Comparison of diagnostic accuracy and utility of non-invasive tests for clinically significant liver disease in a general population with metabolic dysfunction

Kleef, L.A. van; Pustjens, J.; Schattenberg, J.M.; Holleboom, A.G.; Cabezas, M.C.; Tushuizen, M.E.; ... ; Brouwer, W.P.

Received: 17 December 2024 | Accepted: 29 March 2025

DOI: 10.1097/HEP.0000000000001356

ORIGINAL ARTICLE

Comparison of diagnostic accuracy and utility of non-invasive tests for clinically significant liver disease in a general population with metabolic dysfunction

Laurens A. van Kleef[1] | Jesse Pustjens[1] | Jörn M. Schattenberg[2] | Adriaan G. Holleboom[3] | Manuel Castro Cabezas[4,5] | Maarten E. Tushuizen[6] | Robert J. de Knegt[1] | M. Arfan Ikram[7] | Harry L.A. Janssen[1,8] | Sven M. Francque[9,10] | Willem Pieter Brouwer[1]

[1]Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[2]Department of Internal Medicine II, Saarland University Medical Center, Homburg, Germany

[3]Department of Vascular Internal Medicine, Amsterdam Medical Center, Amsterdam, The Netherlands

[4]Department of Internal Medicine, Sint Franciscus Gasthuis, Rotterdam, The Netherlands

[5]Department of Internal Medicine and Endocrinology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[6]Department of Gastroenterology and Hepatology, Leids Universitair Medisch Centrum, Leiden, The Netherlands

[7]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[8]Toronto Centre for Liver Disease, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

[9]InflaMed Centre of Excellence, Laboratory for Experimental Medicine and Paediatrics, Translational Sciences in Inflammation and Immunology, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

[10]Department of Gastroenterology and Hepatology, Antwerp University Hospital, Antwerp, Belgium

Correspondence
Laurens A. van Kleef, Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Centre, Postbus 2040; Dr. Molewaterplein 40, Rotterdam, The Netherlands.
Email: l.vankleef@erasmusmc.nl

Abstract

Background and Aims: Screening for liver disease in the general population requires accurate non-invasive tests (NITs). A head-to-head comparison of NITs for early detection of clinically relevant liver disease among the target population for screening is lacking.

Aproach and Results: Among the meta-cohort (Rotterdam Study and National Health and Nutrition Examination Survey) with metabolic dysfunction aged 18–80 years, 10 NITs were investigated. The diagnostic accuracy for clinically relevant conditions [increased liver stiffness measurement (LSM), at-risk metabolic dysfunction–associated steatohepatitis, advanced fibrosis, or cirrhosis) was assessed. Subgroup analysis included stratification by age group and diabetes/obesity status.

Abbreviations: AASLD, American Association for the Study of Liver Diseases; AGA, American Gastroenterological Association; APRI, AST to platelet ratio index; BMI, body mass index; CAP, controlled attenuation parameter; CORE, cirrhosis outcome risk estimator; DM, diabetes mellitus; DOR, diagnostic odds ratio; EASL, European Association for the Study of the Liver; ELF, enhanced liver fibrosis; FAST, FibroScan-AST score; FDA, Food and Drug Administration; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HbA1c, hemoglobin A1c; HFS, Hepamet fibrosis score; LRS, LiverRisk score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis 5; MASLD, metabolic dysfunction–associated steatotic liver disease; MASH, metabolic dysfunction–associated steatohepatitis; NFS, NAFLD fibrosis score; NHANES, National Health and Nutrition Examination Survey; NITs, non-invasive tests; NPV, negative predictive value; PPV, positive predictive value; SAFE, steatosis-associated fibrosis estimator; SLD, steatotic liver disease; WC, waist circumference.

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.

We analysed 11,404 participants. Metabolic dysfunction–associated fibrosis 5 (MAF-5) obtained the highest AUC for increased LSM (≥8 kPa: 0.80; ≥12 kPa: 0.87) and advanced fibrosis (AUC: 0.90). Fibrotic NASH index and MAF-5 performed best for detecting metabolic dysfunction–associated steatohepatitis (AUC: 0.93 and AUC: 0.92, p = ns) and SAFE for cirrhosis (AUC: 0.92). To obtain 80% sensitivity for LSM ≥8 kPa, the corresponding MAF-5 cut-off resulted in fewer referrals (42%) compared to fibrosis-4 index (77%) and higher specificity (62% vs. 24%); MAF-5 was also superior for detection of LSM ≥12 kPa and advanced fibrosis. Age-dependent scores yielded lower sensitivity among younger individuals, for example, by referring 20% of the population with the highest NIT scores, the fibrosis-4 index, steatosis-associated fibrosis estimator, NAFLD fibrosis score, FORNS, and Hepamet fibrosis score yielded <10% sensitivity for LSM ≥8 kPa among individuals aged 18–35 years, while fibrotic NASH index and MAF-5 obtained 40% and 71%.

Conclusions: Of the 10 investigated NITs, MAF-5 discriminated best between all conditions except cirrhosis, for which the steatosis-associated fibrosis estimator yielded the highest accuracy. The performance of the fibrosis-4 index was poor, implying that referral pathways for significant liver disease in low-prevalence populations can be improved when more accurate NITs such as MAF-5 are employed.

Keywords: epidemiology, external validation, fibrosis risk prediction, general population, MASLD, screening, SLD

INTRODUCTION

The American Association for the Study of Liver Diseases (AASLD), the European Association for the Study of the Liver (EASL), and the American Gastroenterological Association (AGA) recommend referral pathways that aim at detecting F2–F4 fibrosis among individuals who are suspected to live with steatotic liver disease (SLD), and/or have cardiometabolic risk factors.[1–3] Depending on the guideline, this varies between the presence of diabetes mellitus (DM) or obesity (with or without other traits of metabolic syndrome) to less severe metabolic dysfunction phenotypes such as isolated dyslipidaemia or hypertension. In the United States, up to 80% of the general population aged 18–80 years meet the most lenient criteria and are therefore potentially eligible for liver disease screening.[4,5] This has led to concerns about the feasibility of implementing referral pathways.[5–7] The recent introduction of resmetirom as an FDA (Food and Drug Administration)-approved pharmacotherapy for fibrotic metabolic dysfunction–associated steatohepatitis (MASH) has made this implementation even more pertinent.[8] Indeed, the recent EASL guideline lists research into the performance of non-invasive tests (NITs) in lower-prevalence settings as a key agenda topic.[3]

The fibrosis-4 index (FIB-4) is the initial screening tool recommended by these guidelines. Further assessment is required with liver stiffness measurement (LSM) or enhanced liver fibrosis (ELF) test in case the FIB-4 exceeds the threshold of 1.3 or (depending on the guideline) 2.0 among individuals aged ≥65 years. Since the introduction of FIB-4, several other promising tests have been published that may perform better in a low-prevalence setting.[9–18] These tests may be more suited for the intended referral pathways for fibrotic metabolic dysfunction–associated steatotic liver disease (MASLD), as FIB-4–based pathways are either associated with high referral rates or underestimation of fibrosis prevalence.[5,19] Although some studies have singularly compared FIB-4 to the NAFLD fibrosis score (NFS), LiverRisk score (LRS), steatosis-associated fibrosis estimator (SAFE), or metabolic dysfunction–associated fibrosis-5 (MAF-5) in a general population setting, a more comprehensive direct comparison of these and other NITs in a single large meta-study among the target population for screening is not yet available.[9,12,20,21]

Here, we therefore directly compare 10 NITs for ruling in and/or ruling out the risk of an increased LSM, at-risk MASH, advanced fibrosis, and cirrhosis in 2 general population cohorts with metabolic dysfunction.

METHODS

This study comprises a pooled cross-sectional analysis of 2 independent general population cohorts in the Netherlands and the United States.

The first cohort is the Rotterdam Study, which was designed to investigate health and healthy ageing in a general population setting in the Netherlands. For this specific study, we used data that was collected between 2009–2014 and 2016– 2020. Briefly, data were collected on individuals' health, and the examination included, among others, laboratory testing, abdominal ultrasound, and transient elastography. More details on the rationale of the Rotterdam Study, its design, and recent publications are described elsewhere.[22]

The second cohort is the National Health and Nutrition Examination Survey (NHANES) 2017–2020 cohort. NHANES aimed to investigate health and nutritional status in the United States through interviews, physical examinations, and extensive clinical measurements and tests performed by trained research assistants. These tests include, among others, vibration-controlled transient elastography to assess LSM. Details on procedures and the rationale of the NHANES are published elsewhere.[23,24] Data are publicly available from the NHANES database (https://www.cdc.gov/nchs/nhanes/index.htm).

For this study, we aimed to reflect the target population of the current referral pathways. We therefore included only participants with at least one of the metabolic dysfunction criteria (eg, overweight, hypertension, dyslipidaemia, abdominal obesity, (pre)diabetes) as outlined in the SLD consensus statement.[25] We excluded participants who had incomplete data on AST, ALT, and platelets, as these were the most common laboratory parameters included in NITs. Additionally, we excluded participants with heart failure (based on interview data), which is known to falsely elevate LSM in a general population setting.[26]

Endpoints

The endpoints for this study included the presence of clinically relevant (significant) liver fibrosis (significant refers to fibrosis stage F2 or above), at-risk MASH (ie, MASH with a NAFLD activity score of ≥ 4 and fibrosis stage $\geq$F2), advanced fibrosis (ie, $\geq$F3) and cirrhosis.[27,28] As this is a population-based study, these endpoints are based on liver stiffness measurements and liver stiffness-based scores, as liver biopsies in this setting are neither available nor feasible.

Elevated LSM (with 2 thresholds: ≥ 8.0 kPa or ≥ 12.0 kPa) was used as a surrogate marker of clinically relevant (significant) liver fibrosis.[29] At-risk MASH was based on FibroScan AST (FAST) score ≥ 0.35.[30] Finally, advanced fibrosis and cirrhosis were based on Agile 3+ ≥ 0.679 and Agile 4 ≥ 0.565.[31] Formulas are available in the Supplemental Methods, http://links.lww.com/HEP/J791.

LSM was obtained through transient elastography using the FibroScan device equipped with both the M and XL probes (FibroScans). Participants were instructed to fast for at least 3 hours before the assessment. Measurements were considered reliable if at least 10 measurements were obtained with an IQR $<30\%$ in case LSM ≥ 7.0 kPa. Additional details about the procedures are described previously.[22,32,33] Controlled attenuation parameter (CAP) was simultaneously assessed in the Rotterdam Study since 2016 and in the NHANES.

Non-invasive tests

The NITs that were selected to investigate were AST to platelet ratio (APRI), Cirrhosis Outcome Risk Estimator (CORE), FIB-4, fibrotic NASH index (FNI), FORNS, Hepamet fibrosis score (HFS), LRS, MAF-5, NFS, and SAFE.[9–18] Due to a lack of data on albumin, globulin, and HbA1c, the HFS, FNI, NFS, and SAFE scores could not be calculated in the Rotterdam Study. All the aforementioned NITs could be calculated in the NHANES. We calculated the investigated NITs according to their formulas provided in the Supplemental Methods, http://links.lww.com/HEP/J791, or used online calculators when the exact formula has not been published.

Statistical analysis

Data from the Rotterdam Study and the NHANES were pooled, and AUC analysis was performed on detecting increased LSM (≥ 8 and ≥ 12 kPa), at-risk MASH, advanced fibrosis, and cirrhosis for the 10 investigated NITs. NITs were compared with the best-performing test for each outcome to assess whether they were equal or inferior in performance by using the DeLong test.[34] Here, we particularly focused on the AUC of a test as an overall performance estimate, as the AUC of a test is independent of specific cut-offs. Of note, participant selection corresponded with the criteria for screening for liver health according to the AASLD or AGA guidelines: individuals with metabolic dysfunction. In a sensitivity analysis, we focused on the following subgroups: individuals aged 18–35 years, 35–65 years, 65–80 years, participants with obesity, and participants with DM. The final 2 groups,

individuals with diabetes or obesity, reflected the target population for screening by the 2024 EASL guideline on MASLD.[3]

Next, we investigated the proportion of the population that required referral for further workup at the cut-off of the NIT to obtain at least 80% sensitivity for the detection of increased LSM (≥8 and ≥12 kPa) at-risk MASH, advanced fibrosis, and cirrhosis. Additionally, we reported the test characteristics [sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), diagnostic odds ratio (DOR)] at the cut-off that corresponded with 80% sensitivity to allow for a fair comparison (as not all investigated tests were designed to detect the investigated outcomes). For liver disease screening purposes, there is currently no formal acceptable sensitivity threshold published. Therefore, we aimed for sensitivity thresholds that are similar to those used in current colon cancer screening programs.[35] Higher sensitivity thresholds would lead to undesirable high referral rates with corresponding lower PPV. Of note, the PPV and NPV were calculated based on the background-prevalence independent sensitivity and specificity, together with the prevalence of the entire cohort, rather than the subset in which the specific test was available, allowing for a fair comparison.[36]

In an additional sensitivity analysis, the yield of the individual NIT was evaluated in a scenario where 20% of the at-risk population would be referred. The 20% was chosen taking into account the background prevalence of increased LSM, test characteristics in a general population setting, and the workload for already strained healthcare systems. The corresponding sensitivity, specificity, PPV, NPV, and DOR were calculated.

Ethics

Participants of the Rotterdam Study and NHANES 2017–2020 provided informed consent. This study was conducted according to the principles outlined in the Declaration of Helsinki and Istanbul.

RESULTS

We included 13,454 participants from the Rotterdam Study and NHANES 2017–2020, who had data available on LSM. We excluded 642 participants for lacking basic laboratory data, 1.168 participants for not being the target population of screening due to the absence of metabolic dysfunction, and 240 for having potentially unreliable LSM due to prevalent heart failure, leaving 11,404 participants for analysis (Supplemental Figure S1, http://links.lww.com/HEP/J792). In general, 48.8% were male, 12.8% (n = 1464) were aged 18–35 years, 52.6% (n = 5999) were 35–65 years, and 34.6% (3941) were 65–80 years. LSM ≥8 kPa was present in 7.7%, LSM ≥12 kPa in 2.2%, at-risk MASH in 6.8%, advanced fibrosis in 4.3%, and cirrhosis in 0.6%. Baseline characteristics per cohort are presented in Table 1 and events per subgroup in Supplemental Table S1, http://links.lww.com/HEP/J793.

NIT performance in the general population

The discriminative value of NITs for increased LSM, at-risk MASH, advanced fibrosis, and cirrhosis is presented in Figure 1 and Supplemental Table S2, http://links.lww.com/HEP/J793. In general, the investigated tests were more accurate when used to assess the risk of more severe disease. The MAF-5 yielded the highest AUC to detect LSM ≥8 kPa (AUC: 0.80), LSM ≥12 kPa (AUC: 0.87), and advanced fibrosis (AUC: 0.90), followed by SAFE (AUC: 0.74, 0.79, and 0.89, respectively) and LRS (AUC: 0.721, 0.771, and 0.873, respectively). Of note, the SAFE did not differ significantly (using the DeLong test) from MAF-5 in the detection of advanced fibrosis but had significantly lower AUC for increased liver stiffness. The FNI (AUC: 0.93) and MAF-5 (AUC: 0.92) did not differ significantly in the detection of at-risk MASH, while the other NITs yielded inferior performances. Finally, cirrhosis was best detected by the SAFE (AUC: 0.92) and NFS (AUC: 0.91). Important to note is that the FIB-4, the currently recommended screening test, could only accurately assess the presence of cirrhosis (AUC: 0.87, present in 0.6% of the general population) and performed substantially poorer for earlier stages (AUC: 0.58–0.72). Supplemental Table S3, http://links.lww.com/HEP/J793, shows the AUC levels per subgroup for NHANES and the Rotterdam Study separately. In the Rotterdam Study, we observed that MAF-5, together with LRS, was among the best-performing tests. The results in the NHANES cohort did not change substantially.

Test characteristics to obtain at least 80% sensitivity and corresponding referral rates

To obtain 80% sensitivity, the corresponding referral rate to detect LSM ≥8 kPa ranged between 42% (for the MAF-5) and 77% (for the FIB-4) and decreased with more severe liver disease (Figure 2). The MAF-5 also required the lowest number of referrals to obtain 80% sensitivity for LSM ≥12 kPa (23%) and advanced fibrosis (18%), whereas the FNI required the lowest number of referrals to detect 80% of at-risk MASH (15%) and SAFE for cirrhosis (8%). The corresponding cut-offs, specificity, PPV, NPV, and DOR are provided in Supplemental Table S2, http://links.lww.com/HEP/J793. It was observed that for a sensitivity of 80%, the MAF-5 had the highest specificity and DOR compared to all other NITs, ranging from 62% to 87% for the

TABLE 1 Participants' characteristics

Variable	All, n = 11.404	NHANES, n = 6.375	Rotterdam, n = 5.029
Demographics			
Age (y)	59 [46, 68]	52 [36, 64]	65 [56, 70]
Male	5441 (47.7)	3108 (48.8)	2333 (46.4)
Physical examination			
Waist circumference (cm)			
Male	101 (13)	103 (15)	99 (10)
Female	95 (15)	100 (16)	89 (11)
BMI (kg/m^2)	29.1 (6.0)	30.4 (6.9)	27.5 (3.9)
Comorbidity			
Obesity	3956 (34.8)	2821 (44.5)	1135 (22.6)
Diabetes mellitus	1846 (16.2)	1239 (19.4)	607 (12.1)
Biochemistry			
AST (U/L)	21 [18, 26]	19 [16, 24]	24 [20, 28]
ALT (U/L)	19 [14, 27]	18 [13, 26]	20 [16, 27]
GGT	32 (40)	32 (41)	32 (39)
Platelets	255 (65)	248 (66)	263 (62)
Non-invasive tests			
APRI	0.22 (0.20)	0.22 (0.22)	0.23 (0.16)
CORE	0.003 (0.017)	0.003 (0.017)	0.004 (0.016)
FIB-4	1.23 (0.81)	1.09 (0.89)	1.41 (0.64)
FNI	0.12 (0.16)	0.12 (0.16)	—
FORNS	−1.04 (1.63)	−1.42 (1.89)	−0.55 (1.05)
HFS	0.10 (0.14)	0.10 (0.14)	—
LRS	5.36 (1.44)	5.35 (1.52)	5.38 (1.33)
MAF-5	−0.45 (2.01)	−0.24 (2.23)	−0.72 (1.66)
NFS	−1.20 (1.60)	−1.20 (1.60)	—
SAFE	0.79 (98.67)	0.79 (98.67)	—
Outcomes			
LSM ≥8 kPa	881 (7.7)	659 (10.3)	222 (4.4)
LSM ≥12 kPa	252 (2.2)	219 (3.4)	33 (0.7)
At-risk MASH[a]	503 (6.8)	461 (7.2)	42 (4.1)
Advanced fibrosis[a]	500 (4.4)	399 (6.3)	101 (2.0)
Cirrhosis[a]	64 (0.6)	58 (0.9)	6 (0.1)

Note: Data are presented as mean (SD), median [P25–P75], or n and percentage.
[a]At-risk MASH was based on FAST ≥0.35, advanced fibrosis on Agile 3+ ≥0.679, and cirrhosis on Agile 4 ≥0.565.
Abbreviations: APRI, AST platelet ratio; BMI, body mass index; CORE, cirrhosis outcome risk estimator; DOR, diagnostic odds ratio; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HFS, Hepamet fibrosis core; LRS, LiverRisk score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis score-5; NFS, NAFLD fibrosis score; SAFE, steatosis-associated fibrosis estimator.

diagnosis of LSM ≥8 kPa, LSM ≥12 kPa, and advanced fibrosis.

Diagnostic yield of NITs when referring 20% of the at-risk population

Next, simulating a scenario in which 20% of the target population would be referred by selecting a cut-off corresponding with the 80th percentile, the MAF-5 obtained the highest sensitivity for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, and advanced fibrosis, ranging from 63% for LSM ≥8 kPa to 89% for at-risk MASH. The NFS obtained the highest sensitivity at the 80th percentile cut-off for cirrhosis with 88% (Figure 3). The corresponding cut-offs, sensitivity, specificity, NPV, PPV, and DOR are provided in Supplemental Table S4, http://links.lww.com/HEP/J793.

NIT performance among subgroups

The AUC of NITs across subgroups was consistent and typically did not differ >10% within each NIT, compared

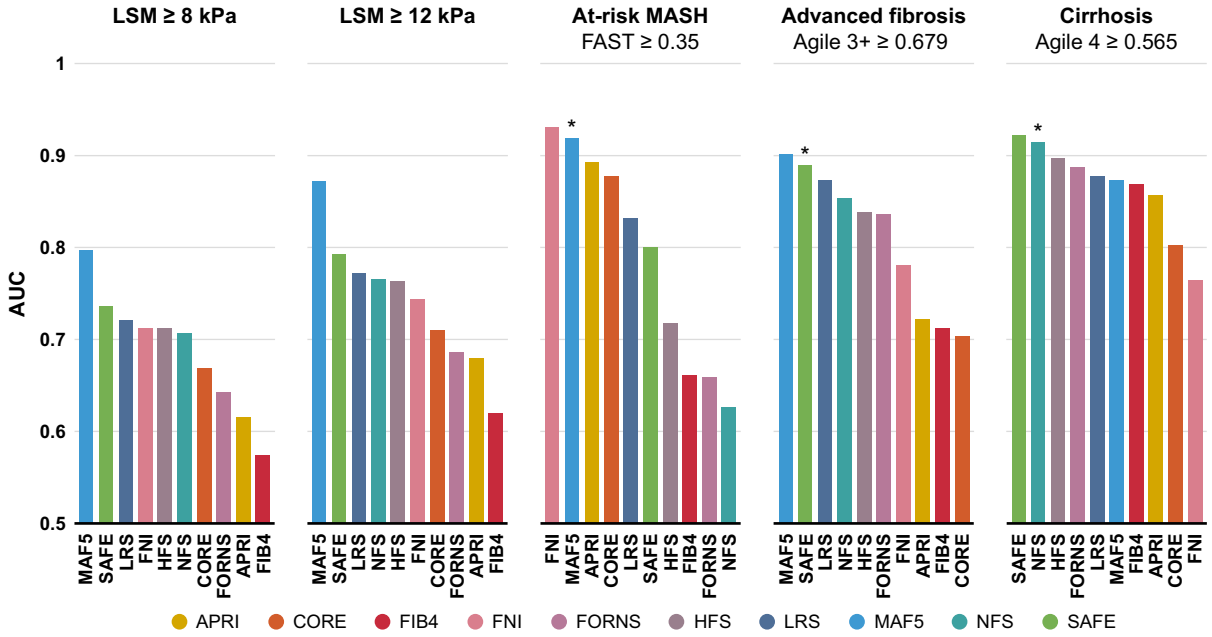

FIGURE 1 AUC per NIT for the detection of increased liver stiffness, at-risk MASH, advanced fibrosis, and cirrhosis in the target population for screening. A comparison of 10 NITs among 11,404 individuals with 881, 252, 503, 500, and 64 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. HFS, FNI, NFS, and SAFE were based on 6,375 participants, with 659, 219, 461, 399, and 58 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. *Non-inferior compared to the best-performing test based on the DeLong test. Abbreviations: APRI, AST platelet ratio; CORE, cirrhosis outcomes risk estimator; DOR, diagnostic odds ratio; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HFS, Hepamet fibrosis core; LRS, LiverRisk score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis score-5; MASH, metabolic dysfunction–associated steatohepatitis; NIT, non-invasive test; NFS, NAFLD fibrosis score; SAFE, steatosis-associated fibrosis estimator.

to the complete cohort for increased liver stiffness (Figure 4). However, the AUC of tests including (markers of) DM in their algorithm dropped >10% for advanced fibrosis (eg, the FNI, HFS, LRS, and MAF-5) in subgroups with DM.

The yield when the NIT cut-off was set at the 80th percentile (reflecting a scenario referring to 20% of the at-risk population) revealed more substantial differences in obtained sensitivity across subgroups (Supplemental Figure S2, http://links.lww.com/HEP/J792). The NITs that include age in their algorithm (FIB-4, LRS, SAFE, NFS, CORE, FORNS, HFS) did not exceed 30% sensitivity for the detection of LSM ≥8 kPa and LSM ≥12 kPa among individuals aged 18–35 years, whereas the sensitivity increased among individuals aged 65–80 years. Differences were less evident when focusing on individuals with obesity or DM. Scores that include (markers of) DM, such as FNI, LRS, and MAF-5, yielded higher sensitivity among individuals with DM when compared to the overall population.

DISCUSSION

We investigated the discriminative value of 10 NITs in the risk stratification of increased LSM, at-risk MASH, advanced fibrosis, and cirrhosis in the target population for screening in a population-based meta-cohort. Screening for clinically relevant (significant) fibrosis and at-risk MASH is important because of the associated long-term reduced prognosis, to educate patients, to prevent prognosis to (pre)cirrhosis and to identify the target population for treatment and clinical trials.[3,37] In this comprehensive study, the MAF-5 was the best-performing NIT for this purpose based on the following targets: LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH (similar to FNI), and advanced fibrosis (similar to SAFE). The FNI and SAFE, however, were significantly inferior for all but 1 or 2 diagnostic criteria. Hence, the MAF-5 could be a universal first-line test with consistent accuracy across relevant subgroups (age groups, BMI, and DM). The SAFE, and not MAF-5, was the best test for diagnosis (or exclusion) of cirrhosis, which is, however, not the primary target condition in screening strategies.

In the last decade, several promising tests have been developed that may be useful in screening algorithms instead of the FIB-4, which is currently endorsed by the EASL, AASLD, and AGA as a first-line screening tool.[9–18] Most tests, including the FIB-4, are not designed for early liver disease but for advanced liver disease in a high-prevalence hospital-based population, drastically affecting their discriminative value in the target population for screening: the general population with cardiometabolic dysfunction. Indeed, we have shown in this study that the FIB-4 is not able to accurately

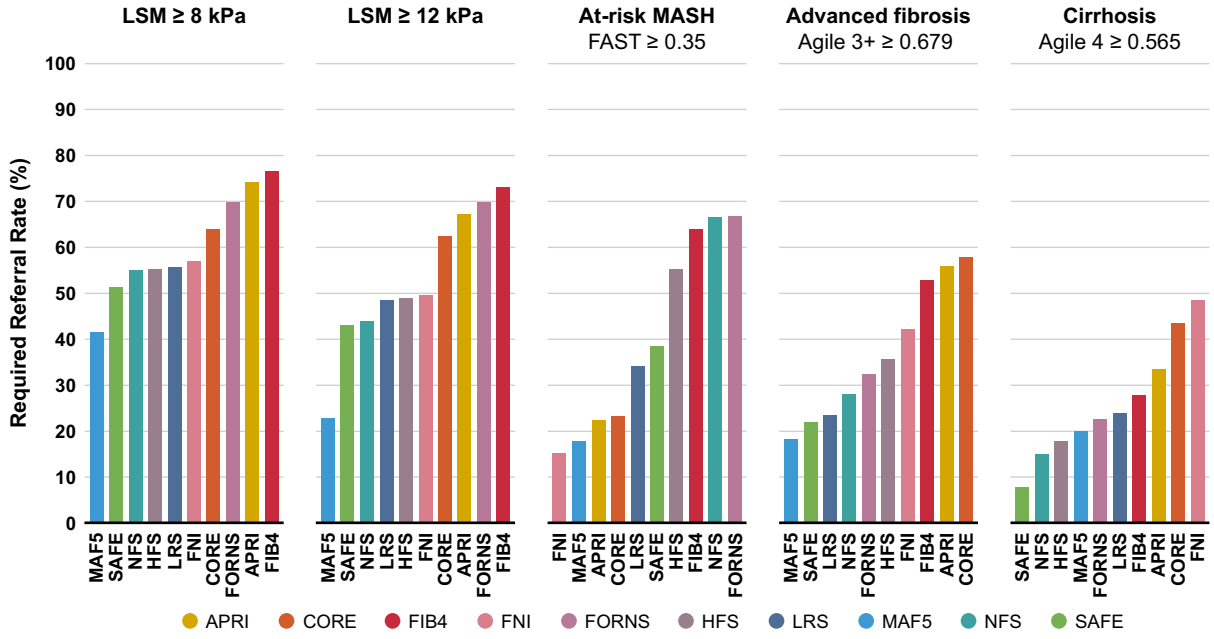

FIGURE 2 Required referral rate per NIT to detect 80% of increased liver stiffness, at-risk MASH, advanced fibrosis, and cirrhosis in the target population for screening. A comparison of 10 NITs among 11,404 individuals with 881, 252, 503, 500, and 64 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. HFS, FNI, NFS, and SAFE were based on 6375 participants, with 659, 219, 461, 399, and 58 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. Abbreviations: APRI, AST platelet ratio; CORE, cirrhosis outcomes risk estimator; DOR, diagnostic odds ratio; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HFS, Hepamet fibrosis core; LRS, LiverRisk score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis score-5; MASH, metabolic dysfunction–associated steatohepatitis; NFS, NAFLD fibrosis score; NIT, non-invasive test; SAFE, steatosis-associated fibrosis estimator.

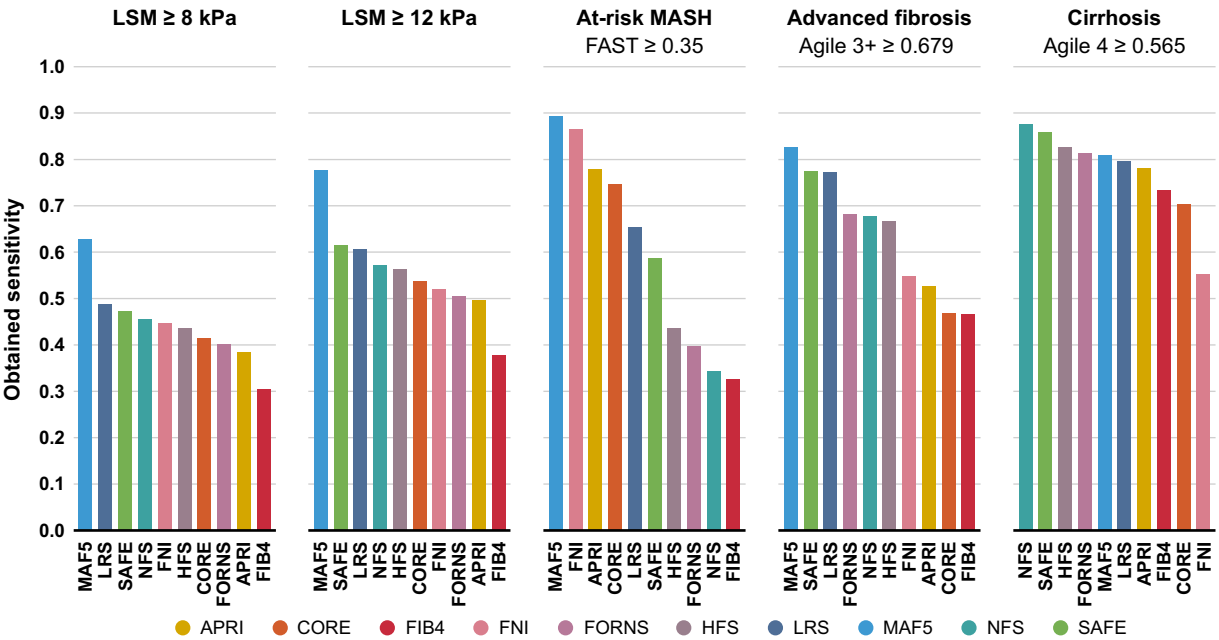

FIGURE 3 The yielded sensitivity when 20% of the at-risk population would be referred for further workup. A comparison of 10 NITs among 11,404 individuals with 881, 252, 503, 500, and 64 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. HFS, FNI, NFS, and SAFE were based on 6375 participants, with 659, 219, 461, 399, and 58 cases for LSM ≥8 kPa, LSM ≥12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. Abbreviations: APRI, AST platelet ratio; CORE, cirrhosis outcomes risk estimator; DOR, diagnostic odds ratio; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HFS, Hepamet fibrosis core; LRS, LiverRisk score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis score-5; MASH, metabolic dysfunction–associated steatohepatitis; NFS, NAFLD fibrosis score; NIT, non-invasive test; SAFE, steatosis-associated fibrosis estimator.

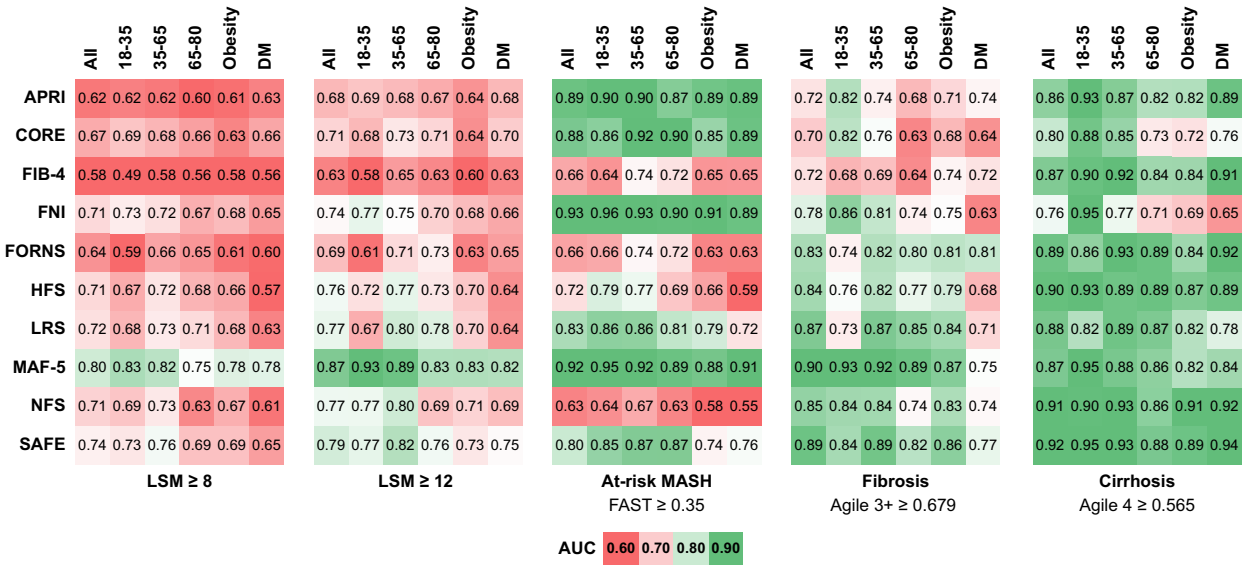

FIGURE 4 AUC per NIT for the detection of increased liver stiffness, at-risk MASH, advanced fibrosis, and cirrhosis in the target population for screening and subgroups. The AUC per NIT for the detection of increased liver stiffness and at-risk MASH, advanced fibrosis, and cirrhosis in the target population for screening and subgroups. The population comprised 11,404 individuals with 881, 252, 503, 500, and 64 cases for LSM ≥ 8 kPa, LSM ≥ 12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. HFS, FNI, NFS, and SAFE were based on 6375 participants, with 659, 219, 461, 399, and 58 cases for LSM ≥ 8 kPa, LSM ≥ 12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis, respectively. Abbreviations: APRI, AST platelet ratio; CORE, cirrhosis outcomes risk estimator; DOR, diagnostic odds ratio; FIB-4, fibrosis-4 index; FNI, fibrotic NASH index; HFS, Hepamet fibrosis core; LRS, LiverRisk Score; LSM, liver stiffness measurement; MAF-5, metabolic dysfunction–associated fibrosis score-5; MASH, metabolic dysfunction–associated steatohepatitis; NFS, NAFLD fibrosis score; SAFE, steatosis-associated fibrosis estimator.

discriminate the presence of early clinically relevant liver disease (such as LSM ≥ 8 kPa, LSM ≥ 12 kPa or at-risk MASH) and only has a good diagnostic value for the presence of cirrhosis, a stage of disease that is ideally prevented and not the target of screening strategies. These findings align with the recent recommendation to not use the FIB-4 for patient identification for clinical trials on MASH due to poor performance.[38] Our study indicates that the identification of patients who are eligible for treatment or can partake in clinical trials can be improved by using other NITs.

The AUC of NITs is independent of the cut-offs used. Therefore, we primarily focused on the AUC of the investigated NITs. Similarly, we deliberately did not investigate the NIT performance (NPV, PPV, sensitivity, specificity, and DOR) at the provided cut-offs in the original publications, to overcome issues with poor performance in low-prevalence populations (especially for those scores that were developed in high-prevalence populations, such as tertiary care biopsy cohorts). Moreover, the investigated NITs were not uniform in their original diagnostic purpose (eg, detection of F ≥ 2, F ≥ 3, or MASH). However, NITs are commonly applied beyond their intended use (eg, FIB-4 was originally developed for the detection of F3–F4 in patients with HIV and HCV).[39,40] Instead of the provided cut-offs, we simulated the consequences of adopting the investigated NITs in a screening program that aims to detect 80% of the cases, aligning the sensitivity of the occult blood tests used for colon cancer screening.[35] Additionally, taking into account the background prevalence, we investigated the yield of referring 20% of the population based on individual NITs. Finally, while the NPV is an important marker of NIT performance, it is less suitable for low-prevalence conditions, and we therefore used the DOR (which is background prevalence independent) in addition to the NPV.[41]

The MAF-5 was the most suited test to detect LSM ≥ 8 kPa, LSM ≥ 12 kPa, and advanced fibrosis in the target population for screening, based on AUC. Additionally, it obtained the highest AUC for at-risk MASH, together with the FNI (AUC: 0.92 and 0.93, respectively, $p = $ ns). Regarding the detection of cirrhosis, the SAFE score had the highest AUC. This indicates that referral pathways may be improved by using the MAF-5, FNI, and/or SAFE for the detection of early but clinically relevant, as well as more advanced liver disease.

Consistent with the AUC analysis, the MAF-5 required the lowest number of participants to be referred for further workup to maintain 80% sensitivity to detect LSM ≥ 8 kPa, LSM ≥ 12 kPa, and advanced fibrosis, the FNI for at-risk MASH, and the SAFE score for cirrhosis. The PPV for these tests ranged between 6% and 15%, which at first glance seems low; however, they are similar when compared to the diagnostic yield of colon cancer screening, which requires more invasive follow-up tests.[42] Similarly, the yield of referring participants > 80th percentile was highest for the MAF-5 in the detection of LSM ≥ 8 kPa, LSM ≥ 12 kPa, at-risk MASH, and advanced fibrosis.

Although the AUC across age subgroups and individuals with DM or obesity is quite consistent compared to the overall population, the yield with a single cut-off for all subgroups differed substantially. With a cut-off at the 80th percentile of the NIT in the overall group, the sensitivity for detection of LSM $\geq$8 kPa dropped to $<$10% for age-dependent scores such as FIB-4, SAFE, NFS, FORNS, and HFS among individuals aged 18–35 years old, making these tests unsuitable for mono-cut-off screening strategies. These findings align with previously identified concerns with age-dependent NITs.[4,6,43] The MAF-5 and FNI were the age-independent tests obtaining the best performance in the population aged 18–35 years with metabolic dysfunction, with the MAF-5 being better at detecting increased LSM while being equal in the detection of at-risk MASH. Liver disease among individuals aged 18–35 years should not be neglected, and although cirrhosis was rare, LSM $\geq$8 kPa and at-risk MASH were present in 5.9% and 6.2%, respectively, illustrating the magnitude of the problem of liver disease even among young individuals. Age-independent tests like MAF-5 and FNI are therefore superior for case finding among young individuals compared to FIB-4, while not introducing additional complexity of screening strategies due to age-dependent cut-offs.

Based on the NPV, one could argue that all investigated tests have adequate performance. However, the worst obtainable NPV is 1 minus background prevalence, which translates to 0.933 for LSM $\geq$8 kPa up to 0.994 for cirrhosis in this study. Hence, especially for low-prevalence conditions, the DOR can be considered a better parameter to assess test performance. The DOR is the ratio between the odds of having the outcome among those being ruled in, compared to those ruled out.[41] This parameter underscores again the poor performance of the FIB-4 (DOR 1.25 for LSM $\geq$8 kPa) compared to other NITs, for example, the MAF-5, SAFE, FNI, and LRS (DOR 6.45, 4.35, 3.68, and 3.47, respectively).

This comprehensive comparison of NITs is important to improve referral pathways that aim at risk-stratifying for early but clinically relevant liver disease in low-prevalence populations, as well as case findings for clinical trials and eligibility for pharmaceutical treatment. The currently recommended test, the FIB-4, was consistently among the worst-performing NITs across every outcome: increased LSM and MASH, advanced fibrosis, and cirrhosis. These findings indicate that replacing the FIB-4 with, for example, the MAF-5, FNI, or SAFE (depending on the aim of the referral pathway) substantially improves the accuracy of these pathways. The prevalence of advanced liver disease among participants being ruled out will be lower, and the overall yield of the referral pathway will consequently increase. Although screening with FIB-4 was already shown to be cost-effective, implementing a more accurate (and not more costly) test will result in an even more cost-effective program and should be considered for guideline updates.[44] The potential of further improving screening strategies by employing sequential NIT testing, utilizing the strength of multiple individual tests, should be evaluated by future studies.

Switching from FIB-4 to other scores, such as MAF-5, may result in a more complex initial test since these are not solely lab-based and require assessment of waist circumference (WC). However, there is an ongoing transition from BMI to WC in obesity diagnosis and in cardiovascular risk management.[45,46] Therefore, WC which is an easily obtainable should be widely available in the future.[45,46] Although a simple initial test is important, preventing unnecessary referrals might be even more essential to (1) prevent flooding of already strained healthcare systems, (2) to keep referring healthcare professionals engaged, and (3) also to prevent unnecessary patient distress associated with a false positive test.

Limitations

Despite this comprehensive comparison of 10 NITs in the target population for screening, the following limitations need to be considered. First, due to a lack of data on albumin, globulin, and HbA1c, the HFS, FNI, NFS, and SAFE could not be calculated in the Rotterdam Study. To account for differences in background prevalence, the PPV and NPV were, however, adjusted to be applicable to the overall background prevalence.[36] Second, NITs including very specific parameters such as the NIS2+, MACK-3, or ELF that include, for example, TIMP-1, CK18, miR-34a-5p, and/or YKL-40, were not investigated in this study, nor were scores that included liver stiffness.[38,47] These tests, although promising, are unlikely to be performed in a low-prevalence setting in primary care and are more suited as second-line tests like the ELF test, which is already positioned as such in the guidelines.[3] Third, this study did not include liver biopsy data and outcomes were based on liver stiffness (which itself is a crucial step in the referral pathways) or liver stiffness-based scores with good diagnostic accuracy in the external validation cohorts like FAST (sensitivity 92%, specificity 49%), Agile 3+ (sensitivity 61%, specificity 87%—90%) and Agile 4 (sensitivity 44%–55%, specificity 97%).[30,31] The either poor specificity (for at-risk MASH) or poor sensitivity (Agile 3+ and Agile 4) may have resulted in misclassification. Fourth, the MAF-5 has been developed in the NHANES 2017–2020 cohort on LSM $\geq$8 kPa and may therefore benefit in this specific comparison. However, also for other endpoints (LSM $\geq$12 kPa, at-risk MASH, advanced fibrosis, and cirrhosis), MAF-5 was consistently among the best-performing tests. Moreover, in a cohort-specific

comparison, the MAF-5 was among the best-performing tests supporting the main findings. Finally, there was no data on confirmation of the diagnosis of (advanced) liver fibrosis, which would be ideal for assessing the potential of NITs in the population-based setting further. This would require an extensive and large-scale study which would logistically and ethically not be feasible.

CONCLUSIONS

We demonstrated that the currently recommended test for screening, the FIB-4, was among the poorest performing NITs in this comprehensive comparison study of 10 NITs in the target population of screening programs. Instead, the MAF-5 was most suited for ruling in and ruling out LSM $\geq$8 kPa, LSM $\geq$12 kPa, advanced fibrosis, the FNI or MAF-5 for MASH, and the SAFE for cirrhosis. Pathways for early but significant liver disease in low-prevalence populations are likely to become substantially more feasible by using the MAF-5, FNI, or SAFE instead of FIB-4.

DATA AVAILABILITY STATEMENT

Data are publicly available from the NHANES database (https://www.cdc.gov/nchs/nhanes/index.htm). Data can be obtained upon request. Requests should be directed toward the management team of the Rotterdam Study (secretariat.epi@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository.

AUTHOR CONTRIBUTIONS

Collection of data, study design, data analysis, and writing of the manuscript: Laurens A. van Kleef, Jesse Pustjens, and Willem Pieter Brouwer. Critical review of the manuscript, writing of the manuscript, approval of the final version, and approval of submission: all authors.

FUNDING INFORMATION

Financial support was provided by the Foundation for Liver and Gastrointestinal Research, Rotterdam, the Netherlands. The funding source did not influence the study design, data collection, analysis, and interpretation of the data, nor the writing of the report and decision to submit for publication.

ACKNOWLEDGMENTS

We gratefully acknowledge the contribution of the participants of the NHANES cohort and Rotterdam study, research assistants, and facilitating personnel.

CONFLICTS OF INTEREST

Jörn M. Schattenberg consults for and is on the speakers' bureau for Boehringer Ingelheim, Gilead, Ipsen, Madrigal, and Novo Nordisk. He consults for Akero, Alentis, Alexion, Altimmune, AstraZeneca, 89bio, Bionorica, GSK, HistoIndex, Inventiva, Kríya Therapeutics, Lilly, MSD, Nordic Bioscience, Northsea Therapeutics, Novartis, Pfizer, Roche, Sanofi, Siemens Healthineers, Summit Clinical, and Vantage Biosciences Research. He is on the speakers' bureau for AbbVie and Worldwide Clinical Trials. He owns stock in Hepta Bio. Adriaan G. Holleboom consults for Echosens, Novo Nordisk, Inventiva, and Gilead. He is on the speakers' bureau for Echosens, Novo Nordisk, Julius Clinical, and Lilly. He is supported by an Amsterdam UMC Innovation grant, the Dutch Gastroenterology Foundation MLDS, Holland~Health TKI-PPP, and Horizon Europe GRIP on MASH. Robert J. de Knegt advises, is on the speakers' bureau for, and has received grants from Gilead and Echosens. He advises AbbVie. He received grants from Inventiva. Harry LA Janssen consults for and received grants from Gilead, GSK, Roche, and Vir. He consults for Aligos, Grifols, Precision Bio, Academic Medical Education, and H.C. Wainwright & Co. He received grants from Janssen. Sven M. Francque consults for and is on the speakers' bureau for AbbVie, Allergan, Bayer, Eisai, Genfit, Gilead, Intercept, Inventiva, Janssen, MSD, Novo Nordisk, Promethera, and Siemens. He received grants paid to his institution from Astellas, Falk Pharma, Genfit, Gilead, GlympsBio, Janssen, Inventiva, MSD, Pfizer, and Roche. He consults for AbbVie, Actelion, Aelin Therapeutics, AgomAb, Aligos Therapeutics, Allergan, Alnylam, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, BMS, CSL Behring, Coherus, Echosens, Dr. Falk Pharma, Eisai, Enyo, Galapagos, Galmed, Genentech, Genfit, Genflow Biosciences, Gilead, Intercept, Inventiva, Janssen, PRO.MED.CS Praha, Julius Clinical, Madrigal, Medimmune, MSD, Mursla Bio, NGM Bio, Novartis, Novo Nordisk, Promethera, Roche, Siemens Healthineers, and Weatherden. He holds a senior clinical investigator fellowship from the Research Foundation Flanders (FWO) (1802154N). Willem Pieter Brouwer is on the speakers' bureau for Eli Lilly. He advises Novo Nordisk. He received grants from 89bio, Boehringer Ingelheim, Novo Nordisk, and Inventiva. The remaining authors have no conflicts to report.

ORCID

Laurens A. van Kleef https://orcid.org/0000–0002–2333–1182
Jesse Pustjens https://orcid.org/0009–0005–5514–4874
Jörn M. Schattenberg https://orcid.org/0000–0002–4224–4703
Adriaan G. Holleboom https://orcid.org/0000–0002–2911–2917
Maarten E. Tushuizen https://orcid.org/0000–0001–6342–9056
Robert J. de Knegt https://orcid.org/0000–0003–0934–6975

M. Arfan Ikram https://orcid.org/0000–0003–0372–8585
Harry L.A. Janssen https://orcid.org/0000–0003–2398–8392
Sven M. Francque https://orcid.org/0000–0002–7527–4714
Willem Pieter Brouwer https://orcid.org/0000–0001–8713–1481

REFERENCES

1. 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.
2. 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.
3. Tacke F, Horn P, Wai-Sun Wong V, Ratziu V, Bugianesi E, Francque S, et al. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol. 2024;81:492–542.
4. van Kleef LA, de Knegt RJ, Ayada I, Pan Q, Brouwer WP. The steatosis-associated fibrosis estimator (SAFE) score: Validation in the general US population. Hepatol Commun. 2023;7:e0075.
5. Udompap P, Therneau TM, Canning RE, Benson JT, Allen AM. Performance of American Gastroenterological Association Clinical Care Pathway for the risk stratification of patients with nonalcoholic fatty liver disease in the US population. Hepatology. 2023;77:931–41.
6. van Kleef LA, Sonneveld MJ, de Man RA, de Knegt RJ. Poor performance of FIB-4 in elderly individuals at risk for chronic liver disease—Implications for the clinical utility of the EASL NIT guideline. J Hepatol. 2022;76:2526–8.
7. Chang M, Chang D, Kodali S, Harrison SA, Ghobrial M, Alkhouri N, et al. Degree of discordance between FIB-4 and transient elastography: An application of current guidelines on general population cohort. Clin Gastroenterol Hepatol. 2024;22:1453–461.e2.
8. Harrison SA, Taub R, Neff GW, Lucas KJ, Labriola D, Moussa SE, et al. Resmetirom for nonalcoholic fatty liver disease: A randomized, double-blind, placebo-controlled phase 3 trial. Nat Med. 2023;29:2919–28.
9. van Kleef LA, Francque SM, Prieto-Ortiz JE, Sonneveld MJ, Sanchez-Luque CB, Prieto-Ortiz RG, et al. MAF-5 predicts liver fibrosis risk and outcome in the general population with metabolic dysfunction. Gastroenterology. 2024;167:357–367.e9.
10. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. Hepatology. 2007;46:32–6.
11. Sripongpun P, Kim WR, Mannalithara A, Charu V, Vidovszky A, Asch S, et al. The steatosis-associated fibrosis estimator (SAFE) score: A tool to detect low-risk NAFLD in primary care. Hepatology. 2023;77:256–67.
12. Serra-Burriel M, Juanola A, Serra-Burriel F, Thiele M, Graupera I, Pose E, et al. Development, validation, and prognostic evaluation of a risk score for long-term liver-related outcomes in the general population: A multicohort study. Lancet. 2023;402:988–96.
13. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007;45:846–54.
14. Ampuero J, Pais R, Aller R, Gallego-Durán R, Crespo J, García-Monzón C, et al. Development and validation of Hepamet fibrosis scoring system—A simple, noninvasive test to identify patients with nonalcoholic fatty liver disease with advanced fibrosis. Clin Gastroenterol Hepatol. 2020;18:216–225.e.
15. Tavaglione F, De Vincentis A, Bruni V, Gallo IF, Carotti S, Tuccinardi D, et al. Accuracy of controlled attenuation parameter for assessing liver steatosis in individuals with morbid obesity before bariatric surgery. Liver Int. 2022;42:374–83.
16. Forns X, Ampurdanès S, Llovet JM, Aponte J, Quintó L, Martínez-Bauer E, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. Hepatology. 2002;36:986–92.
17. Loaeza-del-Castillo A, Paz-Pineda F, Oviedo-Cárdenas E, Sánchez-Ávila F, Vargas-Voráckova F. AST to platelet ratio index (APRI) for the noninvasive evaluation of liver fibrosis. Ann Hepatol. 2008;7:350–7.
18. Strandberg R, Talbäck M, Hammar N, Hagström H. OS-057-YI CORE: A new risk score measuring GGT, AST, and ALT outperforms FIB-4 when predicting the risk of cirrhosis in a primary care setting. J Hepatol. 2024;80:S40.
19. Forlano R, Stanic T, Jayawardana S, Mullish BH, Yee M, Mossialos E, et al. A prospective study on the prevalence of MASLD in people with type-2 diabetes in the community. Cost effectiveness of screening strategies. Liver Int. 2024;44:61–71.
20. Graupera I, Thiele M, Serra-Burriel M, Caballeria L, Roulot D, Wong GLH, et al. Low accuracy of FIB-4 and NAFLD fibrosis scores for screening for liver fibrosis in the population. Clin Gastroenterol Hepatol. 2022;20:2567–76.
21. Chen X, Goh GBB, Huang J, Wu Y, Wang M, Kumar R, et al. Validation of non-invasive fibrosis scores for predicting advanced fibrosis in metabolic-associated fatty liver disease. J Clin Transl Hepatol. 2022;10:589–94.
22. Ikram MA, Kieboom BCT, Brouwer WP, Brusselle G, Chaker L, Ghanbari M, et al. The Rotterdam Study. Design update and major findings between 2020 and 2024. Eur J Epidemiol. 2024;39:183–206.
23. Johnson CL, Paulose-Ram R, Ogden CL, Carroll MD, Kruszon-Moran D, Dohrmann SM, et al. National health and nutrition examination survey: Analytic guidelines, 1999–2010. Vital Health Stat. 2013;Sep (161):1–24.
24. Akinbami LJ, Chen T-C, Davy O, Ogden CL, Fink S, Clark J, et al. National Health and Nutrition Examination Survey, 2017–March 2020 prepandemic file: Sample design, estimation, and analytic guidelines. Vital Health Stat 1. 2022;(190):1–36.
25. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology. 2023;78:1966–86.
26. van Kleef LA, Lu Z, Ikram MA, de Groot NMS, Kavousi M, de Knegt RJ. Liver stiffness not fatty liver disease is associated with atrial fibrillation: The Rotterdam Study. J Hepatol. 2022;77:931–8.
27. 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.
28. Bedossa P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology. 2014;60:565–75.
29. Roulot D, Costes JL, Buyck JF, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut. 2011;60:977–84.
30. 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.
31. Sanyal AJ, Foucquier J, Younossi ZM, Harrison SA, Newsome PN, Chan WK, et al. Enhanced diagnosis of advanced fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile scores. J Hepatol. 2023;78:247–59.
32. Zhang X, Heredia NI, Balakrishnan M, Thrift AP. Prevalence and factors associated with NAFLD detected by vibration controlled transient elastography among US adults: Results from NHANES 2017–2018. PLoS One. 2021;16:e0252164.
33. National Center for Health Statistics. National Health and Nutrition Examination Survey 2017–2020. Liver Ultrasound Transient Elastography Procedures Manual; 2022. http://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_LUX.htm
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics. 1988;44:837–45.
35. Niedermaier T, Tikk K, Gies A, Bieck S, Brenner H. Sensitivity of fecal immunochemical test for colorectal cancer detection differs according to stage and location. Clin Gastroenterol Hepatol. 2020;18:2920–28.e6.
36. Monaghan TF, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, et al. Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value. Medicina. 2021;57:503.
37. Noureddin M, Charlton MR, Harrison SA, Bansal MB, Alkhouri N, Loomba R, et al. Expert panel recommendations: Practical clinical applications for initiating and monitoring resmetirom in patients with MASH/NASH and moderate to noncirrhotic advanced fibrosis. Clin Gastroenterol Hepatol. 2024;22:2367–77.
38. Ratziu V, Harrison SA, Hajji Y, Magnanensi J, Petit S, Majd Z, et al. NIS2+™ as a screening tool to optimize patient selection in metabolic dysfunction-associated steatohepatitis clinical trials. J Hepatol. 2024;80:209–19.
39. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology. 2006;43:1317–25.
40. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol. 2009;7:1104–12.
41. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: A single indicator of test performance. J Clin Epidemiol. 2003;56:1129–35.
42. Bailey SER, Abel GA, Atkins A, Byford R, Davies SJ, Mays J, et al. Diagnostic performance of a faecal immunochemical test for patients with low-risk symptoms of colorectal cancer in primary care: An evaluation in the South West of England. Br J Cancer. 2021;124:1231–6.
43. Sugiyama A, Kurisu A, E B, Ouoba S, Ko K, Rakhimov A, et al. Distribution of FIB-4 index in the general population: Analysis of 75,666 residents who underwent health checkups. BMC Gastroenterol. 2022;22:241.
44. Park H, Yoon EL, Kim M, Kwon SH, Kim D, Cheung R, et al. Cost-effectiveness study of FIB-4 followed by transient elastography screening strategy for advanced hepatic fibrosis in a NAFLD at-risk population. Liver Int. 2024;44:944–54.
45. Busetto L, Dicker D, Frühbeck G, Halford JCG, Sbraccia P, Yumuk V, et al. A new framework for the diagnosis, staging and management of obesity in adults. Nat Med. 2024;30:2395–9.
46. Rubino F, Cummings DE, Eckel RH, Cohen RV, Wilding J, Brown WA, et al. Definition and diagnostic criteria of clinical obesity. Lancet Diabetes Endocrinol. 2025;13:221–62.
47. Chuah K-H, Wan Yusoff WNI, Sthaneshwar P, Nik Mustapha NR, Mahadeva S, Chan W-K. MACK-3 (combination of hoMa, Ast and CK18): A promising novel biomarker for fibrotic non-alcoholic steatohepatitis. Liver Int. 2019;39:1315–24.

How to cite this article: van Kleef LA, Pustjens J, Schattenberg JM, Holleboom AG, Castro Cabezas M, Tushuizen ME, et al. Comparison of diagnostic accuracy and utility of non-invasive tests for clinically significant liver disease in a general population with metabolic dysfunction. Hepatology. 2026;83:591–602. https://doi.org/10.1097/HEP.0000000000001356