Low Accuracy of FIB-4 and NAFLD Fibrosis Scores for Screening for Liver Fibrosis in the Population

Isabel Graupera,[*,a] Maja Thiele,[‡,a] Miquel Serra-Burriel,[§,b] Llorenç Caballeria,[||] Dominique Roulot,[¶] Grace Lai-Hung Wong,[#] Núria Fabrellas,[**] Indra Neil Guha,[‡‡] Anita Arslanow,[§§,¶¶] Carmen Expósito,[||] Rosario Hernández,[||||] Guruprasad Padur Aithal,[‡‡] Peter R. Galle,[§§] Guillem Pera,[||] Vincent Wai-Sun Wong,[#] Frank Lammert,[¶¶] Pere Ginès,[*] Laurent Castera,[##] and Aleksander Krag,[‡,b] for the Investigators of the LiverScreen Consortium

*Liver Unit Hospital Clínic. Institut D'investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Centro de Investigación En Red de Enfermedades Hepáticas Y Digestivas (Ciberehd), Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain; ‡Center for Liver Research, Department of Gastroenterology and Hepatology, Odense University Hospital, and Department for Clinical Research, University of Southern Denmark, Odense, Denmark; §Epidemiology, Biostatistics, and Prevention Institute, University of Zurich, Zurich, Switzerland; ||USR Metropolitana Nord, IDIAP Jordi Gol, ICS Institut Català de la Salut, Barcelona, Spain; ¶Department of Hepatology, AP-HP, Hopital Avicenne, Bobigny, France; Université Paris 13, Sorbonne Paris Cité, Villetaneuse, France; #Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong; **Faculty of Medicine and Health Sciences, School of Nursing, University of Barcelona; Institut D'investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Centro de Investigación En Red de Enfermedades Hepáticas Y Digestivas (Ciberehd), Barcelona, Spain; ‡‡NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, United Kingdom; §§Department of Internal Medicine I, University Medical Centre of the Johannes Gutenberg-University Mainz, Mainz, Germany; ||||CAP La Marina, Institut Català de la Salut de Barcelona, Barcelona, Spain; ¶¶Department of Medicine II, Saarland University Medical Center, Saarland University, Homburg, Germany; and the ##Hôpital Beaujon; Department of Hepatology, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, Clichy, France; Université Paris VII, Inserm Umr 1149, Centre de Recherche Sur L'inflammation, Paris, France

BACKGROUND & AIMS: **Fibrosis-4 (FIB-4) and the nonalcoholic fatty liver disease fibrosis score (NFS) are the 2 most popular noninvasive blood-based serum tests proposed for widespread fibrosis screening. We therefore aimed to describe the accuracy of FIB-4 and NFS to detect elevated liver stiffness as an indicator of hepatic fibrosis in low-prevalence populations.**

METHODS: **This study included a total of 5129 patients with concomitant measurement of FIB-4, NFS, and liver stiffness measurement (LSM) by Fibroscan (Echosens, France) from 5 independent population-based cohorts from Spain, Hong Kong, Denmark, England, and France; 3979 participants from the general population and 1150 from at-risk cohorts due to alcohol, diabetes, or obesity. We correlated LSM with FIB-4 and NFS, and calculated pre- and post-test predictive values of FIB-4 and NFS to detect elevated LSM at 8 kPa and 12 kPa cutoffs. The mean age was 53 ± 12 years, the mean body mass index was 27 ± 5 kg/m^2, and 2439 (57%) were women. One in 10 patients (552; 11%) had liver stiffness ≥8 kPa, but 239 of those (43%) had a normal FIB-4, and 171 (31%) had normal NFS. The proportion of false-negatives was higher in at-risk patients than the general population. FIB-4 was false-negative in 11% of diabetic subjects, compared with 2.5% false-negatives with NFS. Waist circumference outperformed FIB-4 and NFS for detecting LSM ≥8 kPa in the general population. Almost one-third (28%–29%) of elevated FIB-4/NFS were false-positive in both the general population and at-risk cohorts.**

[a]Authors share co-first authorship. [b]Authors share co-corresponding authorship.

Abbreviations used in this paper: ALD, alcoholic liver disease; ALT, alanine aminotransferase; AUC, area under the receiving operating characteristics curve; BMI, body mass index; DK, Denmark; ES, Spain; FIB-4, Fibrosis-4 test; FR, France; HK, Hong Kong; LSM, liver stiffness measurement; NAFLD, nonalcoholic fatty liver disease; NFS, NAFLD fibrosis score; NIT, non-invasive tests; NPV, negative predictive value; PPV, positive predictive value; TE, transient elastography; UK, United Kingdom; ULN, upper limit of normal.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2021.12.034

CONCLUSIONS: **FIB-4 and NFS are suboptimal for screening purposes due to a high risk of overdiagnosis and a non-negligible percentage of false-negatives, especially in patients with risk factors for chronic liver disease. Waist circumference emerged as a potential first step to identify patients at risk for liver fibrosis in the general population.**

Keywords: Alcoholic Liver Disease; Liver Fibrosis; NITs; Nonalcoholic Fatty Liver Disease; Noninvasive Fibrosis Scores; Screening; Transient Elastography.

See editorial on page 2448.

The global burden of liver disease is rising. In the United States, mortality from cirrhosis increases at a 3% rate per year, with the most prominent annual increase, exceeding 10%, among people aged 25 to 34 years.[1] The obesity epidemic alongside high rates of excessive alcohol use cause growing prevalence of nonalcoholic fatty liver disease (NAFLD) and alcohol-related liver disease (ALD).[2,3] Yet, efficient and standardized referral pathways are still lacking, to ensure timely detection of patients who have progressed to significant and advanced fibrosis.[4] Access to testing is not the limiting factor: General practice in the United Kingdom tripled the number of liver blood tests from 2000 to 2015.[5] Therefore, most effort now centers around which blood tests to use for fibrosis screening.[4] Liver stiffness measurement (LSM) by transient elastography (TE) is the reference noninvasive marker due to its high accuracy for advanced fibrosis and its prognostic ability.[6–8] Although LSM is a cost-effective screening tool, its availability is limited in primary care settings.[9] Therefore, the Fibrosis-4 test (FIB-4) and NAFLD fibrosis score (NFS) have been proposed as easy, low-cost alternatives for first-line screening.[10–12]

Sequential use of a cheap, blood-based test first, followed by confirmatory LSM, increases the number of detected patients with cirrhosis,[13] reduces the proportion of futile referrals,[11] and is highly cost-effective.[14] However, tests such as FIB-4 and NFS are developed and validated for fibrosis diagnostics in high-prevalence populations from secondary or tertiary care.[15,16] In primary care, evidence from smaller population samples suggests that there is only a modest correlation between elevated FIB-4 and evidence of liver fibrosis indicated by elevated LSM.[17] And although a Swedish epidemiological study found that elevated FIB-4 predicted 10-year risk of liver-related events in the general population, 65% of events happened in participants with low FIB-4.[18]

We therefore aimed to evaluate the discriminatory ability of FIB-4 and NFS in large population-based study cohorts from several European countries and Hong Kong, for their immediate value as predictors of fibrosis severity evidenced by elevated liver stiffness. Our primary aim was to correlate FIB-4 and NFS to liver stiffness in both the general population and groups at risk of chronic liver disease due to alcohol, diabetes, or obesity. Second, we aimed to describe diagnostic accuracies of FIB-4 and NFS for detection of elevated liver stiffness at 8 kPa and 12 kPa cutoffs, as standards for ruling in and ruling out liver fibrosis in NAFLD and ALD.[19]

Methods

Study Design and Population

This was a cross-sectional study combining 5 independent cohorts that performed prospective screening for liver fibrosis with TE, in different countries: Spain (ES),[20,21] Hong Kong (HK),[22] United Kingdom (UK),[13] France (FR),[23] and Denmark (DK)[24] were included. Data on demographics, physical exam, and clinical and laboratory parameters were recorded alongside with comorbidities at the time of TE. For the present study, we only included participants from the cohorts with TE and noninvasive test (NIT) serological markers determined during the same visit. Only patients with at least 10 successful acquisitions and interquartile range/median $<30\%$ were included. Unsuccessful TE measurement rate was 0.5% in general population cohorts and 3.3% in at-risk cohorts. Subjects gave informed consent to participate in the original screening studies, all of which were conducted in accordance with the Declaration of Helsinki. A Data Sharing Agreement between the investigators was used for the present study.

The presence of risk factors for chronic liver disease increases the prevalence of fibrosis and may change the performance of a diagnostic test. In the present study, we therefore divided participants in to 2 different subgroups: the general population and at-risk. We defined general population cohorts as those that included randomly invited participants from the general population (ES and HK). We defined at risk-cohorts as those that included patients based on the presence of 1 or more risk factors for chronic liver disease (UK, DK, and FR).[25] The cohort from the UK included patients above 18 years with hazardous alcohol use or diabetes, and the cohort from DK comprised patients above 18 years at with hazardous alcohol consumption. Both studies defined hazardous alcohol consumption as $\geq$14 units per week for women and $\geq$21 units per week for men. The cohort from FR was originally from the general population; however, only 26% from the original published cohort had concomitant TE and NIT. All patients from this subset of the FR cohort had at least 1 risk factor for liver disease, and we therefore defined the FR cohort as at-risk. All cohorts excluded patients with previously known liver disease.

Consequently, no included subject had chronic viral hepatitis. Figure 1 presents the flowchart of the enrollment of the patients from each original cohort and the final eligibility of patients that were included.

Statistical Analysis

Categorical data are presented as numbers (percentage). Continuous data are presented as mean (standard deviation) for normally distributed variables and medians (range) for non-normally distributed variables. We compared continuous, parametric variables with the 2-sample *t* test and used the Mann-Whitney test for nonparametric variables. Categorical variables were compared using the χ^2 test or the Fisher exact test. We assessed the correlation between NITs and LSM by their bivariate distribution in scatter plots and correlation coefficients.

We assessed the diagnostic performance of FIB-4 and NFS by computing sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiving operating characteristics curve (AUC) using LSM as reference, with TE $\geq$8 kPa and $\geq$12 kPa as cutoffs. We chose LSM $\geq$8 kPa because several referral pathways have recommended it as the cutoff to identify patients with presumed liver fibrosis who require further evaluation.[6,19] We added analyses for $\geq$12 kPa because recent data suggests that this cutoff is optimal for ruling in advanced liver fibrosis ($\geq$ stage 3).[19] All statistical tests were 2-sided using an α level of 0.05. Cutoff points for NITs were derived from the literature. For FIB-4, cutoffs of $<$1.30 and of $\geq$2.67 were used to their accuracy to rule-out and rule-in elevated liver stiffness, with additional sensitivity analysis for cutoff of 3.25. We also tested the robustness of FIB-4 results by excluding patients with elevated transaminases, body mass index (BMI) $>$35 kg/m^2, and young or old age. For NFS, we used <-1.455 and $\geq$ 0.676. Additionally, we built a multivariable logistic regression model to assess the relative associations of individual variables.

What You Need to Know

Background

The Fibrosis-4 (FIB-4) and nonalcoholic fatty liver disease fibrosis (NFS) scores have been proposed as screening tools, but their accuracy to detect liver fibrosis has never been tested in a general population setting, or in primary care in groups at risk of liver disease.

Findings

FIB-4 and NFS scores correlate poorly with liver stiffness, which result in a substantial proportion of false-positive and false-negatives. This limits their usefulness as screening tools in the primary care setting.

Implications for patients care

New noninvasive methods are needed with better accuracy for fibrosis detection in low-prevalence fibrosis settings.

All analyses were performed in R (R Foundation for Statistical Computing, Vienna, 2019).

Results

Patients' Demographics

Of the 6474 participants from the original cohorts, we included 5129 with concomitant LSM and NITs, 3979 from general population cohorts and 1150 from at-risk cohorts. Figure 1 presents the flowchart of the original study populations and final inclusion of the patients from each cohort.

Table 1 shows the baseline characteristics of the subjects of the 5 cohorts. Subjects from the general population cohorts, ES and HK, had lower proportions of diabetes, arterial hypertension, or hazardous alcohol

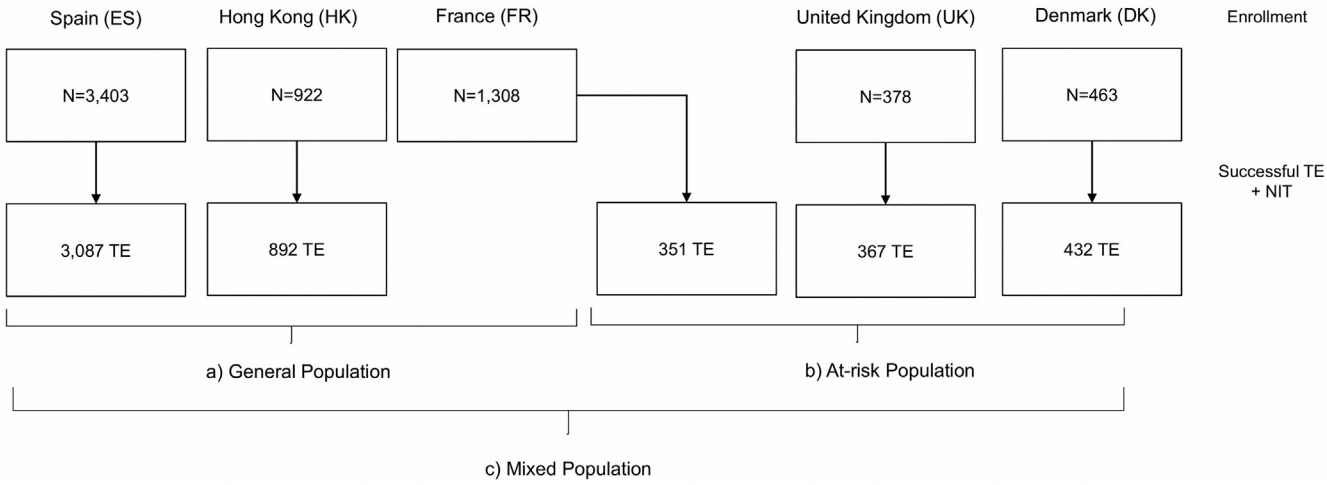


Figure 1. Patient selection flowchart.

Table 1. Patient Characteristics of the Cohorts

	All (N = 5129)	ES (n = 3087)	HK (n = 892)	FR (n = 351)	UK (n = 367)	DK (n = 432)
Women	2678 (52.2)	1772 (57.4)	512 (57.4)	159 (45.3)	120 (32.7)	115 (26.6)
Age, *y*	55.0 (12.2)	55.5 (11.9)	48.0 (10.6)	59.8 (9.24)	61.5 (15.0)	56.5 (10.5)
Alcohol, *yes*	1538 (30.1)	403 (13.1)	198 (22.2)	211 (60.1)	294 (82.8)	432 (100)
BMI, kg/m^2	27.2 (5.15)	28.3 (4.89)	22.7 (3.50)	27.0 (4.34)	28.5 (5.31)	27.3 (5.29)
Waist circumference, *cm*	91.9 (13.5)	94.7 (12.8)	81.5 (10.1)	90.7 (12.4)	–	104 (15.6)
SBP, *mm Hg*	129 (18.2)	126 (17.0)	129 (19.7)	133 (14.3)	–	139 (21.2)
DBP, *mm Hg*	80 (10.6)	80 (10.0)	82 (12.3)	79 (8.70)	–	79 (12.1)
DM, *yes*	642 (13.4)	354 (11.5)	38 (4.3)	–	205 (55.9)	45 (10.4)
Obesity, *yes*	1334 (26.0)	978 (31.7)	25 (2.80)	80 (22.8)	125 (34.6)	126 (29.2)
HT, *yes*	1437 (30.1)	886 (28.7)	136 (15.2)	–	167 (45.6)	248 (57.4)
Glucose, *mg/dL*	5.79 (8.10)	5.87 (10.0)	5.15 (0.93)	5.75 (1.81)	–	6.64 (1.93)
Cholesterol, *umol/L*	5.38 (1.07)	5.53 (1.03)	5.14 (1.00)	5.38 (0.96)	4.63 (1.28)	–
LDL-C	3.28 (0.98)	3.48 (0.90)	3.00 (0.87)	3.35 (0.81)	2.43 (1.14)	3.05 (1.25)
HDL-C	1.45 (0.39)	1.43 (0.34)	1.54 (0.41)	1.46 (0.39)	1.46 (0.56)	1.41 (0.46)
Triglycerides, *umol/L*	1.32 (0.92)	1.23 (0.75)	1.37 (1.21)	1.30 (0.78)	1.75 (1.04)	1.60 (1.14)
AST, *UI/L*	25.4 (16.7)	23.7 (9.16)	21.4 (12.3)	22.6 (14.8)	28.0 (19.0)	45.9 (37.8)
ALT, *UI/L*	26.0 (18.2)	24.0 (14.5)	25.9 (16.1)	23.4 (16.1)	28.3 (19.4)	40.1 (33.6)
GGT, *UI/L*	49.6 (120)	33.4 (39.0)	–	30.9 (36.4)	–	180 (314)
Bilirubin, *umol/L*	11.9 (5.71)	11.3 (4.89)	13.5 (5.98)	–	11.9 (5.06)	12.3 (9.25)
Albumin, *g/L*	43.6 (3.77)	44.0 (3.24)	45.2 (2.59)	–	37.8 (2.90)	41.9 (4.80)
Ferritin, n*g/L*	206 (294)	119 (119)	483 (478)	165 (179)	–	292 (366)
Platelets, $\times 10^9/L$	243 (61.6)	243 (59.4)	244 (54.3)	252 (58.6)	241 (62.3)	241 (87.7)
LSM, *kPa*	6.0 (6.4)	5.0 (2.2)	4.8 (2.6)	6.1 (2.6)	8.0 (7.4)	14.1 (17.3)
LSM ≥8	552 (10.8)	169 (5.5)	55 (6.2)	42 (12.0)	99 (27.0)	187 (43.3)
LSM ≥12	209 (4.1)	36 (1.2)	13 (1.5)	7 (2.0)	46 (12.5)	107 (24.8)

Note: Data are presented as number (%) or mean (standard deviation).
Abd, Abdominal; ALT, alanine transaminase; AST, aspartate transaminase; BMI, body mass index; DBP, diastolic blood pressure; DK, Denmark; DM, diabetes mellitus; ES, Spain; FIB-4, Fibrosis-4 Index for Liver Fibrosis; FR, France; GGT, gamma-glutamyl transferase; Hb, hemoglobin; HDL-C, high-density lipoprotein cholesterol; HK, Hong Kong; HT, hypertension; LDL-C, low-density lipoprotein cholesterol; LSM, liver stiffness measurement; NFS, Nonalcoholic Fatty Liver Disease Fibrosis Score; SBP, systolic blood pressure; UK, United Kingdom.

consumption compared the cohorts from FR, UK, and DK. Overweight and obesity were present in almost two-thirds of the population (38% with BMI between 25 and 29.9 and 25% with BMI ≥30). Subjects of the at-risk populations had higher transaminase levels and lower albumin. Median alanine aminotransferase (ALT) levels were below the gender-specific cutoffs, and only 0.1% of men and 0.4% of women presented ALT levels 2 to 3 times the upper limit of normal (ULN).

LSM and Performance of NITs

There was a significant but poor correlation between LSM values and FIB-4 and NFS, both in the general population cohorts (Figure 2*a*) and at-risk population cohorts (Figure 2*b*). The correlation between FIB4 and TE was 45%, whereas the correlation between NFS and TE was 29% (both *P*-values < .001). We observed the lowest median LSM values in the general population cohorts from HK (4.8 ± 2.6 kPa) and ES (5.0 ± 2.2 kPa), compared with higher median LSM in the at-risk cohorts, at 6.1 ± 2.6, 8.0 ± 7.4, and 14.1 ± 17.3 kPa for FR, UK, and DK cohorts, respectively (Table 1). The prevalence of LSM ≥8 kPa as indirect evidence of significant liver fibrosis in our cohorts was 5.6% for the general population and 29% for at-risk patients, whereas LSM ≥12 kPa occurred in 1.2% of the general population and 14% of at-risk patients (Table 1).

As seen in Figure 2, the majority of patients with LSM <8 kPa were well-classified by NITs. However, 239 of 3308 patients (7.2%) with FIB-4 <1.3 had LSM ≥8 kPa. Something similar was seen with NFS, where we observed 5.7% false-negatives (171 of 3019 with low NFS had LSM ≥8 kPa). At-risk population cohorts

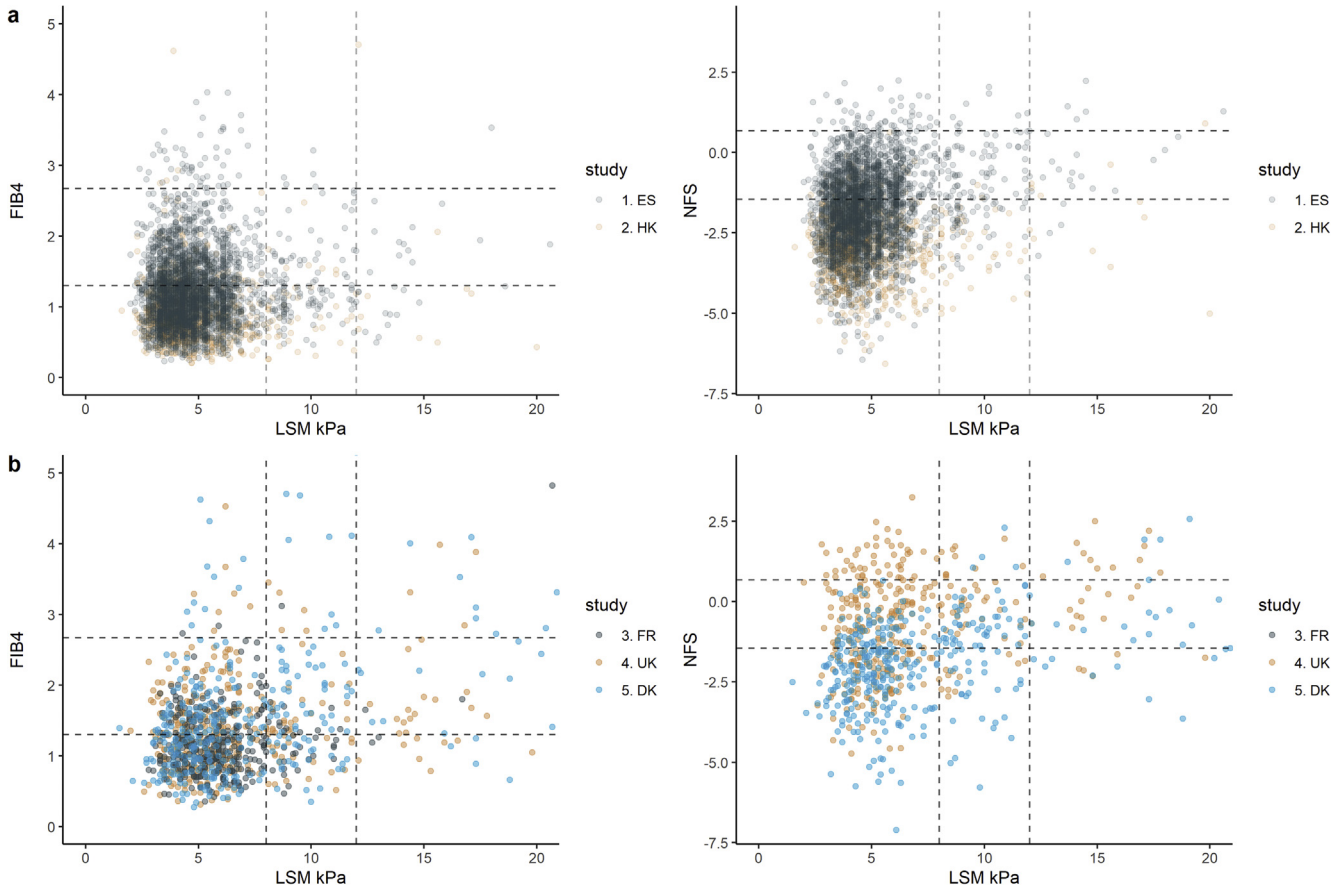


Figure 2. Bivariate distributions of LSM and NITs by population. *a*, General population; *b*, At-risk population. *a*, Correlation LSM – FIB-4: 16.53%, LSM – NFS: 16.67%; *b*, Correlation LSM - FIB-4: 46.93%, LSM – NFS: 39.35%.

exhibited a higher proportion of false-negative patients (8%–9%) than the general population cohorts (2%–4%) with both NIT scores. Tables 2 and 3 display the frequencies and distributions of LSM ≥8 and ≥12 kPa in the 2 types of population cohorts.

We also observed a significant proportion of false-positives with both noninvasive scores in the general population and at-risk population cohorts. There were 1179 participants (29%) with FIB-4 ≥1.3 and LSM <8 kPa and 1130 participants (28%) with NFS ≥−1.45 among the general population, and similar percentages in at-risk cohorts with 329 patients (28%) for FIB-4 and 225 (28%) for NFS.

Supplementary Tables 1 to 3 show the sensitivity, specificity, NPV, and PPV of FIB-4 and NFS to detect patients above specific cutoffs of liver stiffness. In the general population, NFS performed better than FIB-4, with higher sensitivity to detect patients above 8 kPa. However, in at risk-populations, there were no differences between NFS and FIB-4 performances for detection of increased liver stiffness. The sensitivity and specificity for the cutoff commonly used to rule out fibrosis (LSM <8 kPa) with FIB-4 in the general population was 37% and 69%, respectively, whereas NFS had a sensitivity of 52% and specificity of 69%. In at-risk populations, FIB-4 and NFS sensitivity to rule out fibrosis was 70% and 77%, and specificity to rule in was 60% and 55%.

The comparison of the baseline characteristics of patients misclassified due to false-negative results in NITs (FIB-4 and NFS) with those well-classified (Supplementary Table 4) showed that false-negative results of NITs were associated to obesity, diabetes, arterial hypertension, hazardous alcohol consumption, higher waist circumference, and higher values of ALT. We also found younger age to predict false negative results by FIB-4. In 341 patients aged <35 years, 14 of 16 (87%) had FIB-4 <1.30 despite having LSM ≥8 kPa. In contrast, older age predicted FIB-4 false-positives. In 1118 patients >65 years, 658 of 782 (84%) had FIB-4 ≥1.30 despite having LSM <8 kPa. Excluding participants with ALT >2 × ULN did not change results, nor did excluding obese patients with BMI >35 kg/m^2.

Performance of Waist Circumference for Liver Fibrosis Screening

We next performed a multivariate analysis of individual risk factors associated with elevated LSM. Waist circumference appeared as one of the most significant

Table 2. Classification of Patients From General Population Cohorts by NIT and LSM

LSM	FIB-4 ≥1.3 Negative n = 2717	FIB-4 ≥1.3 Positive n = 1262	FIB-4 ≥2.7 Negative n = 3910	FIB-4 ≥2.7 Positive n = 69
≥8 kPa				
Positive	141 (5.19)	83 (6.58)	213 (5.45)	11 (15.9)
Negative	2576 (94.8)	1179 (93.4)	3697 (94.6)	58 (84.1)
≥12 kPa				
Positive	25 (0.92)	24 (1.90)	43 (2.43)	6 (8.70)
Negative	3256 (99.1)	1238 (98.1)	3867 (98.9)	60 (91.3)

LSM	NFS ≥−1.45 Negative n = 2677	NFS ≥−1.45 Positive n = 1244	NFS ≥−0.67 Negative n = 3847	NFS ≥−0.67 Positive n = 74
≥8 kPa				
Positive	105 (3.92)	114 (9.16)	200 (5.20)	19 (25.7)
Negative	2572 (96.1)	1130 (90.8)	3647 (94.8)	55 (74.3)
≥12 kPa				
Positive	14 (0.52)	32 (2.57)	36 (0.94)	10 (13.5)
Negative	2663 (99.5)	1212 (97.4)	3811 (99.1)	64 (86.5)

Note: Data are presented as number (percent).
FIB-4, Fibrosis-4 test; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NIT, noninvasive testing.

Table 3. Classification of Patients From At-risk Cohorts by NIT and LSM

LSM	Low FIB-4 cutoff FIB-4 <1.3 n = 591	Low FIB-4 cutoff FIB-4 ≥1.3 n = 559	High FIB-4 cutoff FIB-4 <2.7 n = 1027	High FIB-4 cutoff FIB-4 ≥2.7 n = 123
8 kPa				
≥8 kPa	98 (16.6)	230 (41.1)	226 (22.0)	102 (82.9)
<8 kPa	493 (83.4)	329 (58.9)	801 (78.0)	21 (17.1)
12 kPa				
≥12 kPa	27 (4.57)	133 (23.8)	77 (7.50)	83 (67.5)
<12 kPa	564 (95.4)	426 (76.2)	950 (92.5)	40 (32.5)

LSM	Low NFS cutoff NFS <−1.45 n = 342	Low NFS cutoff NFS ≥−1.45 n = 444	High NFS cutoff NFS <−0.67 n = 665	High NFS cutoff NFS ≥−0.67 n = 121
8 kPa				
≥8 kPa	66 (19.3)	219 (49.3)	209 (31.4)	76 (62.8)
<8 kPa	276 (80.7)	225 (50.7)	456 (68.6)	45 (37.2)
12 kPa				
≥12 kPa	23 (6.73)	130 (29.3)	91 (13.7)	62 (51.2)
<12 kPa	319 (93.3)	314 (70.7)	574 (86.3)	59 (48.8)

Note: Data are presented as number (percent).
FIB-4, Fibrosis-4 test; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NIT, noninvasive testing.

factors (Supplementary Figure 1). When compared with FIB-4 and NFS, waist circumference had a significantly higher AUC in the general population for liver fibrosis detection at TE ≥8 kPa (FIB-4 AUC, 0.572; NFS, 0.643; waist circumference, 0.716). The diagnostic accuracy of waist circumference also exceeded that of FIB-4 and NFS for TE ≥12 kPa, albeit not significantly so (AUC FIB-4, 0.642; NFS, 0.727; waist circumference, 0.756). Figure 3 shows the comparison of AUCs for waist circumference, FIB-4, and NFS in the combined cohorts, in the general population cohorts and the at-risk population cohorts.

NITS Performance in Patients With Specific Risk Factors: Diabetes or Hazardous Alcohol Consumption

Previous studies suggest that FIB-4 and NFS underperform in patients with diabetes,[26,27] so we explored how FIB-4 and NFS worked in the specific at-risk subpopulation of patients with diabetes. FIB-4 resulted in a high rate of false-negatives for the cutoff of LSM ≥8 kPa in the diabetic population (72 patients; 12%), but not NFS (16 patients; 2.5%) (Supplementary Table 6). In patients with a hazardous alcohol consumption, 30% had a false-negative FIB-4 (469 of 1538 patients), and 29% had a false-negative NFS.

Discussion

In the present study, we tested the ability of the 2 most used noninvasive scores, FIB-4 and NFS, for screening for elevated liver stiffness as the main predictive factor of liver fibrosis progression to decompensation and liver-related mortality.[28–31] Our large cohort study reports a significant proportion of false-negative subjects with FIB-4 or NFS, and a larger number of false-positives. A screening strategy based on either of these 2 scores would both miss patients that need referral and risk overdiagnosis and futile referrals. Waist circumference showed up as a potential useful tool to identify patients from the general population with high risk of liver fibrosis.

FIB-4 and NFS have been developed in high-prevalence populations at secondary centers, but their performance in population-based studies where screening is relevant has not been tested. To our knowledge, this is the largest study, with more than 5000 patients, that compares FIB-4 and NFS with TE in a population setting, constituting a representative spectrum of disease. From the random population with a 5.6% prevalence of elevated liver stiffness, to 43% with elevated liver stiffness in the population pre-selected for

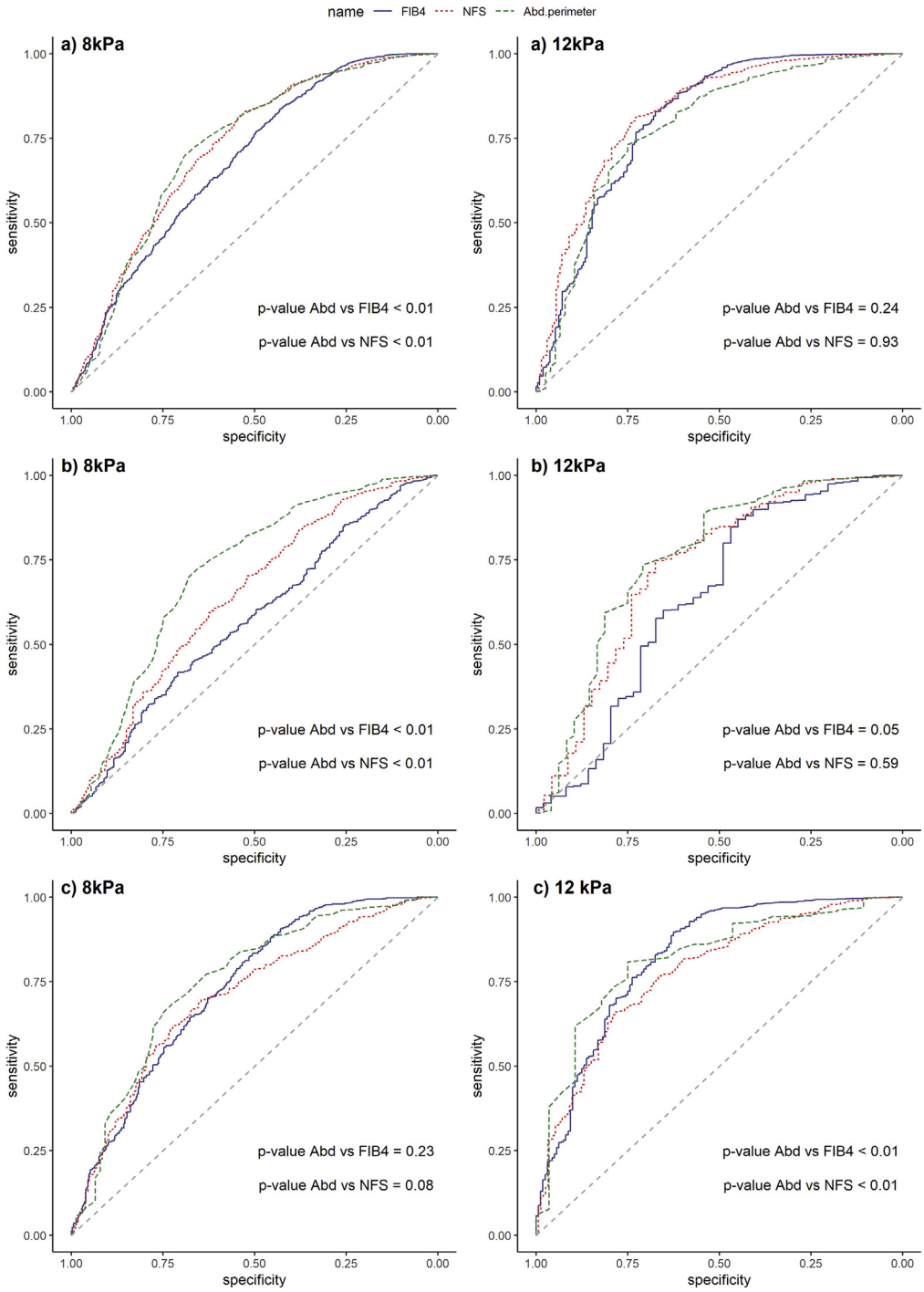


Figure 3. AUCs of FIB-4, NFS, and waist circumference. *a*, Combined population; *b*, General population; *c*, At-risk population. *a*, 8kPa: FIB-4, 0.683; NFS, 0.723; WC, 0.756. 12kPa: FIB-4, 0.805; NFS, 0.826; WC, 0.783. *a*, 8kPa: FIB-4, 0.572; NFS, 0.643; WC, 0.716. 12kPa: FIB-4, 0.642; NFS, 0.727; WC, 0.756. *a*, 8kPa: FIB-4, 0.729; NFS, 0.705; WC, 0.749. 12kPa: FIB-4, 0.821; NFS, 0.777; WC, 0.715. Abd, Abdominal perimeter; WC, waist circumference.

screening due to the most common lifestyle risk factors: obesity, type-2 diabetes, and alcohol consumption.

More than one-third of patients were misclassified by FIB-4 and NFS, worse in at-risk cohorts than the general population. Specifically, using TE $\geq$8 kPa, a total of 2% to 4% of all subjects had false-negative values of FIB-4/NFS and would have been missed. This proportion increased to 8% to 9% of the at-risk cohorts. In patients with diabetes, hazardous alcohol consumption, or age $<$35 years using FIB-4, the proportion of false-negatives was even higher. This is important, because screening tools should be especially suited for fibrosis detection in at-risk patients. Despite the high number of false-negatives, NPVs for FIB-4 and NFS were above 90% in the general population cohorts. This may give the impression of a good test performance to rule out disease. However, NPV is a misleading metric in low prevalence cohorts, because even a coin toss would have a high NPV. If the prevalence is 5.6%, as in the general population cohort presented here, flipping a coin would result in a NPV of 94.4%. Accordingly, the FIB-4 NPV of 95% or NFS NPV of 96% as seen in Supplementary Table 2 is only marginally better than a coin toss. We are therefore cautious of suggesting FIB-4 or NFS as screening tools. We instead suggest to continue searching for more

sensitive tests. An optimal screening program will likely consist of 3 phases: preselection based on risk factors for chronic liver disease, followed by screening with a highly sensitive test to rule in disease, but with adequate specificity to avoid over-diagnosis. Finally, this would be followed by a confirmatory test with high specificity in those patients who screened positive.

The proportion of false-positives was also high in our study, at 28% to 29%. This was especially true for participants >65 years, where FIB-4 frequently overestimated the risk of fibrosis. Overdiagnosis is potentially harmful to healthy subjects and leads to futile use of health care resources.[32,33]

An interesting finding of our study is that waist circumference was strongly associated with increased liver stiffness, and outperformed FIB-4 and NFS for detection of patients with TE ≥8 kPa in the general population, but not in at-risk cohorts. Waist circumference is easy to measure, cheap, and was recently shown to be an independent predictor of all-cause mortality in patients with ALD and NAFLD.[34] However, its ability as a diagnostic and prognostic marker needs validation.

We are aware that our reference standard, LSM with TE, is a surrogate for liver fibrosis and risk of liver-related events. However, outcome assessments in patients with TE ≥8 kPa require a very long follow-up period. Additionally, liver biopsy is not feasible in the screening setting and is an imperfect gold standard.[35,36] Several studies support the applicability and validity of liver stiffness as a useful surrogate for liver fibrosis in low prevalence population.[22–24] With respect to TE false-positives, it is worth noting that increased LSM due to inflammation would also raise FIB-4/NFS. Moreover, exclusion of patients with ALT above 2 × ULN or BMI >35 did not change our results. Another limitation of the present study is that we could not assess the Enhanced Liver Fibrosis Test's role in a 2-step approach, as recently proposed.[11,14]

Currently there are several projects under evaluation in Europe and the United States (LiverScreen, Renown, Seal, Scarred liver project) to establish the best strategy for fibrosis screening.[37] Data coming from these consortia will help define the best approach for liver fibrosis detection and personalized referral pathways.

Conclusion

In conclusion, although several studies have reported a reduction in referrals after introduction of NITs, our data show a significant percentage of false-negatives and false-positives with FIB-4 and NFS in the general population setting, even higher in the at-risk population. Therefore, although FIB-4 and NFS represent a step in the right direction, we should be cautious implementing them without first searching for better first-line screening tools and referral pathways.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2021.12.034.

References

1. Tapper EB, Parikh ND. Mortality due to cirrhosis and liver cancer in the United States, 1999-2016: observational study. BMJ 2018;362:k2817.
2. Kim D, Li AA, Gadiparthi C, et al. Changing trends in etiology-based annual mortality from chronic liver disease, from 2007 through 2016. Gastroenterology 2018;155:1154–1163.e3.
3. Pimpin L, Cortez-Pinto H, Negro F, et al. , EASL HEPAHEALTH Steering Committee. Burden of liver disease in Europe: epidemiology and analysis of risk factors to identify prevention policies. J Hepatol 2018;69:718–735.
4. Ginès P, Graupera I, Lammert F, et al. Screening for liver fibrosis in the general population: a call for action. Lancet Gastroenterol Hepatol 2016;1:256–260.
5. O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. BMJ 2018;363:k4666.
6. Nguyen-Khac E, Thiele M, Voican C, et al. Non-invasive diagnosis of liver fibrosis in patients with alcohol-related liver disease by transient elastography: an individual patient data meta-analysis. Lancet Gastroenterol Hepatol 2018;3:614–625.
7. Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1717–1730.
8. Boursier J, Vergniol J, Guillet A, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. J Hepatol 2016;65:570–578.
9. Serra-Burriel M, Graupera I, Torán P, et al. Investigators of the LiverScreen Consortium. Transient elastography for screening of liver fibrosis: cost-effectiveness analysis from six prospective cohorts in Europe and Asia. J Hepatol 2019;71:1141–1151.
10. Anstee QM, Lawitz EJ, Alkhouri N, et al. Noninvasive tests accurately identify advanced fibrosis due to NASH: baseline data from the STELLAR trials. Hepatology 2019;70:1521–1530.
11. Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. J Hepatol 2019;71:371–378.
12. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1264–1281.e4.
13. Harman DJ, Ryder SD, James MW, et al. Direct targeting of risk factors significantly increases the detection of liver cirrhosis in primary care: a cross-sectional diagnostic study utilising transient elastography. BMJ Open 2015;5:e007516.
14. Asphaug L, Thiele M, Krag A, et al. Cost-effectiveness of noninvasive screening for alcohol-related liver fibrosis. Hepatology 2020;71:2093–2104.
15. Sterling RK, Lissen E, Clumeck N, et al. APRICOT Clinical Investigators. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317–1325.

16. Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007;45:846–854.
17. Foschi FG, Domenicali M, Giacomoni P, et al. , Bagnacavallo Study Group. Is there an association between commonly employed biomarkers of liver fibrosis and liver stiffness in the general population? Ann Hepatol 2020;19:380–387.
18. Hagström H, Talbäck M, Andreasson A, et al. Ability of noninvasive scoring systems to identify individuals in the population at risk for severe liver disease. Gastroenterology 2020; 158:200–214.
19. Papatheodoridi M, Hiriart JB, Lupsor-Platon M, et al. Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease. J Hepatol 2021; 74:1109–1116.
20. Fabrellas N, Hernández R, Graupera I, et al. Prevalence of hepatic steatosis as assessed by controlled attenuation parameter (CAP) in subjects with metabolic risk factors in primary care. A population-based study. PLoS One 2018;13:e0200656.
21. Caballería L, Pera G, Arteaga I, et al. High prevalence of liver fibrosis among European adults with unknown liver disease: a population-based study. Clin Gastroenterol Hepatol 2018; 16:1138–1145.e5.
22. Wong VWS, Chu WCW, Wong GLH, et al. Prevalence of nonalcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. Gut 2012;61:409–415.
23. Roulot D, Costes J-L, Buyck J-F, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut 2011;60:977–984.
24. Thiele M, Detlefsen S, Sevelsted Møller L, et al. Transient and 2-dimensional shear-wave elastography provide comparable assessment of alcoholic liver fibrosis and cirrhosis. Gastroenterology 2016;150:123–133.
25. Thiele M, Madsen BS, Hansen JF, et al. Accuracy of the enhanced liver fibrosis test vs FibroTest, elastography, and indirect markers in detection of advanced fibrosis in patients with alcoholic liver disease. Gastroenterology 2018; 154:1369–1379.
26. Bril F, McPhaul MJ, Caulfield MP, et al. Performance of plasma biomarkers and diagnostic panels for nonalcoholic steatohepatitis and advanced fibrosis in patients with type 2 diabetes. Diabetes Care 2020;43:290–297.
27. Castera L. Non-invasive tests for liver fibrosis in NAFLD: creating pathways between primary healthcare and liver clinics. Liver Int 2020;40(Suppl 1):77–81.
28. Rasmussen DN, Thiele M, Johansen S, et al. GALAXY; MicrobLiver consortia. Prognostic performance of 7 biomarkers compared to liver biopsy in early alcohol-related liver disease. J Hepatol 2021;75:1017–1025.
29. Petta S, Sebastiani G, Viganò M, et al. Monitoring occurrence of liver-related events and survival by transient elastography in patients with nonalcoholic fatty liver disease and compensated advanced chronic liver disease. Clin Gastroenterol Hepatol 2021;19:806–815.e5.
30. Shili-Masmoudi S, Wong GL-H, Hiriart J-B, et al. Liver stiffness measurement predicts long-term survival and complications in non-alcoholic fatty liver disease. Liver Int 2020;40:581–589.
31. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015;149:389–397.e10.
32. Haymart MR, Miller DC, Hawley ST. Active surveillance for low-risk cancers - a viable solution to overtreatment? N Engl J Med 2017;377:203–206.
33. Johansson M, Jørgensen KJ, Brodersen J. The benefits of screening–and its harms. Lancet 2016;388:563–564.
34. Decraecker M, Dutartre D, Hiriart J-B, et al. Long-term prognosis of patients with alcohol-related liver disease or non-alcoholic fatty liver disease according to metabolic syndrome or alcohol use. Liver Int 2021 Oct 22. https://doi.org/10.1111/liv.15081:Online ahead of print.
35. Sanyal AJ, Harrison SA, Ratziu V, et al. The natural history of advanced fibrosis due to nonalcoholic steatohepatitis: data from the simtuzumab trials. Hepatology 2019;70:1913–1927.
36. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol 2020;73:1322–1332.
37. Ginès P, Castera L, Lammert F, et al. Population screening for liver fibrosis: towards early diagnosis and intervention for chronic liver diseases. Hepatology 2022;75:219–228.

Reprint requests
Address requests for reprints to: Miquel Serra-Burriel, PhD, Epidemiology, Biostatatistics, and Prevention Institute, University of Zurich, Hirschengraben 84, 8001, Zurich, Switzerland. e-mail: miquel.serraburriel@uzh.ch; fax: +41 44 634 6383. Aleksander Krag, MD, PhD, Centre for Liver Research, Department of Gastroenterology and Hepatology, Odense University Hospital, J.B. Winsløws Vej 4, Entrance 126, 2nd floor, 5000 Odense C, Denmark. e-mail: Aleksander.Krag@rsyd.dk; phone: +45 21 409 915.

CRediT Authorship Contributions

Isabel Graupera (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – original draft: Lead; Writing – review & editing: Equal)

Maja Thiele (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Miquel Serra-Burriel (Conceptualization: Equal; Data curation: Equal; Formal analysis: Lead; Writing – original draft: Equal; Writing – review & editing: Equal)

Llorenç Caballeria (Data curation: Equal; Methodology: Equal; Resources: Equal; Writing – review & editing: Equal)

Dominique Roulot (Data curation: Equal; Formal analysis: Equal; Methodology: Equal; Writing – review & editing: Equal)

Grace Lai-Hung Wong (Data curation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Nuria Fabrellas (Data curation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting)

Indra Neil Guha (Data curation: Supporting; Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting)

Anita Arslanow (Data curation: Equal; Methodology: Equal; Writing – review & editing: Supporting)

Carmen Exposito (Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting)

Rosario Hernández (Investigation: Equal; Methodology: Equal; Writing – review & editing: Supporting)

Guruprasad Padur Aithal (Investigation: Supporting; Methodology: Supporting)

Peter R. Galle (Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Equal)

Guillem Pera (Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Supporting)

Vincent Wai-Sun Wong (Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Frank Lammert (Conceptualization: Equal; Investigation: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Pere Ginès (Conceptualization: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Writing – review & editing: Equal)

Laurent Castera (Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

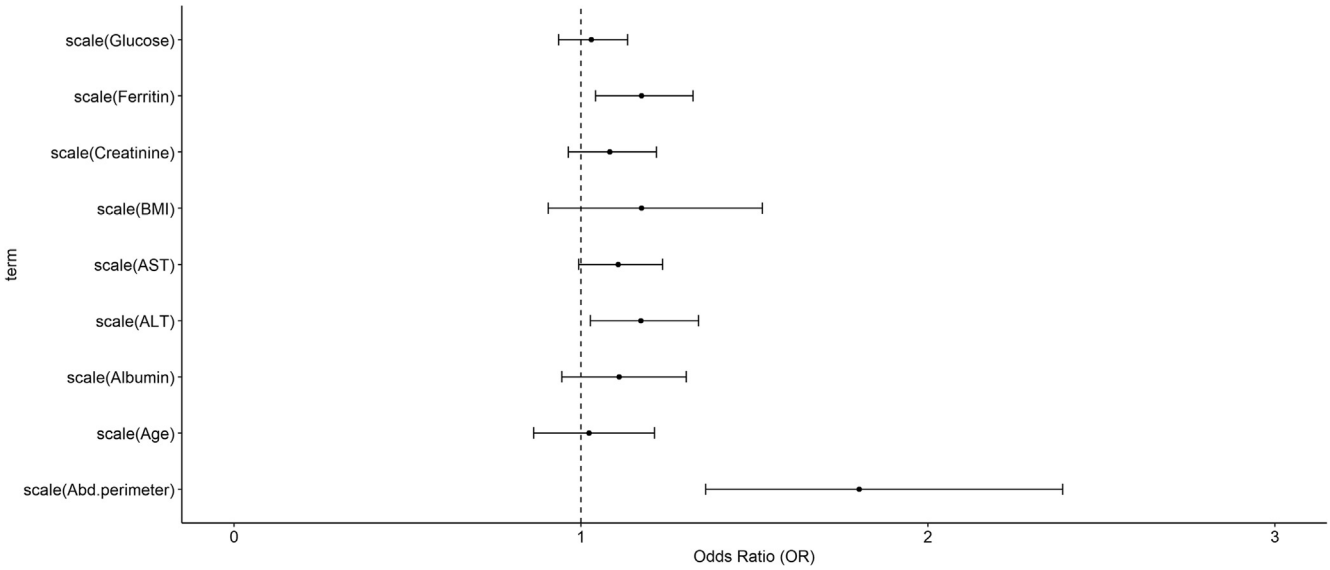
Aleksander Krag (Conceptualization: Lead; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Visualization: Equal; Writing – original draft: Lead; Writing – review & editing: Lead)

Conflicts of interest

These authors disclose the following: Isabel Graupera has received lecture fees from Gilead and Novartis. Maja Thiele has received lecture fees from Echosens and Siemens, and advisory board fees from GE Healthcare. Pere Gines reports grants and personal fees from Grifols, grants and personal fees from Gilead, grants from Mallinckrodt, personal fees from Promethera, personal fees from Martin Pharmaceuticals, grants from Ferring Pharmaceuticals, and grants and personal fees from Sequana. Vincent Wong has served as a consultant or advisory board member for AbbVie, Allergan, Echosens, Gilead Sciences, Janssen, Perspectum Diagnostics, Pfizer, and Terns; he has also received lecture fees from Bristol-Myers Squibb, Echosens, Gilead Sciences, and Merck. Aleksander Krag reports advisory board and lecture fees from Norgine and Siemens. The remaining authors disclose no conflicts.

Funding

The LiverScreen project is funded by the European Union's Horizon 2020 Research and Innovation Program (grant agreement number 847989). The Danish study was funded by Innovation Fund Denmark, Horizon 2020 (grant agreement number 668031) and the MicrobLiver project from the Novo Nordic Foundation Challenge program under grant number NNF15OC0016692. Part of this work was supported by the Fundación de Investigación sanitaria and cofunded by Instituto Carlos III (ISCIII)-Subdirección General de Evaluación and the European Regional Development Fund with the projects PI18/00862 to Isabel Graupera, and PI18/01330 to Nuria Fabrellas and PI18/00662 to Llorenç Caballeria. This work was also supported by a grant from Agencia de Gestió d'Ajuts Universitaris i de Recerca (AGAUR_SGR 01281).



Supplementary Figure 1. Multivariate model coefficients. Note: Logistic regression estimates. Includes study-fixed effects. Abd, Abdominal; AST, aspartate aminotransferase.

Supplementary Table 1. Performance Metrics of the Combined Population

	LSM kPa >8			LSM kPa >9.1			LSM kPa >10			LSM kPa >12			LSM kPa >15		
FIB-4	Threshold = 1.3	95% CI													
Sens.	57	52	61	62	57	67	66	61	71	75	69	81	82	75	87
Spec.	67	66	68	67	65	68	67	65	68	66	65	68	66	65	67
PPV	17	15	19	13	12	15	12	10	13	9	7	10	7	6	8
NPV	93	92	94	96	95	96	97	96	97	98	98	99	99	99	99
	Threshold = 2.676														
Sens.	20	17	24	27	22	31	30	25	35	43	36	50	54	46	62
Spec.	98	98	99	98	98	98	98	98	98	98	97	98	98	97	98
PPV	59	52	66	54	46	61	51	44	58	46	39	54	43	36	51
NPV	91	90	92	94	94	95	95	95	96	98	97	98	99	98	99
	Threshold = 3.25														
Sens.	14	17	24	19	15	24	22	17	27	33	26	39	41	33	49
Spec.	99	99	100	99	99	100	99	99	99	99	99	99	99	99	99
PPV	74	65	82	69	59	77	66	56	75	63	53	72	58	48	68
NPV	91	90	91	94	93	94	95	94	96	97	97	98	98	98	99
NFS	Threshold = −1.455														
Sens.	66	62	70	72	67	77	74	69	79	81	75	87	84	78	90
Spec.	68	66	69	67	66	69	67	65	68	66	65	68	66	64	67
PPV	20	18	22	15	14	17	14	12	15	10	8	11	7	6	9
NPV	94	93	95	97	96	97	97	97	98	99	99	99	99	99	100
	Threshold = 0.676														
Sens.	19	16	23	24	19	28	26	22	32	36	30	43	42	34	50
Spec.	98	97	98	97	97	98	97	97	98	97	97	98	97	97	98
PPV	49	42	56	44	37	51	42	35	49	37	30	44	32	25	39
NPV	91	90	92	94	93	95	95	94	96	97	97	98	98	98	98

Note: Data are presented as percentages.
CI, Confidence interval; FIB-4, Fibrosis-4 Index for Liver Fibrosis; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

Supplementary Table 2. Performance Metrics of the General Population Cohorts

	LSM kPa >8			LSM kPa >9.1			LSM kPa >10			LSM kPa >12			LSM kPa >15		
FIB-4	Threshold = 1.3	95% CI													
Sens.	37	31	44	42	34	51	45	35	54	49	34	64	65	41	85
Spec.	69	67	70	69	67	70	69	67	70	68	67	70	68	67	70
PPV	7	5	8	5	3	6	4	3	5	2	1	3	1	1	2
NPV	95	94	96	97	96	98	98	97	98	99	99	99	100	99	100
	Threshold = 2.676														
Sens.	5	2	9	7	3	12	8	4	15	12	5	25	25	9	49
Spec.	98	98	99	98	98	99	98	98	99	98	98	99	98	98	99
PPV	16	8	27	13	6	23	13	6	23	9	3	18	7	2	16
NPV	95	94	95	97	96	97	97	97	98	99	99	100	100	99	100
	Threshold = 3.25														
Sens.	2	0	5	3	1	7	4	1	9	8	2	20	15	3	38
Spec.	100	99	100	100	99	100	100	99	100	100	99	100	100	99	100
PPV	20	6	44	20	6	44	20	6	44	20	6	44	15	3	38
NPV	94	94	95	97	96	97	97	97	98	99	98	99	100	99	100
NFS	Threshold = −1.455														
Sens.	52	45	59	59	51	68	61	52	71	70	54	82	63	38	84
Spec.	69	68	71	69	68	71	69	68	71	69	67	70	68	67	70
PPV	9	8	11	6	5	8	5	4	7	3	2	4	1	0	2
NPV	96	95	97	98	97	98	98	98	99	100	99	100	100	99	100
NFS	Threshold = 0.676														
Sens.	9	5	13	12	7	19	14	8	22	22	11	36	32	13	57
Spec.	99	98	99	98	98	99	98	98	99	98	98	99	98	98	99
PPV	26	16	37	22	13	33	20	12	31	14	7	23	8	3	17
NPV	95	94	95	97	96	97	98	97	98	99	99	99	100	99	100

Note: Data are presented as percentages.
CI, Confidence interval; FIB-4, Fibrosis-4 Index for Liver Fibrosis; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

Supplemental Table 3. Performance Metrics of the At-risk Population Cohorts

	LSM kPa >8			LSM kPa >9.1			LSM kPa >10			LSM kPa >1			LSM kPa >15		
FIB-4	Threshold = 1.3	95% CI													
Sens.	70	65	75	73	68	79	77	71	83	83	76	89	84	77	90
Spec.	60	57	63	58	55	62	58	55	61	57	54	60	56	53	59
PPV	41	37	45	33	29	37	30	26	33	24	20	28	20	17	24
NPV	83	80	86	89	86	91	92	89	94	95	93	97	96	95	98
	Threshold = 2.676														
Sens.	31	26	36	38	32	44	42	35	49	52	44	60	59	50	67
Spec.	97	96	98	97	95	98	96	95	97	96	95	97	96	94	97
PPV	83	75	89	76	68	84	72	64	80	67	58	76	63	54	72
NPV	78	75	80	85	83	87	88	86	90	93	91	94	95	93	96
	Threshold = 3.25														
Sens.	23	19	28	28	23	34	31	25	38	40	32	48	45	36	54
Spec.	99	97	99	98	97	99	98	97	99	98	96	98	97	96	98
PPV	86	77	93	80	70	87	76	66	85	73	62	82	68	57	78
NPV	76	74	79	83	81	85	86	84	88	91	89	93	93	91	95
NFS	Threshold = −1.455														
Sens.	77	72	82	80	74	85	81	75	86	85	78	90	88	81	93
Spec.	55	51	60	53	49	57	52	48	56	50	46	54	50	46	54
PPV	49	45	54	41	36	45	36	32	41	29	25	34	25	21	30
NPV	81	76	85	87	83	90	89	85	92	93	90	96	95	93	97
	Threshold = 0.676														
Sens.	27	22	32	31	25	37	33	27	40	41	34	49	43	35	52
Spec.	91	88	93	91	88	93	91	88	93	91	88	93	90	88	92
PPV	63	54	71	57	48	66	55	45	64	51	40	60	46	37	56
NPV	69	65	72	77	73	80	80	77	83	86	83	89	89	86	91

Note: Data are presented as percentages.
CI, Confidence interval; FIB-4, Fibrosis-4 Index for Liver Fibrosis; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

Supplementary Table 4. False Positive and Negative Characteristics (Additional)

	LSM kPa 8 – FIB-4 1.3			LSM kPa 8 – NFS −1.45		
Variable	False negative n = 239	False positive n = 1508	Correct class n = 3382	False negative n = 171	False positive n = 1355	Correct class n = 3181
Age	54.4 (11.8)	63.0 (8.90)	51.5 (11.9)	50.8 (11.4)	63.3 (8.97)	51.2 (11.9)
Alcohol	0.42 (0.49)	0.31 (0.46)	0.29 (0.45)	0.49 (0.50)	0.28 (0.45)	0.27 (0.44)
BMI	30.4 (6.49)	27.4 (4.64)	26.8 (5.18)	27.0 (5.53)	29.7 (4.82)	26.1 (4.98)
Abd. perimeter	101 (17.2)	93.8 (12.3)	90.6 (13.5)	94.0 (15.8)	98.7 (12.0)	89.1 (13.1)
SBP	133 (20.1)	132 (17.9)	127 (18.0)	131 (21.0)	131 (17.5)	127 (18.5)
DBP	82.3 (12.4)	81.2 (10.5)	80.0 (10.6)	81.8 (12.6)	81.3 (10.1)	80.1 (10.9)
DM	0.33 (0.47)	0.15 (0.35)	0.12 (0.32)	0.09 (0.29)	0.28 (0.45)	0.07 (0.26)
Obesity	0.53 (0.50)	0.24 (0.43)	0.25 (0.43)	0.31 (0.46)	0.42 (0.49)	0.19 (0.39)
HT	0.44 (0.50)	0.41 (0.49)	0.24 (0.43)	0.35 (0.48)	0.45 (0.50)	0.24 (0.42)
Glucose	6.38 (2.01)	5.64 (0.86)	5.82 (9.89)	5.90 (1.29)	5.91 (1.18)	5.75 (10.1)
Creatinine	0.89 (0.26)	0.92 (0.23)	0.84 (0.22)	0.81 (0.22)	0.92 (0.24)	0.84 (0.23)
Cholesterol	5.02 (1.03)	5.42 (1.06)	5.39 (1.08)	5.18 (1.03)	5.30 (1.11)	5.42 (1.07)
LDL-C	2.90 (0.95)	3.31 (0.96)	3.30 (0.99)	3.01 (0.97)	3.23 (1.00)	3.31 (0.99)
HDL-C	1.30 (0.36)	1.50 (0.41)	1.44 (0.38)	1.40 (0.41)	1.43 (0.39)	1.47 (0.38)
Triglycerides	1.67 (0.96)	1.27 (0.80)	1.32 (0.96)	1.54 (0.93)	1.34 (0.79)	1.31 (0.98)
AST	26.1 (11.8)	27.1 (14.7)	24.6 (17.7)	34.9 (32.1)	24.1 (12.5)	25.7 (17.2)
ALT	35.3 (34.7)	24.1 (15.5)	26.1 (17.3)	39.9 (39.1)	22.3 (11.5)	27.0 (18.4)
GGT	81.7 (156)	37.2 (54.4)	53.9 (140)	171 (317)	32.7 (36.8)	55.5 (136)
Bilirubin	11.9 (5.35)	12.3 (5.33)	11.7 (5.88)	12.8 (6.29)	11.9 (5.19)	11.8 (5.90)
Protein	76.6 (5.34)	72.7 (4.53)	74.9 (5.64)	77.7 (5.51)	71.8 (4.96)	75.2 (5.25)
Albumin	43.2 (4.22)	43.4 (3.50)	43.7 (3.85)	43.5 (4.34)	42.5 (3.80)	44.1 (3.62)
Ferritin	260 (362)	176 (213)	215 (318)	324 (443)	154 (189)	226 (325)
Leucocytes	7.56 (2.15)	6.20 (1.69)	6.77 (1.90)	7.31 (2.28)	6.45 (1.81)	6.68 (1.88)
Hb	13.2 (2.43)	13.9 (1.88)	13.5 (2.04)	12.5 (2.80)	14.0 (1.71)	13.5 (2.07)
Platelets	271 (67.2)	206 (43.5)	258 (60.8)	276 (74.1)	208 (45.7)	255 (61.1)
LSM, *kPa*	11.7 (7.08)	4.76 (1.22)	6.13 (7.35)	11.9 (6.76)	4.82 (1.20)	6.17 (7.65)

Units expressed as BMI (kg/m^2), Waist Circumference (cm), SBP (mmHg), DBP (mmHg), Glucose (mmol/L), Creatinine (mg/dL), Cholesterol (mmol/L), Cholesterol_HDL (mmol/L), Triglycerides (mmol/L), AST (IU/L), ALT (IU/L), GGT (IU/L), Bilirrubin (μmol/L), Leucocytes (10^9/L), Hemoglobin (g/dL), Platelets (1,000/μL), LSM (kPa).

Note: Data are presented as mean (standard deviation)

Abd, Abdominal; ALT, alanine transaminase; AST, aspartate transaminase; BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes mellitus; FIB-4, Fibrosis-4 Index for Liver Fibrosis; GGT, gamma-glutamyl transferase; Hb, hemoglobin; HDL-C, high-density lipoprotein cholesterol; HT, hypertension; LDL-C, low-density lipoprotein cholesterol; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; SBP, systolic blood pressure.

Supplementary Table 5. Diagnostic Performance of Ethnic-/Gender-specific Waist Circumference Thresholds

	Combined population		General population		At-risk population	
	Value	95% CI	Value	95% CI	Value	95% CI
LSM ≥8 kPa						
Apparent prevalence	0.1	(0.09–0.10)	0.06	(0.05–0.06)	0.3	(0.26–0.33)
True prevalence	0.61	(0.60–0.63)	0.68	(0.67–0.70)	0.26	(0.23–0.29)
Sensitivity	0.08	(0.07–0.09)	0.07	(0.06–0.08)	0.15	(0.10–0.20)
Specificity	0.88	(0.86–0.89)	0.98	(0.97–0.98)	0.65	(0.61–0.69)
Positive predictive value	0.49	(0.45–0.54)	0.87	(0.82–0.91)	0.13	(0.09–0.18)
Negative predictive value	0.37	(0.36–0.39)	0.33	(0.31–0.34)	0.69	(0.65–0.73)
Positive likelihood ratio	0.61	(0.51–0.73)	3.09	(2.10–4.54)	0.42	(0.29–0.60)
Negative likelihood ratio	1.06	(1.03–1.08)	0.95	(0.94–0.96)	1.31	(1.20–1.42)
LSM ≥12 kPa						
Apparent prevalence	0.03	(0.03–0.04)	0.01	(0.01–0.02)	0.15	(0.12–0.17)
True prevalence	0.61	(0.60–0.63)	0.68	(0.67–0.70)	0.26	(0.23–0.29)
Sensitivity	0.02	(0.01–0.02)	0.02	(0.01–0.02)	0.03	(0.01–0.06)
Specificity	0.94	(0.93–0.95)	1	(0.99–1.00)	0.81	(0.78–0.84)
Positive predictive value	0.3	(0.23–0.37)	0.9	(0.77–0.97)	0.04	(0.01–0.10)
Negative predictive value	0.38	(0.36–0.39)	0.32	(0.31–0.33)	0.71	(0.67–0.74)
Positive likelihood ratio	0.27	(0.19–0.37)	3.99	(1.59–10.06)	0.13	(0.06–0.32)
Negative likelihood ratio	1.05	(1.04–1.06)	0.99	(0.98–0.99)	1.2	(1.15–1.26)

Note: Thresholds: Caucasic men, 94 cm; Caucasic women, 80 cm; Asian men, 90 cm; Asian women, 80 cm.
CI, Confidence interval; LSM, liver stiffness measurement.

Supplementary Table 6. Classification of Patients With Diabetes by NIT and LSM

	FIB-4 ≥1.3		FIB-4 ≥2.7	
LSM	Negative n = 337	Positive n = 305	Negative n = 600	Positive n = 42
8 kPa				
Positive	72 (21.4)	99 (32.5)	141 (23.5)	30 (71.4)
Negative	265 (78.6)	206 (67.5)	459 (76.5)	12 (28.6)
12 kPa				
Positive	28 (8.31)	53 (17.4)	58 (9.67)	23 (54.8)
Negative	309 (91.7)	252 (82.6)	542 (90.3)	19 (45.2)
	NFS ≥−1.45		NFS ≥−0.67	
LSM	Negative n = 98	Positive n = 536	Negative n = 490	Positive n = 144
8 kPa				
Positive	16 (16.3)	153 (28.5)	108 (22.0)	61 (42.4)
Negative	82 (83.7)	383 (71.5)	382 (78.0)	83 (57.6)
12 kPa				
Positive	3 (3.06)	78 (14.6)	39 (7.96)	42 (29.2)
Negative	95 (96.9)	458 (85.4)	451 (92.0)	102 (70.8)

Note: Data are presented as number (%).
FIB-4, Fibrosis-4 test; LSM, liver stiffness measurement; NFS, nonalcoholic fatty liver disease fibrosis score; NIT, noninvasive testing.