].

NAFLD and type 2 diabetes (T2D) have a complex bidirectional interplay: NAFLD increases the risk of T2D development [7], and T2D is a risk factor for NAFLD occurrence, severity, and progression toward liver cirrhosis and its complications [8]. Consequently, the estimated prevalence of NAFLD in people with T2D is about 55%, and -most relevant- nonalcoholic steatohepatitis (NASH) and advanced fibrosis can be observed in about 37% and 17%, respectively, of patients with T2D [9].

The high prevalence of NAFLD and of NAFLD-related liver damage in patients with T2D led clinical guidelines to encourage screening for advanced fibrosis in patients with metabolic dysfunctions including those with T2D [10,11]. For this purpose, noninvasive scores like FIB-4 and NFS, and liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE) have been largely validated as accurate tools to exclude advanced fibrosis in NAFLD [12,13], and their rational use is recommended by international guidelines [10]. However, preliminary evidence suggests a poor accuracy of these scores/tools in patients with T2D [12,14,15], finally leading to high referral rates for expert evaluation [16]. Moreover, the AGILE 3+ score, based on the simultaneous combination of AST/ALT ratio, platelet count, T2D status, sex, age and LSM by VCTE, has been recently developed and proposed for the diagnosis of advanced fibrosis in NAFLD [17,18], but data on its performance in the diabetic population are lacking. On the other side, the identification of NAFLD patients with NASH, and -most relevant- with fibrotic NASH, especially

in high-risk groups like patients with T2D, is an important need for inclusion in phase 2b and phase 3 clinical trials assessing pharmacological treatment of NASH patients [19,20].

Our aim was thus to explore the diagnostic accuracy of simple serum based noninvasive scores and LSM by VCTE for the diagnosis of advanced fibrosis, NASH, and fibrotic NASH in a large cohort of patients with histological diagnosis of NAFLD and T2D. A comparison with NAFLD patients without T2D was also made.

Patients & METHODS

Patients

For the present study we utilized the subgroup of 1780 patients with histological diagnosis of NAFLD and T2D from a previously published individual patient data meta-analysis of 37 studies, that aimed to assess the accuracy of LSM by VCTE and noninvasive scores for ruling-out advanced fibrosis in biopsy-confirmed NAFLD patients [12]. All authors who had provided data for the original individual patient data meta-analysis (IPDMA) were contacted with details of the present study, and their data were only included with their agreement. In the present analysis we considered all but five studies included in the IPDMA [21-25] because the authors of those five studies have not responded to email that was asking for their consent for participating in this sub-analysis. Search details, inclusion criteria and quality assessment of the studies were reported in the original study [12]; literature search for eligible studies for this IPDMA stopped at April 2020.
Briefly, studies reporting data on adults (≥18 years) with NAFLD after exclusion of other causes of liver diseases and paired liver histology and LSM by VCTE were eligible. All studies were considered if the interval of time between liver biopsy, LSM and noninvasive scores was within 6 months. The diagnosis of T2D was made according to the American Diabetes Association [26], using a value of fasting blood glucose ≥ 126 mg/dl, or based on the use of anti-diabetic therapy. In patients with a previous diagnosis of T2D, current medications were documented. Finally, only

studies reporting histological classification of liver fibrosis based on the non-alcoholic steatohepatitis Clinical Research Network (NASH CRN) staging system [27] were considered.

Patients and the public were not involved in the conduct of this study as there was no direct patient participation in the study.

Assessment of liver histology

Liver histology was based on local reporting from the original studies based on the NASH CRN staging system [27]. NASH was defined by the presence of a nonalcoholic steatohepatitis activity score (NAS) >3 with at least grade 1 in each component; Fibrotic NASH was defined by presence of NASH plus fibrosis stage F2-F4; Advanced fibrosis was defined by presence of fibrosis stage F3-F4.

Noninvasive assessment of liver fibrosis

The FIB-4 (comprising age, AST, ALT, and PLT) score was calculated using the original reported formula and patients were classified as low risk of advanced fibrosis if FIB-4 <1.30, intermediate risk if FIB-4 was between 1.30 and 2.67, and high risk if FIB-4 >2.67 [28].

The NFS (comprising age, BMI, AST, ALT, albumin, PLT, and T2D status) score was calculated using the original reported formula and patients were classified as low risk of advanced fibrosis if NFS <-1.455, intermediate risk if NFS was between -1.455 and 0.675, and high risk if NFS >0.675 [29].

The AST-to-platelet ratio index (APRI) was also computed [30].

Vibration-controlled transient elastography was performed with the FibroScan (Echosens, Paris, France) medical device. For this meta-analysis, if only one VCTE-based LSM was available then this was included in the main analysis irrespective of probe type and BMI. Where two VCTE-based LSM were available (one with each probe), the main analysis included the M-probe measurement for BMI < 30 kg/m^2 and the XL probe measurement for BMI ≥30 kg/m^2. Therefore,

all LSM cut-offs were determined independent of probe type. LSM <7.9 kPa was defined as indicating a low risk of F3-F4 fibrosis; LSM 7.9-9.6 kPa as an intermediate risk; LSM >9.6 kPa as a high risk [31].

The AGILE 3+ score (comprising age, sex, AST, ALT, PLT, T2D status, and LSM) was calculated using the original reported formula and patients were classified as low risk of advanced fibrosis if AGILE 3+ <0.45, intermediate risk if AGILE 3+ was between 0.45 and 0.67, and high risk if AGILE 3+ ≥0.68 [17,18].

Statistics

Data for continuous variables were expressed as mean and standard deviation or median and interquartile range, and data for categorical variables were expressed as frequency and percentage. Differences between continuous data were assessed by Student's t test or by the Mann-Whitney U test. Differences between categorical variables were assessed by the χ^2 test.

The accuracy of each score for detection of advanced fibrosis (F3-F4), NASH, and fibrotic NASH (NASH plus F2-F4 fibrosis), was assessed using the area under receiver operating characteristic curves described as AUROC. AUROCs were compared using De Long's test statistic. Cut-off points of LSM, NFS, FIB-4 and AGILE-3 + for the advanced fibrosis model were derived from literature. Specific cut-offs with sensitivity >90% for ruling-out or specificity >90% for ruling in all outcomes were calculated, and for this purpose the cohort was split in a into a training (cohorts with >=100 enrolled patients) and a validation (cohorts with <100 enrolled patients). Accordingly, false negative and false positive rates of the single test, as well as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated.

Finally, we also evaluated the accuracy of sequential combination strategies based on FIB-4 or NFS as first test, and LSM or AGILE 3+ as second test in those with FIB-4 or NFS values higher than the rule-out cut-offs.

The main analysis was conducted to maximize data for each NIT. For a valid comparison of the performance of NITs, a separate analysis was conducted in the subgroup of patients with a complete dataset.

Subgroup analysis was performed according to age (<35 years, 35-65 years, >65 years), body mass index (BMI; BMI<30 kg/m^2, BMI ≥30 kg/m^2), and aminotransferase levels (normal versus abnormal ALT). For this last subgroup analysis, the upper limit of normal for ALT is 19 IU\L for women and 30 IU\L for men [32]. We also evaluated the overall accuracy of LSM in the subgroup of patients with BMI <30 kg/m^2 measured with the M probe and with BMI >=30 kg/m^2 measured with XL probe.

As AUROC focuses only on the predictive accuracy of a model, not considering cases where a false-negative result is more harmful than a false-positive result, we also performed a decision curve analysis (DCA) for identifying threshold probabilities at which use of non-invasive criteria will translate into maximum net benefit of detecting advanced fibrosis [33,34].

DCA evaluated prediction models in comparison with default strategies of performing liver biopsy in all patients or none allowing an assessment of overall yield of prediction rules. DCA estimates a "net benefit" for each of prediction rule, defined as

net benefit = sensitivity × prevalence – (1–specificity) × (1–prevalence) × w

where *w* is the odds of true diagnosis (i.e., advanced fibrosis in this case) across different threshold probabilities. In this setting, net benefit represents a composite of the benefit gained by performing liver biopsy for true advanced fibrosis in patients classified as high-risk according to non-invasive scores (true positive) and risk/ discomfort incurred due to liver biopsy in those without advanced fibrosis but who were classified as high-risk according to non-invasive scores (false positive). Threshold probability represents a theoretical risk level where the expected benefit of treatment is equal to the expected risk of avoiding treatment (e.g., benefit of liver biopsy equals risk of not

performing it). Thus, net benefit is assessed across a range of threshold probabilities to identify the best diagnostic strategy for different risk-scenarios.

All data were analysed using R Studio. DCA was implemented in R using code derived from Zhang et al [35,36]. In addition to the base packages in R, tidy verse, survival, survminer, boot, reshape2, and readxl packages were used.

RESULTS

Features of patients with NAFLD and T2D

Baseline characteristics of the 1780 patients with NAFLD and T2D stratified for advanced fibrosis, NASH and fibrotic NASH are shown in **Supplemental table 1 http://links.lww.com/HEP/E554**.

LSM was determined in 1692 (95%), FIB-4 in 1681 (94%), NFS in 1001 (56%) and AGILE 3+ in 1603 (90%.) patients. Overall, 46.2% of patients had advanced fibrosis, 77.4% of patients had NASH and 55.8% of patients had fibrotic NASH.

Supplemental table 2 and 3 http://links.lww.com/HEP/E554 report baseline characteristics of the 748 patients with a complete dataset.

Diagnostic accuracy of noninvasive scores/tools for advanced fibrosis in patients with NAFLD and T2D

LSM, FIB-4, NFS, APRI and AGILE 3+ had AUROCs of 0.82, 0.75, 0.72, 0.68, 0.82 for advanced fibrosis (**Table 1**, **Figure 1A**). FIB-4 had a similar acceptable diagnostic accuracy as NFS (p=0.30) and worked significantly better than APRI (p<0.001). LSM and AGILE 3+ had good performance and performed similarly (p=0.60) and significantly better than all serum-based tests (p<0.001 for all comparisons). These results were confirmed when performing a head-to-head comparison in the cohort with a complete dataset (**Supplemental Figure 1A** http://links.lww.com/HEP/E555 and **Supplemental Table 4 http://links.lww.com/HEP/E554**).

Considering the poor accuracy of APRI, further analyses did no longer consider this score.

Analyses considering cut-offs from the literature for the diagnosis of advanced fibrosis are reported in **Table 2**. Proportions of patients classified as having low, intermediate and high risk of advanced fibrosis were 46%, 40% and 14% by using FIB-4, 27.6%, 55.8% and 16.6% by using NFS, 35.8%, 12.9% and 51.2% by using LSM, and 36.4%, 20.1% and 43.5%, respectively, by using AGILE 3+ respectively. Consequently, FIB-4 had the highest proportion of patients at low risk of advanced fibrosis, and LSM the lowest proportion of patients falling into the uncertainty area. NFS and LSM had the highest sensitivity (88%), LSM the highest NPV (84%), FIB-4 the highest specificity (93%) and the highest PPV (75%) (**Table 2**). Similar results were observed when comparing the scores in the cohort with a complete dataset (**Supplemental Table 5 http://links.lww.com/HEP/E554**).

We further evaluated the performance of LSM, FIB-4, NFS and AGILE 3+ to diagnose advanced fibrosis in sequential combinations. When selecting threshold combinations for FIB-4 (<1.3) and NFS (<−1.455) available in the literature and pairing them with the best threshold pair for LSM (<7.9 kPa and ≥9.6 kPa) or AGILE-3+ <0.45 and ≥0.68), the FIB-4→LSM strategy lead to the highest proportion of patients identified as being at low risk of advanced fibrosis, and the NFS→LSM strategy to the lowest proportion of patients falling in the uncertainty area (**Table 3**). Furthermore, NFS→LSM and NFS→AGILE 3+ strategies lead to the highest sensitivity (79% and

80%, respectively) and NPV (84% and 82%, respectively), while the FIB-4→LSM and FIB-4→AGILE 3+ strategies lead to the highest specificity (86% and 84%, respectively) and PPV (79% and 76%, respectively) (**Table 3**). Similar results were observed when comparing the scores in the cohort with a complete dataset (**Supplemental Table 6 http://links.lww.com/HEP/E554**).

The net benefit of FIB-4, NFS, LSM and AGILE 3+ scores for ruling out advanced fibrosis at 5%, 10% and 15% threshold probabilities of missing advanced fibrosis is shown in **Figure 2**. At the risk thresholds of 5% and 10% of missing advanced fibrosis, all scores/tools showed no benefit for ruling out advanced fibrosis compared to the strategy of performing liver biopsy in all patients, while, at the risk threshold of 15% the observed net benefit was modest and LSM outperformed AGILE 3+, NFS and FIB-4. Results obtained for ruling in advanced fibrosis are showed in **Supplemental Figure 2 http://links.lww.com/HEP/E555**. When considering strategies based on the combination of FIB-4 or NFS with LSM or AGILE 3+, these showed no benefit for ruling out advanced fibrosis at the risk thresholds of 5% and 10% of missing advanced fibrosis (**Supplemental Figure 3 http://links.lww.com/HEP/E555**).

Identification of best cut-offs for advanced fibrosis in patients with NAFLD and T2D

Considering the unique opportunity to have a large cohort of patients with histological diagnosis of NAFLD and T2D we split the population into a training and a validation set to search for best rule-out and rule-in cut-offs for advanced fibrosis. Differences between training and validation cohorts are reported in **Supplemental Table 7 http://links.lww.com/HEP/E554.**

These analyses are reported in **Table 4**. Notably, the accuracy of the new proposed cut-offs was replicated in the validation set where NPV and PPV of about 80% were maintained at the cost of an uncertainty area of about 35%-38% for LSM and AGILE 3+, and of about 45%-58% for NFS and FIB-4 (**Table 4**).

Net benefit of FIB-4, NFS, LSM and AGILE 3+ scores by using these new cut-offs for ruling out advanced fibrosis at 5%, 10% and 15% threshold probabilities of missing advanced fibrosis is showed in **Figure 3A**.

Comparison of diagnostic accuracy of noninvasive scores/tools for advanced fibrosis between NAFLD patients with or without T2D

The baseline characteristics of the 2278 NAFLD patients with T2D arising from the same studies considered in this IPDMA, respect to NAFLD patients without T2D are shown in **Supplemental table 8 http://links.lww.com/HEP/E554.**

Supplemental Table 9 http://links.lww.com/HEP/E554 reports the comparison of AUROCs of LSM, FIB-4, NFS, and AGILE 3+ according to T2D status. All noninvasive scores performed similarly in NAFLD patients with T2D compared to those without for predicting advanced fibrosis (**Supplemental Table 9 http://links.lww.com/HEP/E554**).

When considering cut-offs from the literature for the diagnosis of advanced fibrosis, in nondiabetic patients respect to population with NAFLD and T2D, noninvasive scores had lower sensitivity -except for FIB-4 that was similar- but higher specificity, and presented a lower

uncertainty area -except for LSM that was similar **(Supplemental Table 10 http://links.lww.com/HEP/E554)**.

The diagnostic accuracy of LSM, FIB-4, NFS and AGILE 3+ in sequential combinations is reported in **Supplemental Table 11 http://links.lww.com/HEP/E554**. Respect to NAFLD patients with T2D, in nondiabetics the sequential combination of FIB-4 or NFS with LSM or AGILE 3+, generated a lower uncertainty area and higher specificity but lower sensitivity especially for NFS-based algorithms (**Supplemental Table 11 http://links.lww.com/HEP/E554).**

The net benefit of FIB-4, NFS, LSM and AGILE 3+ scores alone or in combination for ruling out advanced fibrosis at 5%, 10% and 15% threshold probabilities of missing advanced fibrosis is shown in **Supplemental Figure 4 http://links.lww.com/HEP/E555**. Respect to population of NAFLD with T2D, the net benefit for ruling-out advanced fibrosis at 5%,10% and 15% risk threshold was higher for all noninvasive scores.

The results about the search for best rule-out and rule-in cut-offs for advanced fibrosis are reported in **Supplemental Table 12 http://links.lww.com/HEP/E554, Supplemental Figure 5 http://links.lww.com/HEP/E555, Supplemental Table 13 http://links.lww.com/HEP/E554**. Respect to diabetic population, best rule-in and rule-out cut-offs for nondiabetic patients generated a smaller uncertainty area; FIB-4 best cut-offs were similar between diabetic and nondiabetic patients, while rule-in cut-off for LSM was lower in nondiabetic compared to diabetic patients.

Diagnostic accuracy of noninvasive scores/tools for NASH and fibrotic NASH in patients with NAFLD and T2D

LSM, FIB-4, NFS, APRI and AGILE-3 + had corresponding AUROCs of 0.71, 0.65, 0.66, 0.70, 0.69 for identifying NASH (**Table 1**, **Figure 1B**), and of 0.79, 0.71, 0.71, 0.70, 0.77 for fibrotic NASH (**Table 1**, **Figure 1C**). Consistently, all noninvasive tools tested here poorly predicted the presence of NASH, while LSM and AGILE 3+ have an acceptable accuracy for detecting fibrotic

NASH, LSM being significantly better than all the other simple serum-based tests ($p<0.01$ for all) and with a similar performance as AGILE 3 + ($p=0.87$). These trends were confirmed when performing a head-to-head comparison of LSM, FIB-4, NFS and AGILE-3 + in the cohort with a complete dataset (**Supplemental table 4 http://links.lww.com/HEP/E554**).

The best rule-out and rule-in cut-offs for NASH as well as their operating characteristics are reported in **Supplemental Table 14 http://links.lww.com/HEP/E554**. All scores/tools showed no benefit for ruling out NASH (**Figure 3B**).

Table 4 shows the best rule-out and rule-in cut-offs for fibrotic NASH as well as their operating characteristics in both training and validation sets. LSM identified 1 patient in 3/4 as at high risk of fibrotic NASH at a specificity and a PPV of 91% and 80%, respectively, leading to an uncertainty area of 45.3%. At the risk threshold of 5%, 10% and 15% of missing fibrotic NASH, all nonivasive scores had not benefit for ruling out fibrotic NASH (**Figure 3C**).

Comparison of diagnostic accuracy of noninvasive scores/tools for NASH and fibrotic NASH between patients with NAFLD with or without T2D

All noninvasive scores performed similarly poor for diagnosing NASH in both diabetic and nondiabetic cohorts. Otherwise, their accuracy for predicting fibrotic NASH was significantly better in nondiabetic compared to diabetic patients (**Supplemental Table 9 http://links.lww.com/HEP/E554**).

Respect to diabetic population, best rule-in and rule-out cut-offs for fibrotic NASH in nondiabetic patients generated a smaller uncertainty area for LSM; FIB-4 best cut-offs were similar between diabetic and nondiabetic patients, while rule-in cut-off for LSM was lower in nondiabetic compared to diabetic patients **Supplemental Table 12 http://links.lww.com/HEP/E554 and Supplemental Figure 5**).

Subgroup analyses in patients with NAFLD and T2D

Subgroup analyses for the diagnosis of advanced fibrosis, NASH and fibrotic NASH are reported in **Table 5**. When looking at the diagnosis of advanced fibrosis, all scores/tools had a trend for a better accuracy in patients older than 35 years and performed significantly better in non-obese patients; FIB-4 and AGILE 3+ had significantly higher accuracy in patients with normal ALT, while in the same sub-groups NFS and LSM had a non-significant trend for a better performance (**Table 5**). Finally, looking at the diagnosis of fibrotic NASH, the only significant difference was for a higher accuracy of LSM in patients with normal ALT.

Finally, when looking at patients where LSM was performed by using the M probe in non-obese and the XL probe in obese patients, the overall accuracy for advanced fibrosis, NASH and fibrotic NASH was 0.86, 0.72 and 0.81 for LSM, and 0.85, 0.71 and 0.80 for AGILE 3+; these results were similar to those obtained in the entire cohort. In this sub-group we confirmed a higher accuracy of LSM in non-obese patients compared to obese patients (AUROCs 0.865 vs 0.802, p=0.04).

DISCUSSION

In this study on a large cohort of patients with histological diagnosis of NAFLD and T2D, we provided evidence that LSM and AGILE 3+ have a good diagnostic accuracy for advanced fibrosis and an acceptable diagnostic accuracy for fibrotic NASH, while AGILE 3+ did not provide any additional relevant diagnostic insights over and above LSM alone. Overall, both LSM and AGILE 3+ outperformed FIB-4 and NFS that showed an acceptable performance. The sequential combination of serum-based tests with LSM-based tests for advanced fibrosis allowed to limit the number of LSM-based tests -mostly with FIB-4-. Furthermore, DCA showed that the net benefit for the ruling out advanced fibrosis and fibrotic NASH was modest for all tools. In comparison to NAFLD patients without T2D, the overall accuracy of NITs for advanced fibrosis was similar even if with a lower net benefit mainly related to lower specificity and higher uncertainty area, while the accuracy for fibrotic NASH was lower.

In this large IPDMA on patients with NAFLD and T2D, simple serum-based tests lead to a high uncertainty area ranging from 40% for FIB-4 to 55% for NFS, the latter leading to the highest sensitivity (88%) and FIB-4 to the highest specificity (93%). Otherwise, LSM and AGILE 3+ were characterized by a low uncertainty area ranging from 12% to 20% with highest sensitivity of 88% for LSM and highest specificity of 78% for AGILE 3+. Overall, these data, according to European and American guidelines [10,11], suggest, when available and in tertiary setting, to use LSM-based tests as first tests, while demanding the use of simple serum scores where LSM is not available. AGILE 3+ can be an alternative to LSM but, at least in a diabetic population, does not provide any additional relevant diagnostic insights. We also tested the strategy of using serum-based scores as triage and to refer for LSM-based tests when patients were at intermediate-to-high risk by simple serum-based scores. This strategy led to a relevant reduction in the proportion of patients to an uncertainty area ranging from 8% to 16%, keeping the highest sensitivities for NFS-based algorithms (79%-80%) and the highest specificities for FIB-4-base algorithms (84%-86%). These results confirm that the sequential combination strategies can be useful also in the setting of NAFLD with T2D, even a big proportion of patients (about 50% for FIB-4 and about 70% for NFS) is worthy of being referred for LSM-based assessment. From a clinical point of view, FIB-4 based strategies may be preferred because they can spare more LSM compared with NFS (**Supplemental Figure 6 http://links.lww.com/HEP/E555**). Notably, in our decision curve analysis noninvasive tests and their combinations showed similar modest net benefit for ruling out advanced fibrosis at threshold probabilities of 5% and 10% of missing advanced fibrosis.

In our study we also identified the best rule-in (90% specificity) and rule-out (90% sensitivity) thresholds for advanced fibrosis to be applied in the setting of diabetic patients. The higher sensitivity and specificity of these new cut-offs were at the cost of a higher uncertainty area ranging from about 42% for LSM to about 61% for NFS. Consequently, at the moment traditional cut-offs applied in general NAFLD population should be recommended also in the diabetic setting.

Our study also observed that in patients with NAFLD and T2D, LSM and its related score AGILE 3+ had a significantly lower accuracy for the diagnosis of advanced fibrosis in obese patients compared to nonobese patients and in those with elevated ALT compared to their counterpart. This finding confirms what was already been reported in the overall NAFLD population [37]. When looking at BMI, our results can raise the doubt that the lower accuracy of LSM in obese patients could be due to the use of the M instead of the XL probe. In an attempt to solve this question, we confirmed that the accuracy of LSM in the sub-group of patients where it was measured by M probe in non-obese and XL probe in obese was higher in non-obese patients compared to obese patients. Further studies assessing skin-to-capsule distance, could add insights about this topic. On the other side, evidence in NAFLD already demonstrated that high ALT levels affect the accuracy of LSM for fibrosis by overestimating liver damage [37]. When looking at NFS and FIB-4 we observed that these scores performed better in patients older than 35 years, who were non-obese and had normal ALT values. These data have already been reported in the NAFLD population, and they can be explained by the fact that these variables –included in the scores- are associated between them and with advanced fibrosis, but are also present in the absence of advanced fibrosis, therefore sometimes lowering the accuracy of non-invasive scores.

The comparison of NAFLD population with T2D to that without, showed that noninvasive scores have a similar diagnostic accuracy for advanced fibrosis in terms of AUROCs, even if they have lower sensitivity -except for FIB-4- but higher specificity and lower uncertainty area -except for LSM- in NAFLD without T2D. This trend was confirmed also when looking at sequential combination strategies, finally leading to a higher -even if modest- net benefit of noninvasive scores for ruling-out advanced fibrosis in NAFLD patients without T2D. Notably, we also observed that 90% specificity rule-in cut-off of LSM for advanced fibrosis was higher in NAFLD patients with T2D respect to those without (14.6 KPa versus 11.8 KPa), this result being worthy to further validation clinical practice.

International regulatory agencies identified patients with fibrotic NASH as those eligible for clinical trials testing new pharmacological agents for NASH. In our study we found that LSM and AGILE 3+, although originally developed or diagnosing liver fibrosis, overall outperformed NFS and FIB-4, and had an acceptable accuracy for the diagnosis of fibrotic NASH. When looking at the best rule-in and rule-out cut-offs the use of LSM and AGILE 3+ identified about 1 patient in 3/4 at high risk of fibrotic NASH, keeping a specificity >90%. As for advanced fibrosis, in our DCA only LSM showed a small benefit for ruling out fibrotic NASH at threshold probabilities of 5%, 10% and 15% of missing fibrotic NASH, confirming the need for liver biopsy for a correct identification of patients with fibrotic NASH, especially if the patient is eligible for inclusion in clinical trials. When comparing NAFLD population with T2D to those without, we found that the diagnostic accuracy of all noninvasive scores for fibrotic NASH was significantly lower in NAFLD patients with T2D. Notably, when looking at LSM we also observed a lower uncertainty area, and a higher rule-in cut-off for fibrotic NASH in NAFLD patients with T2D respect to those without (14 KPa versus 11.8 KPa). Different scores like NIS4 [38] or MACK3 [39] or cT1-AST-fasting glucose (cTAG) [40] have been recently proposed for the non-invasive identification of NAFLD patients with fibrotic NASH, but limited external validation and, most importantly, the use of not easily available unconventional variables limit their use in clinical practice. Along this line MRE-based indices like MEFIB [41] and MAST [42] showed high PPV for fibrotic NASH but the cost and availability of MRE limit their validation and their use in clinical practice. Otherwise, a simpler score, called FAST [43], based on the combination of LSM, controlled attenuation parameter (CAP) and AST, has been recently shown to have good accuracy with an AUROC ranging from 0.74 to 0.85, a proportion of at high risk patients ranging from 4% to 36%, and a specificity ranging from 82% to 99%. FAST could not be investigated in this IPDMA dataset because CAP was not available. Further studies in the setting of diabetic patients may demonstrate the superiority of FAST score - the today standard for the diagnosis of fibrotic NASH- to LSM alone.

The main limitation of the present study lies in its potentially limited validity of the results in different populations and settings. Since our study includes patients referred to tertiary hepatological referral centers for suspected liver damage, it is possible that the obtained results could not be replicated in general diabetic populations differing for age, biochemical alterations, severity of T2D and metabolic comorbidities. Along this line, the observed relatively high PPV and relatively low NPV of studied scores can be related to the high prevalence of advanced fibrosis, NASH and fibrotic NASH in our population respect to what observed in diabetic general population. Consequently, it could be possible that other cut-offs might be better in general diabetic populations with a lower prevalence of the investigated outcomes. The exclusion from IPDMA of studies using screening strategies other than LSM, and the hypo representation of North or South America populations where T2D is highly prevalent, could further limit the generalizability of our results. The allowed 6-month interval between NITs and liver biopsy could also affect the interpretation of the results: this interval time could be considered substantial for NASH and fibrotic NASH where inflammation and steatosis components can significantly change within this timeframe. Moreover, the observed performance of NITs in sub-group analyses can be affected by the spectrum bias effect. Lack of an external validation cohort, of central biopsy reading, and potentially hidden alcohol abuse in some patients could further limit the interpretation of our results.

In conclusion, we demonstrated that in the setting of NAFLD patients with T2D, LSM and AGILE 3+ have a similar good and an acceptable diagnostic accuracy for the diagnosis of advanced fibrosis and fibrotic NASH, respectively. In a context of their limited availability, the sequential combination of serum-based with LSM-based tests for advanced fibrosis lead to a reduction of about 40% to 60% in necessary LSM tests keeping sensitivity and specificity ≥80% for NFS-based and FIB-4-based combinations.

References

1) Le MH et al. 2019 Global NAFLD Prevalence: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2021 Dec 7:S1542-3565(21)01280-5.
2) Harris R, et al. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. Lancet Gastroenterol Hepatol. 2017 Apr;2(4):288-297.
3) Gu W, et al. Trends and the course of liver cirrhosis and its complications in Germany: Nationwide population-based study (2005 to 2018). Lancet Reg Health Eur. 2021 Nov 4;12:100240. doi:10.1016/j.lanepe.2021.100240.
4) Vitale A, et al. Epidemiological trends and trajectories of MAFLD-associated hepatocellular carcinoma 2002-2033: the ITA.LI.CA database. Gut. 2021 Dec 21:gutjnl-2021-324915. doi: 10.1136/gutjnl-2021-324915.
5) Taylor RS, et al. Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. Gastroenterology. 2020 May;158(6):1611-1625.e12.
6) Pennisi G, et al Liver-related and extrahepatic events in patients with non-alcoholic fatty liver disease: a retrospective competing risks analysis. Aliment Pharmacol Ther. 2022 Mar;55(5):604-615.
7) Mantovani A, et al. Nonalcoholic fatty liver disease and risk of incident diabetes mellitus: an updated meta-analysis of 501 022 adult individuals. Gut. 2021 May;70(5):962-969.
8) Kanwal F, et al. Effect of Metabolic Traits on the Risk of Cirrhosis and Hepatocellular Cancer in Nonalcoholic Fatty Liver Disease. Hepatology. 2020 Mar;71(3):808-819.
9) Younossi ZM, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. J Hepatol. 2019 Oct;71(4):793-801.
10) European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol. 2021 Sep;75(3):659-689.
11) American Diabetes Association. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: *Standards of Medical Care in Diabetes-2020*. Diabetes Care. 2020;**43**:S37–S47.
12) Mózes FE, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. Gut. 2022 May;71(5):1006-1019.

13) Selvaraj EA, et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: A systematic review and meta-analysis. J Hepatol. 2021 Oct;75(4):770-785.

14) Bril F, et al. Performance of Plasma Biomarkers and Diagnostic Panels for Nonalcoholic Steatohepatitis and Advanced Fibrosis in Patients With Type 2 Diabetes. Diabetes Care. 2020 Feb;43(2):290-297.

15) Boursier J, et al. Impact of Type 2 Diabetes on the Accuracy of Noninvasive Tests of Liver Fibrosis With Resulting Clinical Implications. Clin Gastroenterol Hepatol. 2022 Mar 11:S1542-3565(22)00248-8.

16) Blank V, et al. Current NAFLD guidelines for risk stratification in diabetic patients have poor diagnostic discrimination. Sci Rep. 2020 Oct 27;10(1):18345.

17) Younossi Z, et al. Development and validation of Agile 3+: novel FibroScan based score for the diagnosis of advanced fibrosis in patients with nonalcoholic fatty liver disease. JOURNAL OF HEPATOLOGY VOLUME 75, SUPPLEMENT 2, PAGES S191–S866, OS 555.

18) Pennisi G, et al. AGILE 3+ Score for the diagnosis of advanced fibrosis and for predicting liver-related events in NAFLD. Clin Gastroenterol Hepatol. 2022 Jul 13:S1542-3565(22)00646-2.

19) US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: developin drugs for treatment. Guidance for industry [draft guidance]. US Food and Drug Administration. Published December 2018. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/noncirrhotic-nonalcoholicsteatohepatitis-liver-fibrosis-developing-drugs-treatment. Accessed June 14, 2021.

20) Loomba R, et al. Expert Panel Review to Compare FDA and EMA Guidance on Drug Development and Endpoints in Nonalcoholic Steatohepatitis. Gastroenterology. 2022 Mar;162(3):680-688. doi: 10.1053/j.gastro.2021.10.051.

21) Okajima A, et al. Liver stiffness measurement to platelet ratio index predicts the stage of liver fibrosis in non-alcoholic fatty liver disease. Hepatol Res 2017;47:721–30.

22) Ooi GJ, et al. Evaluating feasibility and accuracy of non-invasive tests for nonalcoholic fatty liver disease in severe and morbid obesity. Int J Obes 2018;42:1900–11.

23) Seki K, et al. Assessment of transient elastography in Japanese patients with non-alcoholic fatty liver disease. Hepatol Res 2017;47:882–9.

24) Ziol M, et al. Relationships between fibrosis amounts assessed by morphometry and liver stiffness measurements in chronic hepatitis or steatohepatitis. Eur J Gastroenterol Hepatol 2009;21:1261–8.

25) Garg H, et al. Utility of transient elastography (fibroscan) and impact of bariatric surgery on nonalcoholic fatty liver disease (NAFLD) in morbidly obese patients. Surg Obes Relat Dis. 2018 Jan;14(1):81-91.

26) Introduction: Standards of Medical Care in Diabetes-2020. Diabetes Care. 2020 Jan;43(Suppl 1):S1-S2. doi: 10.2337/dc20-Sint. PMID: 31862741

27) Kleiner DE, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology2005; 411:313–21

28) McPherson S, et al. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. Gut. 2010 Sep;59(9):1265-9.

29) Angulo P, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007 Apr;45(4):846-54.

30) Lin Z-H, et al. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. Hepatology 2011;53:726–36.

31) Wong VW, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology. 2010 Feb;51(2):454-62.

32) Prati D, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med 2002; 137:1-10.

33) Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74.

34) Vickers AJ, et al. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18.

35) R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

36) Zhang Z, et al. Decision curve analysis: a technical note. Ann Transl Med. 2018;6:308.

37) Petta S, et al. Impact of Obesity and Alanine Aminotransferase Levels on the Diagnostic Accuracy for Advanced Liver Fibrosis of Noninvasive Tools in Patients With Nonalcoholic Fatty Liver Disease. Am J Gastroenterol. 2019 Jun;114(6):916-928.

38) Harrison SA, 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 Nov;5(11):970-985.
39) Boursier J, et al. Screening for therapeutic trials and treatment indication in clinical practice: MACK-3, a new blood test for the diagnosis of fibrotic NASH. Aliment Pharmacol Ther. 2018 May;47(10):1387-1396.
40) Dennis A, et al. A composite biomarker using multiparametric magnetic resonance imaging and blood analytes accurately identifies patients with non-alcoholic steatohepatitis and significant fibrosis. Sci Rep. 2020 Sep 17;10(1):15308.
41) Jung J, et al. MRE combined with FIB-4 (MEFIB) index in detection of candidates for pharmacological treatment of NASH-related fibrosis. Gut. 2021 Oct;70(10):1946-1953.
42) Noureddin M, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. J Hepatol. 2022 Apr;76(4):781-787.
43) Newsome PN, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020 Apr;5(4):362-373.

Figure 1. ROC curves for prediction of F3-F4 fibrosis (A), NASH (B) and fibrotic NASH (C) using APRI, NFS, FIB-4, LSM and AGILE 3+ in NAFLD patients with diabetes.

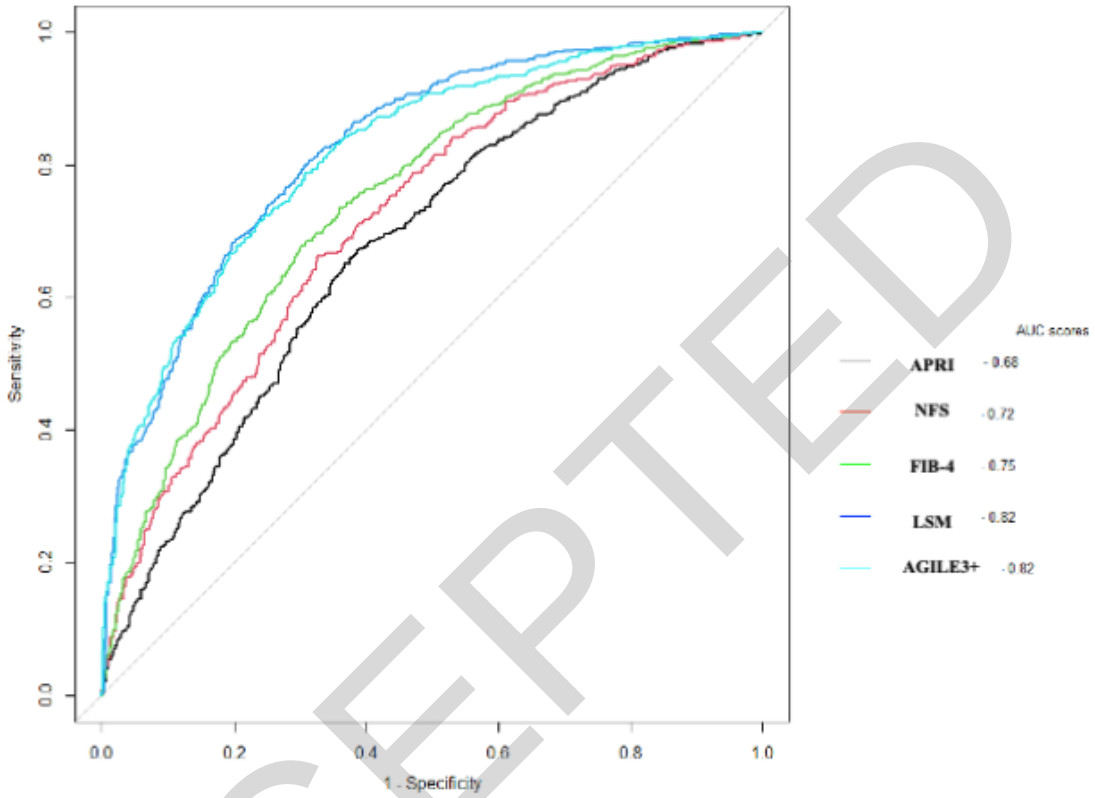

Figure 1A

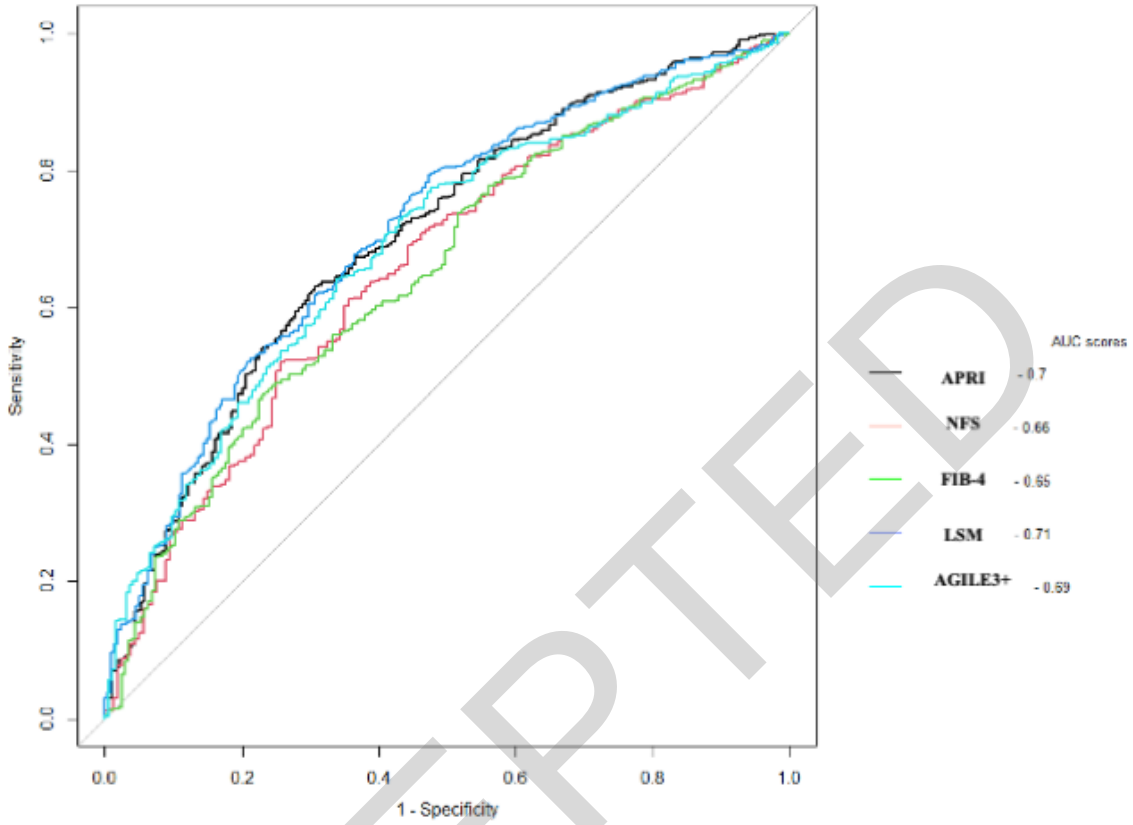

Figure 1B

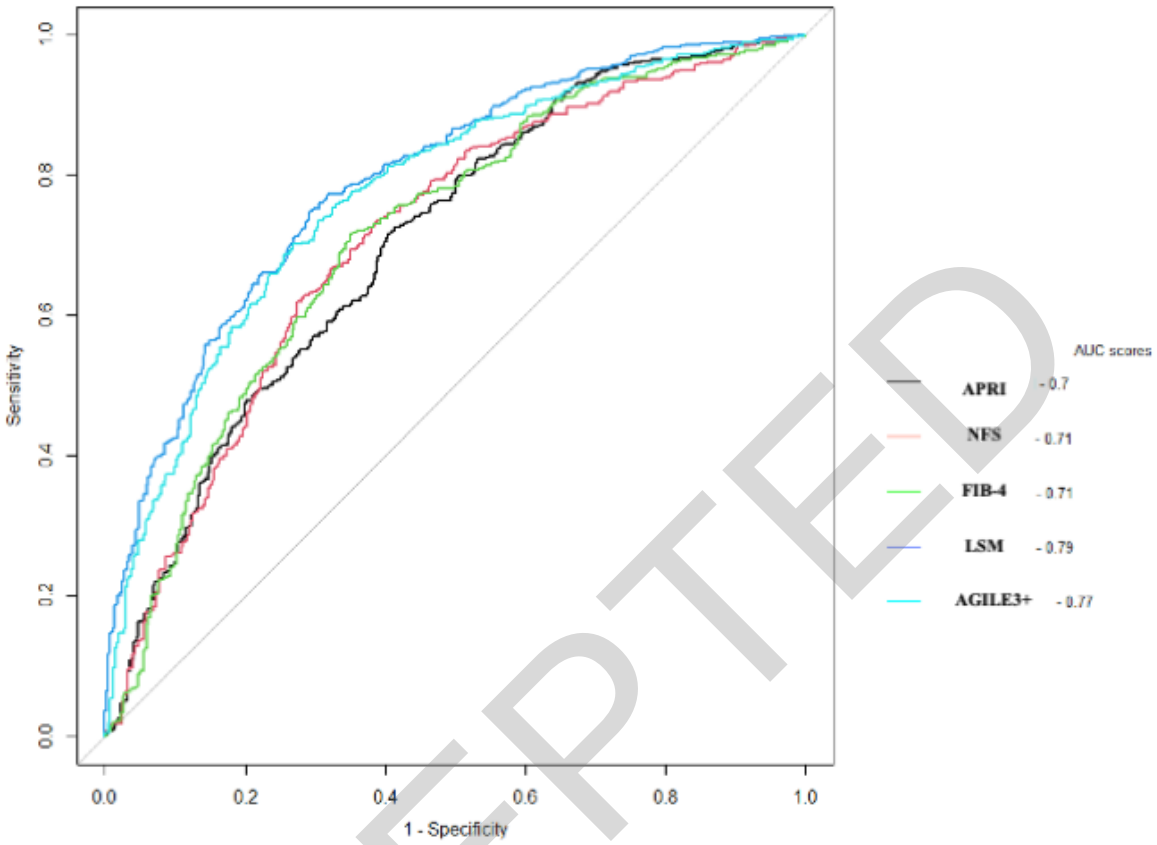

Figure 1C

Figure 2. Net benefits by decision curve analyses of NFS, FIB-4, LSM and AGILE 3+ for ruling out advanced liver fibrosis in NAFLD patients with diabetes.

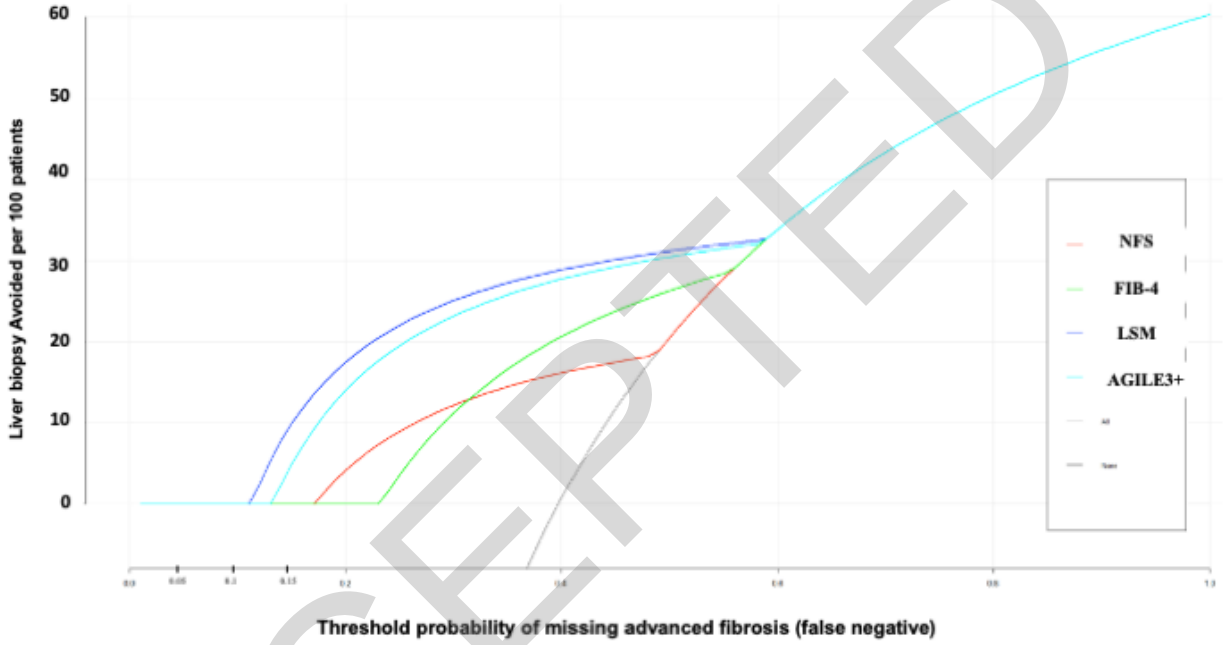

Figure 2

Figure 3. Net benefits by decision curve analyses of NFS, FIB-4, LSM and AGILE 3+ for ruling out advanced liver fibrosis (A), NASH (B) and fibrotic NASH (C) by using new identified rule-out cut-off in NAFLD patients with diabetes.

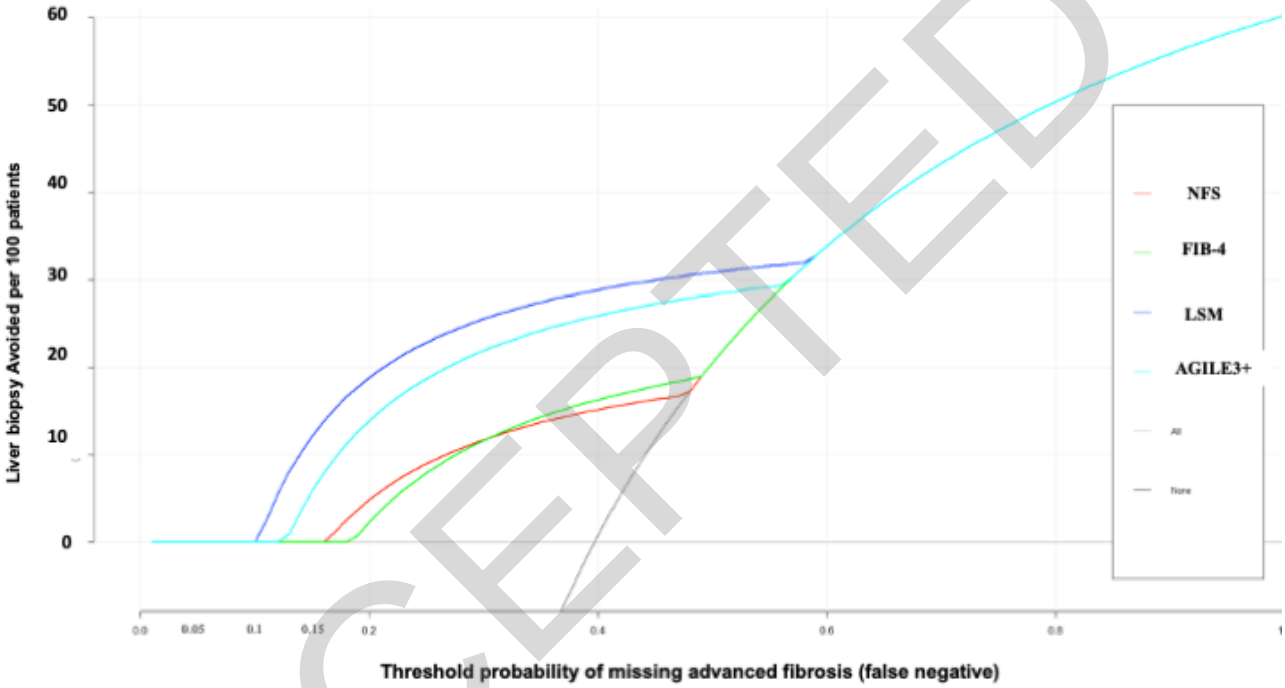

Figure 3A

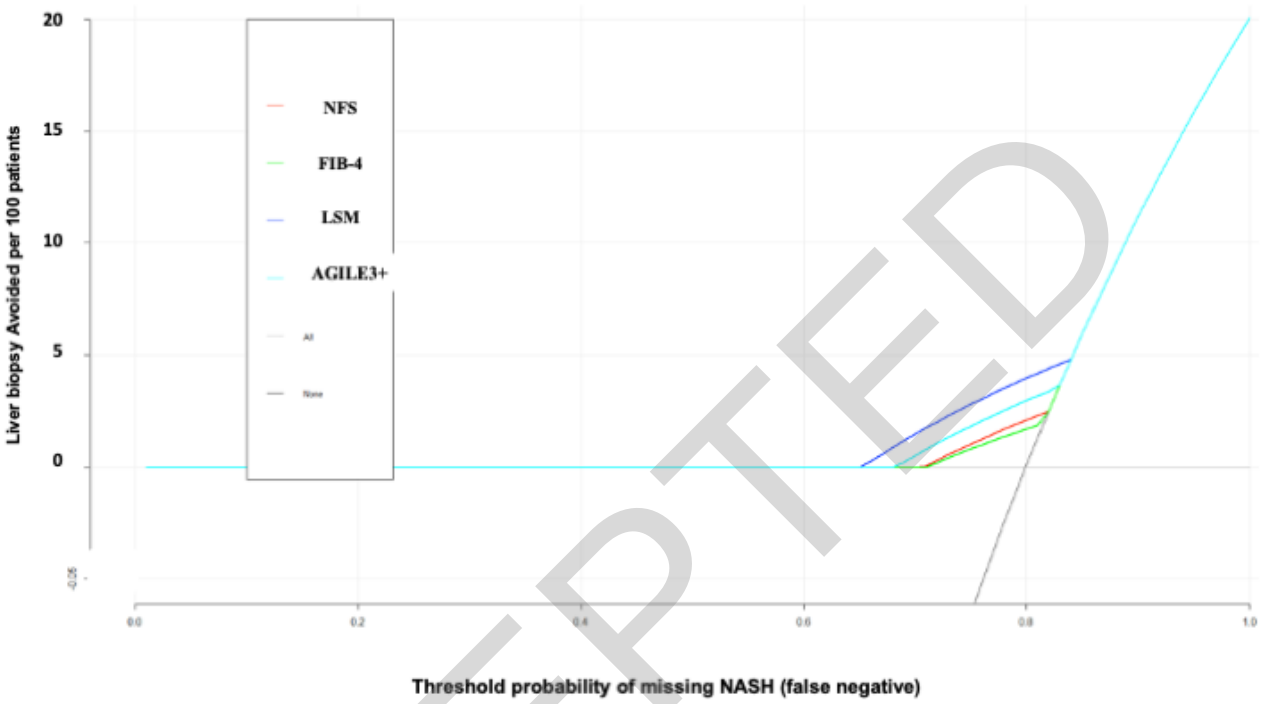

Figure 3B

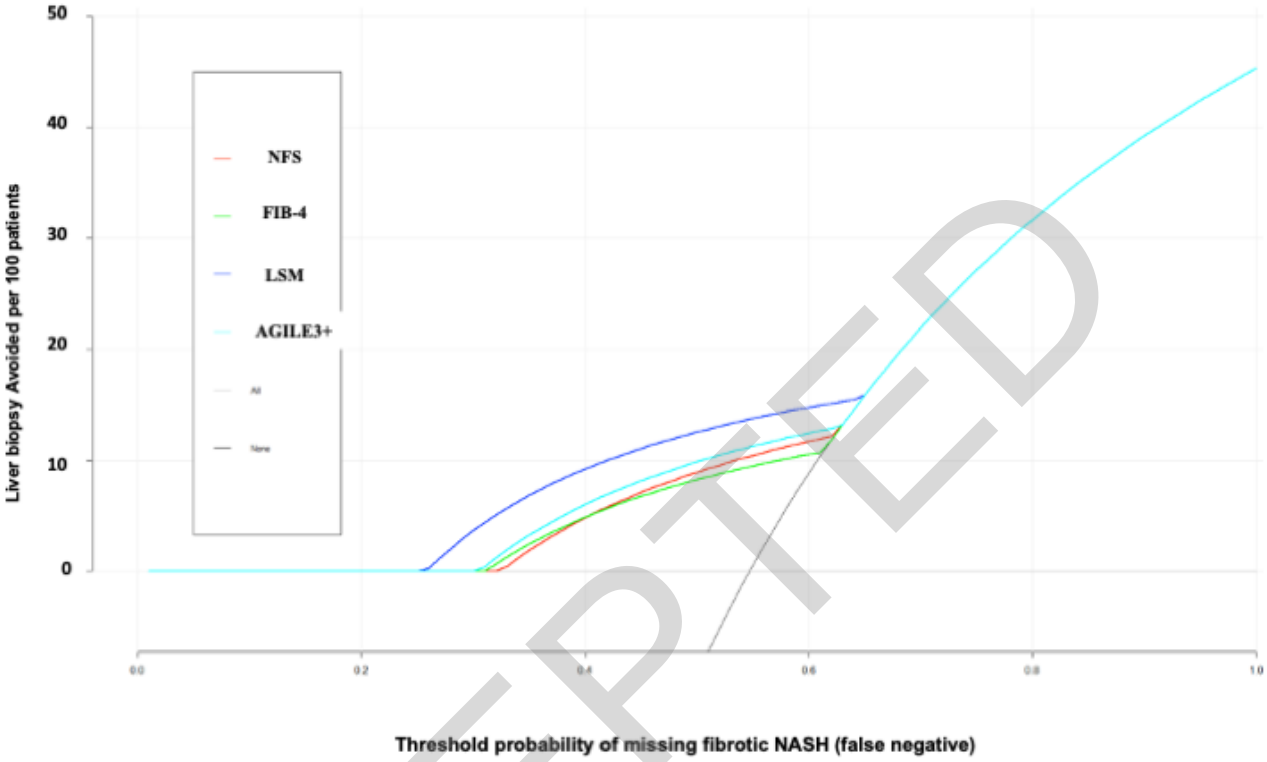

Figure 3C

Table 1. Comparison in the entire population of NAFLD patients with T2D of AUCs of LSM and different scores for diagnosing F3-F4 fibrosis. NASH and Fibrotic NASH.

	F3-F4 Fibrosis		**NASH**		**FIBROTIC NASH**	
	AUC	***p*-value**	**AUC**	***p*-value**	**AUC**	***p*-value**
APRI vs. NFS	0.68 - 0.72	0.163	0.70 - 0.66	0.844	0.70 - 0.71	0.879
APRI vs. FIB4	0.68 - 0.75	<0.001	0.70 - 0.65	0.054	0.70 - 0.71	0.879
APRI vs. LSM	0.68 - 0.82	<0.001	0.70 - 0.71	0.883	0.70 - 0.79	0.001
APRI vs. AGILE 3+	0.68 - 0.82	<0.001	0.70 - 0.69	0.883	0.70 - 0.77	0.019
NFS vs. FIB4	0.72 - 0.75	0.304	0.66 - 0.65	0.883	0.71 - 0.71	0.879
NFS vs. LSM	0.72 - 0.82	<0.001	0.66 - 0.71	0.507	0.71 - 0.79	0.004
NFS vs. AGILE 3+	0.72 - 0.82	<0.001	0.66 - 0.69	0.883	0.71 - 0.77	0.055
FIB4 vs. LSM	0.75 - 0.82	<0.001	0.65 - 0.71	0.317	0.71 - 0.79	0.004
FIB4 vs. AGILE 3+	0.75 - 0.82	<0.001	0.65 - 0.69	0.883	0.71 - 0.77	0.058
LSM vs. AGILE 3+	0.82 - 0.82	0.608	0.71 - 0.69	0.883	0.79 - 0.77	0.879

Abbreviations: APRI: AST-to-platelet ratio index; NFS: NAFLD fibrosis score; LSM: Liver stiffness measurement.

Table 2. Diagnostic accuracy in terms of sensitivity, specificity, PPV, NPV and 95% confidence intervals of LSM and different scores for diagnosing advanced fibrosis (F3-F4) by using literature suggested cut-offs in the entire population of NAFLD patients with T2D.

		STAGC				
		Sensitivity	***Specificity***	***PPV***	***NPV***	***Uncertainty Area***
NFS	***<-1.455 (n=276; 27.6%)***	0.88 (0.84 – 0.91)	0.38 (0.34 – 0.42)	0.50 (0.46 – 0.53)	0.82 (0.77 – 0.87)	55.8%
	>0.676 (n=166; 16.6%)	0.27 (0.3 – 0.32)	0.91 (0.88 – 0.93)	0.67 (0.60 – 0.75)	0.65 (0.61 - 0.68)	
FIB4	***<1.3 (n=773; 46%)***	0.73 (0.69 – 0.76)	0.62 (0.59 - 0.65)	0.62 (0.59 - 0.66)	0.72 (0.69 - 0.75)	40%
	>2.67 (n=236; 14%)	0.23 (0.20 - 0.26)	0.93 (0.92 - 0.95)	0.75 (0.69 - 0.80)	0.58 (0.56 - 0.61)	
LSM	***<8 KPa (n=606; 35.8%)***	0.88 (0.86 - 0.90)	0.56 (0.53 - 0.60)	0.64 (0.61 - 0.66)	0.84 (0.81 - 0.87)	12.9%
	>9.6 KPa (n=867; 51.2%)	0.77 (0.74 - 0.80)	0.71 (0.68 - 0.74)	0.70 (0.66 - 0.73)	0.78 (0.75 - 0.81)	
AGILE 3+	***<0.45 (n=583; 36.4%)***	0.87 (0.84 - 0.89)	0.56 (0.53 - 0.60)	0.63 (0.60 - 0.66)	0.83 (0.80 - 0.86)	20.1%
	>0.68 (n=698; 43.5%)	0.69 (0.65 - 0.72)	0.78 (0.76 - 0.81)	0.73 (0.70 - 0.77)	0.74 (0.71 - 0.77)	

Abbreviations: NFS: NAFLD fibrosis score; LSM: Liver stiffness measurement.

Table 3. Diagnostic accuracy in terms of sensitivity, specificity, PPV, NPV and 95% confidence intervals of combination of FIB-4 or NFS with LSM or AGILE 3+ for diagnosing advanced fibrosis (F3-F4) in the entire population of NAFLD patients with T2D by using literature suggested cut-offs.

		Se	***Sp***	***PPV***	***NPV***	***Uncertainty Area***
FIB4 → LSM	*Rule-out (n=974; **59.5%**)*	0.66 (0.62 - 0.69)	0.81 (0.78 - 0.84)	0.75 (0.71 - 0.78)	0.74 (0.71 - 0.76)	16%
	Rule-in (n=566; 34.5%)	0.59 (0.56 - 0.63)	**0.86** (0.84 - 0.89)	**0.79** (0.75 - 0.82)	0.71 (0.68 - 0.74)	
NFS → LSM	*Rule-out (n=481; 51.7%)*	**0.79** (0.75 - 0.83)	0.72 (0.69 - 0.76)	0.66 (0.61 - 0.70)	**0.84** (0.80 - 0.87)	**8.2%**
	*Rule-in (n=374**; 40.2%**)*	0.70 (0.65 - 0.75)	0.80 (0.76 - 0.83)	0.70 (0.65 - 0.75)	0.80 (0.76 - 0.83)	
FIB4 → AGILE 3+	*Rule-out (n=874; 53.3%)*	0.68 (0.65 - 0.72)	0.72 (0.69 - 0.75)	0.67 (0.64 - 0.71)	0.73 (0.70 - 0.76)	10%
	Rule-in (n=601; 36.7%)	0.61 (0.57 - 0.64)	**0.84** (0.81 - 0.86)	**0.76** (0.72 - 0.79)	0.71 (0.69 - 0.74)	
NFS → AGILE 3+	*Rule-out (n=434; 46.6%)*	**0.80** (0.75 - 0.84)	0.64 (0.60 - 0.68)	0.60 (0.56 - 0.64)	**0.82** (0.79 - 0.86)	15.1%
	Rule-in (n=357; 38.3%)	0.67 (0.62 - 0.71)	0.81 (0.77 - 0.84)	0.70 (0.65 - 0.75)	0.78 (0.75 - 0.82)	

Abbreviations: NFS: NAFLD fibrosis score; LSM: Liver stiffness measurement.
Used cut-offs: FIB-4 rule-out <1.30, rule-in >2.67; NFS rule-out <-1.455, rule-in >0.675; LSM rule-out <7.9 KPa, rule-in >9.6 KPa; AGILE 3+ rule-out <0.45, rule-in ≥0.68.

Table 4. Diagnostic accuracy for advanced fibrosis and fibrotic NASH in terms of sensitivity. specificity. PPV. NPV and 95% confidence intervals of NFS, FIB-4, LSM and AGILE 3+ in the training and validation sets of NAFLD patients with T2D according to the best identified cut-offs

Advanced Fibrosis – Training Set						
		Se	*Sp*	*PPV*	*NPV*	*Uncertainty Area*
NFS	*Rule-out <-1.539 (N=119; 21.5%)*	0,90 (0,86 - 0,94)	0,31 (0,26 - 0,36)	0,51 (0,46 - 0,55)	0,80 (0,71 - 0,87)	61.4%
	Rule-in >0.766 (N=95; 17.1%)	0,27 (0,21 - 0,33)	0,90 (0,86 - 0,93)	0,68 (0,58 - 0,78)	0,61 (0,56 - 0,65)	
FIB4	*Rule-out <0.973 (N=294; 23.9%)*	0,90 (0,87 - 0,92)	0,38 (0,34 - 0,42)	0,59 (0,56 - 0,62)	0,79 (0,74 - 0,83)	54.7%
	Rule-in >2.310 (N=264; 21.4%)	0,33 (0,29 - 0,37)	0,90 (0,87 - 0,92)	0,77 (0,71 - 0,82)	0,57 (0,54 - 0,60)	
LSM	*Rule-out <7.9 KPa (N=390; 31.0%)*	0,90 (0,87 - 0,92)	0,52 (0,48 - 0,56)	0,65 (0,62 - 0,68)	0,84 (0,80 - 0,88)	42.4%
	Rule-in >14.6 KPa (N=334; 26.6%)	0,44 (0,40 - 0,48)	0,90 (0,88 - 0,93)	0,82 (0,77 - 0,86)	0,62 (0,59 - 0,65)	
AGILE 3+	*Rule-out <0.426 (N=338; 28.8%)*	0,90 (0,87 - 0,92)	0,48 (0,44 - 0,52)	0,63 (0,60 - 0,67)	0,83 (0,78 - 0,86)	43.4%
	Rule-in >0.848 (N=327; 27.8%)	0,46 (0,42 - 0,50)	0,90 (0,87 - 0,92)	0,82 (0,78 - 0,86)	0,62 (0,59 - 0,66)	

Advanced Fibrosis – Validation Set						
NFS	*Rule-out <-1.539 (N=135; 30.2%)*	0,91 (0,85 - 0,95)	0,20 (0,15 - 0,25)	0,40 (0,35 - 0,45)	0,79 (0,68 - 0,88)	57.9%
	Rule-in >0.766 (N=53; 11.9%)	0,25 (0,18 - 0,32)	0,92 (0,89 - 0,95)	0,65 (0,52 - 0,77)	0,68 (0,63 - 0,73)	
FIB4	*Rule-out <0.973 (N=163; 36.3%)*	0,88 (0,82 - 0,92)	0,41 (0,35 - 0,47)	0,46 (0,41 - 0,52)	0,85 (0,78 - 0,91)	45.2%

	Rule-in >2.310 (N=83; 18.5%)	0,24 (0,17 - 0,31)	0,95 (0,92 - 0,97)	0,74 (0,60 - 0,85)	0,68 (0,63 - 0,73)	
LSM	*Rule-out <7.9 KPa (N=190; 43.6%)*	0,81 (0,74 - 0,86)	0,46 (0,40 - 0,52)	0,47 (0,41 - 0,52)	0,80 (0,73 - 0,86)	37.4%
	Rule-in >14.6 KPa (N=83; 19.0%)	0,35 (0,28 - 0,43)	0,91 (0,87 - 0,94)	0,70 (0,59 - 0,79)	0,71 (0,66 - 0,75)	
AGILE3	*Rule-out <0.426 (N=191; 44.6%)*	0,86 (0,79 - 0,91)	0,61 (0,54 - 0,66)	0,56 (0,49 - 0,62)	0,88 (0,82 - 0,92)	34.8%
	Rule-in >0.848 (N=88; 20.6%)	0,42 (0,35 - 0,51)	0,95 (0,91 - 0,97)	0,82 (0,72 - 0,90)	0,74 (0,69 - 0,78)	

Fibrotic NASH - Training Set						
		Se	*Sp*	*PPV*	*NPV*	*Uncertainty Area*
NFS	*Rule-out <-1.539 (N=73; 18.0%)*	0,90 (0,85 - 0,93)	0,30 (0,23 - 0,37)	0,65 (0,59 - 0,70)	0,67 (0,55 - 0,78)	62.5%
	Rule-in >0.674 (N=79; 19.5%)	0,26 (0,21 - 0,32)	0,90 (0,85 - 0,94)	0,80 (0,69 - 0,88)	0,46 (0,41 - 0,52)	
FIB4	*Rule-out <0.845 (N=106; 18.3%)*	0,90 (0,86 - 0,93)	0,31 (0,25 - 0,38)	0,67 (0,63 - 0,71)	0,67 (0,57 - 0,76)	63.3%
	Rule-in >2.306 (N=107; 18.4%)	0,24 (0,20 - 0,29)	0,90 (0,86 - 0,94)	0,79 (0,71 - 0,87)	0,43 (0,39 - 0,48)	
LSM	*Rule-out <6.6 KPa (N=117; 19.3%)*	0,90 (0,86 - 0,93)	0,33 (0,27 - 0,39)	0,67 (0,62 - 0,71)	0,69 (0,60 - 0,77)	52.1%
	Rule-in >14.0KPa (N=173; 28.5%)	0,41 (0,36 - 0,47)	0,90 (0,86 - 0,94)	0,87 (0,81 - 0,91)	0,51 (0,46 - 0,56)	
AGILE3	*Rule-out <0.288 (N=101; 18.5%)*	0,90 (0,86 - 0,93)	0,32 (0,26 - 0,39)	0,68 (0,63 - 0,72)	0,67 (0,57 - 0,76)	52.7%
	Rule-in >0.819	0,41 (0,36 -	0,90 (0,86 - 0,94)	0,87 (0,81 -	0,49 (0,44 -	

	(N=157; 28.8%)	0,47)		0,92)	0,54)	

Fibrotic NASH - Validation Set						
		Se	*Sp*	*PPV*	*NPV*	*Uncertainty Area*
NFS	*Rule-out <-1.539 (N=64; 16.2%)*	0,80 (0,74 - 0,86)	0,42 (0,35 - 0,49)	0,58 (0,52 - 0,64)	0,68 (0,58 - 0,76)	69.6%
	Rule-in >0.674 (N=56; 14.2%)	0,13 (0,09 - 0,19)	0,89 (0,84 - 0,93)	0,55 (0,40 - 0,70)	0,50 (0,45 - 0,56)	
FIB4	*Rule-out <0.845 (N=109; 27.6%)*	0,79 (0,73 - 0,84)	0,34 (0,28 - 0,41)	0,55 (0,49 - 0,61)	0,61 (0,52 - 0,71)	55.9%
	Rule-in >2.306 (N=65; 16.5%)	0,22 (0,16 - 0,28)	0,89 (0,84 - 0,93)	0,66 (0,53 - 0,77)	0,53 (0,47 - 0,58)	
LSM	*Rule-out <6.6 KPa (N=125; 32.6%)*	0,83 (0,77 - 0,88)	0,48 (0,41 - 0,55)	0,61 (0,55 - 0,67)	0,74 (0,65 - 0,81)	45.3%
	Rule-in >14.0 KPa (N=85; 22.1%)	0,35 (0,29 - 0,43)	0,91 (0,86 - 0,95)	0,80 (0,70 - 0,88)	0,59 (0,53 - 0,64)	
AGILE3	*Rule-out <0.288 (N=124; 32.9%)*	0,78 (0,71 - 0,83)	0,43 (0,36 - 0,51)	0,58 (0,51 - 0,64)	0,66 (0,57 - 0,74)	46.2%
	Rule-in >0.819 (N=79; 20.9%)	0,32 (0,25 - 0,39)	0,90 (0,85 - 0,94)	0,76 (0,65 - 0,85)	0,57 (0,51 - 0,63)	

Abbreviations: NFS: NAFLD fibrosis score; LSM: Liver stiffness measurement; NASH: nonalcoholic steatohepatitis.

Table 5. Comparison. in the entire cohort of diabetic NAFLD patients with T2D of AUCs of LSM and different scores for diagnosing F3-F4 fibrosis, F2-F4 fibrosis, NASH and Fibrotic NASH according to obesit

	Advanced Fibrosis		**NASH**		**Fibrotic NASH**	
	AUC	***p*-value**	**AUC**	***p*-value**	**AUC**	***p*-value**
	BMI <30 Kg/m2 vs BMI≥30 Kg/m2					
APRI	0.75 - 0.66	0.034	0.69 - 0.50	0.004	0.74 - 0.71	0.531
NFS	0.80 - 0.70	0.030	0.71 - 0.59	0.061	0.77 - 0.73	0.531
FIB4	0.88 - 0.79	0.016	0.75 - 0.63	0.059	0.84 - 0.76	0.145
LSM	0.87 - 0.79	0.021	0.72 - 0.59	0.038	0.82 - 0.77	0.455
AGILE3	0.75 - 0.66	0.034	0.69 - 0.50	0.004	0.74 - 0.71	0.531
	Normal ALT vs Abnormal ALT					
APRI	0.74 - 0.59	0.005	0.72 - 0.70	0.727	0.73 - 0.63	0.117
NFS	0.74 - 0.66	0.091	0.68 - 0.58	0.231	0.73 - 0.67	0.256
FIB4	0.80 - 0.68	0.011	0.72 - 0.64	0.308	0.78 - 0.72	0.256
LSM	0.87 - 0.82	0.091	0.80 - 0.70	0.231	0.86 - 0.76	0.028
AGILE3	0.88 - 0.79	0.015	0.78 - 0.65	0.153	0.84 - 0.77	0.139
	Age < 35 years vs Age 35-65 years					
APRI	0.47 - 0.72	0.125	0.45 - 0.71	0.208	0.67 - 0.73	0.981
NFS	0.54 - 0.74	0.166	0.83 - 0.66	0.570	0.51 - 0.73	0.245
FIB4	0.68 - 0.77	0.526	0.70 - 0.69	0.946	0.54 - 0.75	0.392
LSM	0.44 - 0.85	0.058	0.72 - 0.74	0.946	0.81 - 0.82	0.981
AGILE3	0.59 - 0.85	0.125	0.63 - 0.71	0.946	0.81 - 0.81	0.981
	Age < 35 years vs Age >65 years					
APRI	0.47 - 0.65	0.523	0.45 - 0.85	0.019	0.67 - 0.77	0.927
NFS	0.54 - 0.65	0.720	0.83 - 0.81	0.899	0.51 - 0.75	0.215
FIB4	0.68 - 0.69	0.948	0.70 - 0.83	0.893	0.54 - 0.77	0.338
LSM	0.44 - 0.84	0.073	0.72 - 0.85	0.899	0.81 - 0.85	0.954
AGILE3	0.59 -	0.346	0.63 -	0.624	0.81 -	0.954

	0.82		0.84		0.82	
	Age 35-65 years vs Age >65 years					
APRI	0.72 - 0.65	0.525	0.71 - 0.85	0.021	0.73 - 0.77	0.833
NFS	0.74 - 0.65	0.350	0.66 - 0.81	0.021	0.73 - 0.75	0.833
FIB4	0.77 - 0.69	0.350	0.69 - 0.83	0.021	0.75 - 0.77	0.833
LSM	0.85 - 0.84	0.798	0.74 - 0.85	0.021	0.82 - 0.85	0.833
AGILE3	0.85 - 0.82	0.798	0.71 - 0.84	0.021	0.81 - 0.82	0.833

Liver stiffness measurement.

y, age and ALT levels.

Abbreviations: APRI: AST-to-platelet ratio indes; NFS: NAFLD fibrosis score; LSM: