Clinical Gastroenterology and Hepatology 2022;■:■–■

Performance of Noninvasive Tests of Fibrosis Among Asians, Hispanic, and non-Hispanic Whites in the STELLAR Trials

Vincent Wai-Sun Wong,[*] Won Young Tak,[‡] George Boon Bee Goh,[§] Pin-Nan Cheng,[||] Eric J. Lawitz,[¶] Zobair M. Younossi,[#] Raj Vuppalanchi,[**] Ziad Younes,[‡‡] Naim Alkhouri,[§§] Lulu Wang,[||||] Jialuo Liu,[||||] Kathryn Kersey,[||||] Robert P. Myers,[||||] Stephen A. Harrison,[¶¶] Zachary Goodman,[#] Michael Trauner,[##] Manuel Romero-Gomez,[***] Quentin M. Anstee,[‡‡‡] Mindie H. Nguyen,[§§§] and Takeshi Okanoue[||||||]

[*]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong; [‡]School of Medicine Kyungpook National University, Kyungpook National University Hospital, Daegu, Republic of Korea; [§]Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore; [||]Department of Internal Medicine, National Cheng Kung University Hospital, Tainan, Taiwan; [¶]Texas Liver Institute, University of Texas Health San Antonio, San Antonio, Texas; [#]Inova Fairfax Hospital, Falls Church, Virginia; [**]Indiana University Medical Center, Indianapolis, Indiana; [‡‡]Gastro One, Germantown, Tennessee; [§§]Arizona Liver Health, Chandler, Arizona; [||||]Gilead Sciences, Inc, Foster City, California; [¶¶]Pinnacle Clinical Research, San Antonio, Texas; [##]Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria; [***]Hospital Universitario Virgen del Rocio, Sevilla, Spain; [‡‡‡]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, UK; Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; [§§§]Division of Gastroenterology and Hepatology, Stanford University Medical Center, Palo Alto, California; and [||||||]Saiseikai Suita Hospital, Suita City, Osaka, Japan

BACKGROUND & AIMS: **The effect of race on routinely available noninvasive tests of fibrosis is incompletely understood. This study evaluated the performance of noninvasive tests among white and Asian patients in the STELLAR trials (NCT03053050 and NCT03053063), which evaluated selonsertib in patients with advanced (F3-F4) fibrosis due to nonalcoholic steatohepatitis (NASH).**

METHODS: **Baseline liver biopsies were centrally read using the NASH Clinical Research Network system, and 4 noninvasive tests (Nonalcoholic fatty liver disease fibrosis score [NFS], Fibrosis-4 index [FIB-4], Enhanced Liver Fibrosis test [ELF], and liver stiffness by vibration-controlled transient elastography) were measured. The performance of these tests to discriminate advanced fibrosis was evaluated using areas under the receiver operating characteristics curves with 5-fold cross-validation repeated 100 times.**

RESULTS: **Among 3207 patients screened with evaluable liver histology, 2281 were whites and 762 were Asians. Seventy-two percent of whites and 67% of Asians had advanced fibrosis. The areas under the receiver operating characteristics curves of the noninvasive tests for advanced fibrosis were similar in whites and Asians: 0.73 and 0.75 for NFS, 0.78 and 0.80 for FIB-4, 0.79 and 0.81 for ELF, and 0.80 and 0.83 for liver stiffness, respectively. At the published cutoffs, the tests had similar sensitivities and specificities in the 2 groups. However, the sensitivities of NFS, FIB-4, and ELF were low in both white and Asian patients younger than 40 years.**

CONCLUSIONS: **In the global phase III STELLAR trials, the diagnostic performance of routinely available noninvasive tests for the detection of advanced fibrosis due to NASH was acceptable and similar between white and Asian patients.**

Keywords: Cirrhosis; Liver Histology; NIT.

Abbreviations used in this paper: **AUROC, area under the receiver operating characteristics curve; BMI, body mass index; CI, confidence interval; CV, cross-validation; ELF, Enhanced Liver Fibrosis panel; FIB-4, Fibrosis-4 index; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; VCTE, vibration-controlled transient elastography.**

1542-3565
https://doi.org/10.1016/j.cgh.2022.01.015

Nonalcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease, affecting at least a quarter of the global adult population.[1] NAFLD has also become increasingly common among Asians.[2] Systematic reviews have reported NAFLD in 33.9% of Asians in 2012 to 2017 and 32.9% of Chinese in 2018.[3–5] Although Asian patients with NAFLD tend to have less severe disease,[6] Asia is expected to harbor the biggest number of patients with cirrhosis and hepatocellular carcinoma secondary to nonalcoholic steatohepatitis (NASH) by 2030.[7,8]

Because only a fraction of patients with NAFLD will progress to liver-related morbidity and mortality, noninvasive tests as initial assessment are preferable.[9] Additionally, most phase IIb and III studies for NASH currently utilize histology in enrollment criteria and primary endpoints.[10] The high screen failure rate typical of these studies means patients are subject to an unnecessary invasive procedure; it also adds costs to drug development. Therefore, it is important to understand the performance of noninvasive tests not only in routine care but also in clinical trial settings.

Historically, most noninvasive tests were derived and validated in whites. There are reasons to suspect that the tests may perform differently in Asians. Asians tend to develop NAFLD and metabolic complications at a lower body mass index (BMI), and lean or nonobese NAFLD is more often described in Asians.[11–13] This may particularly affect scores that include BMI as a component (eg, the NAFLD fibrosis score).[14,15] This is highly relevant because nonobese NAFLD was reported in about 40% of the global NAFLD population.[16] Likewise, some studies showed that liver stiffness measurement by vibration-controlled transient elastography (VCTE) might be less accurate in patients with extreme BMI.[17–19] Other factors such as ethnicity and age may also impact noninvasive tests.[20,21]

With this background, we aimed to compare the accuracy of noninvasive tests in a large cohort of white and Asian patients. We also compared Hispanic and non-Hispanic whites, as well as the potential confounding effects of BMI and age on the performance of noninvasive tests.

What You Need to Know

Background

Clinicians need to use noninvasive tests to select patients with nonalcoholic fatty liver disease (NAFLD) for treatment and clinical trials. Most noninvasive tests were derived and validated in white patients.

Findings

Using data of 2281 white and 762 Asian patients being screened for the phase III STELLAR trials, we showed that the overall accuracies of the NAFLD fibrosis score, Fibrosis-4 index, Enhanced Liver Fibrosis test, and liver stiffness measurement by vibration—controlled transient elastography were similar between the 2 racial groups, as well as between Hispanic and non-Hispanic whites. The same cutoffs could also be used with similar sensitivities and specificities.

Implications for patient care

Clinicians can use these routinely available noninvasive tests in both white and Asian patients with NAFLD without adopting different cutoffs. This would simplify patient management and allow comparison of test results across studies.

Methods

Study Design Overview

This was a post hoc analysis of the screening data of the STELLAR-3 (NCT03053050) and STELLAR-4 (NCT03053063) trials evaluating the efficacy of selonsertib, an apoptosis signal-regulating kinase 1 inhibitor, in patients with NASH with bridging fibrosis (F3) and compensated cirrhosis (F4), respectively (Supplementary Methods).[22,23] The protocols conformed to ethical guidelines and were approved by the appropriate national and institutional review committees. All patients provided written informed consent.

Clinical Assessments

Investigators obtained history from and performed physical examination on the patients. Race was self-reported as white, Asian, black, American Indian, Pacific Islander, or other. BMI was calculated as body weight (kg) divided by height (m) squared. Blood was drawn after overnight fasting for liver biochemistry, metabolic parameters, and complete blood count. This study evaluated 4 noninvasive tests, namely the NAFLD fibrosis score, Fibrosis-4 index (FIB-4), Enhanced Liver Fibrosis panel (ELF), and liver stiffness measurement by VCTE (Supplementary Methods).

Liver Histology

Liver histology served as the reference standard in this study. All biopsy samples were centrally read and scored (by Z.G.) using the NASH Clinical Research Network system.[24] The primary endpoint was to identify advanced fibrosis, which included patients with bridging fibrosis (F3) and cirrhosis (F4).

Statistical Analysis

Continuous variables were expressed as median (interquartile range). The performance of noninvasive tests was evaluated using areas under the receiver

operating characteristics curves (AUROC) from univariate logistic regressions through 5-fold cross-validation (CV) repeated 100 times. Mean AUROC and 95% confidence interval (CI) are reported in the tables.[25] Specifically, we first calculate averaged AUROC over 5-fold CV ($AUROC_{CV}$) and variance of AUROC (σ^2_{CV}) in each replicate, then mean AUROC was averaged $AUROC_{CV}$ and 95% CI was based on normal approximation with pooled variance over 100 replicates. AUROC comparisons were based on the χ^2 distribution. We evaluated approaches using a single test with 1 cutoff and a single test with 2 cutoffs and calculated the corresponding sensitivities, specificities, and positive and negative predictive values. All test performance analyses were evaluated by racial subgroups. To address controversies about test confounders, we also performed subgroup analysis by ethnicity (Hispanic vs non-Hispanic), BMI (<30 and ≥30 kg/m^2) or age (<40, 40–64, and ≥65 years) within each racial subgroup. The Cochran-Mantel-Haenszel test stratified by diabetes status was used to evaluate the association between categorical variables. The Wilcoxon rank-sum test stratified by diabetes was used to compare baseline laboratory tests in Asian vs white and Hispanic vs non-Hispanic. Due to the exploratory nature of these analyses, there were no multiplicity adjustment for *P*-values.

Results

Analysis Population

A total of 4024 patients underwent screening (2155 for STELLAR-3 and 1869 for STELLAR-4), of whom 3207 patients had evaluable histology (1001 [31%] were based on historical biopsies). This included 2281 whites, 762 Asians and 164 patients of other races. Because the vast majority were whites and Asians, we focused on these 2 racial groups. Among 2281 white patients, 539 (24%) were Hispanic. Supplementary Table 1 shows the breakdown of racial groups from each recruiting country or region. The at-risk *PNPLA3* GG genotype was found in 57.6% of Hispanic whites, 25.6% of non-Hispanic whites, and 38.9% of Asians (Supplementary Figure 1).

Compared with white patients, Asians were male predominant and fewer had hypertension, and had lower BMI, higher alanine aminotransferase, aspartate aminotransferase, and bilirubin levels but lower gamma-glutamyl transferase level and platelet count (Table 1). The groups had similar age and hemoglobin A_{1c}. Overall, 76% of the entire cohort had NAFLD activity score ≥4, and 71% had advanced fibrosis (72% in whites and 67% in Asians; $P = .044$).

Noninvasive Tests in White and Asian Patients

Overall and in each racial subgroup, the results of all 4 noninvasive tests increased with fibrosis stage (Table 2, Figure 1). At each fibrosis stage, the median NAFLD fibrosis score was lower in Asian than in white patients but only statistically significant at F4 ($P = .042$); the median FIB-4 score was higher in Asian than in white patients at F2 ($P = .015$), F3 ($P < .001$), and F4 ($P < .001$); the median ELF score was higher in Asian than in white patients at F3 ($P = .002$) and F4 ($P = .023$). Median liver stiffness was similar in the 2 groups.

The AUROC of the 4 noninvasive tests for F3–F4 fibrosis in white patients were 0.73 to 0.80 and 0.75 to 0.83 for Asian patients (Table 3, Supplementary Table 2). Reliable liver stiffness measurements were obtained in 92% of whites and 81% of Asians. When dual published cutoffs were used (low cutoffs to rule out and high cutoffs to rule in F3–F4 fibrosis), the NAFLD fibrosis score, FIB-4, and ELF had similar sensitivities and specificities in white and Asian patients. Although liver stiffness had similarly high sensitivity in both white (84%) and Asian (83%) patients, the test was more specific in Asians (68% vs 78%). In addition, the positive predictive values of all 4 tests were over 90% in both racial groups. However, the negative predictive values of all 4 tests were lower in white patients than in Asian patients, with the biggest difference observed for liver stiffness (40% vs 58%).

Based on dual cutoffs, 40% to 50% of white and Asian patients were in the gray zone (indeterminate results where tests are nondiagnostic) by the NAFLD fibrosis score, FIB-4, and ELF (Table 3). By contrast, only 8% to 10% of white and Asian patients were in the gray zone by liver stiffness measurement.

Among whites, the AUROC of the 4 noninvasive tests for F3–F4 fibrosis were 0.73 to 0.87 for Hispanics and 0.72 to 0.77 for non-Hispanics (Table 3, Supplementary Table 2). Using dual cutoffs, the NAFLD fibrosis score, FIB-4, and ELF had similar sensitivities and specificities in both groups. Although liver stiffness measurement had similarly high sensitivity (84%) in both groups, the specificity was higher in Hispanics than non-Hispanics (75% vs 65%). Although the positive predictive values of all 4 tests were over 90% in both groups, the negative predictive values were particularly low in non-Hispanics, ranging from 33% to 58%.

Noninvasive Tests in White and Asian Patients With Different BMI

The test performance was similar in white and Asian patients with BMI <30 or ≥30 kg/m^2 with AUROCs ranging from 0.73 to 0.85 (Table 4, Supplementary Table 3). The sensitivities and specificities of the NAFLD fibrosis score, FIB-4, and ELF were similar in patients with different BMI in both groups. However, the specificity of liver stiffness measurement was lower in whites with BMI ≥30 kg/m^2 than those with BMI <30 kg/m^2 (61% vs 78%), but the same was not observed in Asians (86% vs 79%). As in the main analysis, the

Table 1. Clinical Characteristics of White and Asian Patients

		White					
	Hispanic (n = 539)	Non-Hispanic (n = 1734)	All white (n = 2281)	Asian (n = 762)	White and Asian (N = 3043)	White vs Asian *P* value	Hispanic vs non-Hispanic *P* value
Age, *y*	57 (50–63)	59 (52–64)	58 (52–64)	58 (49–64)	58 (51–64)	.3043	.0111
Female sex	318 (59)	983 (57)	1308 (57)	359 (47)	1667 (55)	< .0001	.2714
Body mass index, *kg/m²*	32.82 (29.1–37.23)	34.18 (30.36–38.76)	33.94 (30.04–38.48)	28.21 (25.3–31.52)	32.41 (28.26–37.14)	< .0001	2e-04
Diabetes	299 (55)	1070 (62)	1375 (60)	440 (59)	1815 (60)	.4289	.0084
Hypertension	182 (34)	875 (50)	1060 (46)	268 (35)	1328 (44)	< .0001	< .0001
ALT, *U/L*	38 (26–63.5)	43 (30–62.5)	42 (29–63)	50 (32–77)	43.5 (30–67)	< .0001	.0299
AST, *U/L*	36 (24–57)	39.33 (29.5–55.67)	38.83 (28–56)	46 (32–65.5)	40 (29–58.5)	< .0001	.0018
GGT, *U/L*	50 (32–92)	64 (38–118)	61 (36–112)	53 (34–87)	58 (35–106)	< .0001	< .0001
Bilirubin, *mg/dL*	0.5 (0.4–0.7)	0.55 (0.4–0.8)	0.55 (0.4–0.8)	0.7 (0.5–0.9)	0.6 (0.4–0.8)	< .0001	5e-04
Direct bilirubin, *mg/dL*	0.15 (0.1–0.2)	0.20 (0.1–0.2)	0.2 (0.1–0.2)	0.2 (0.1–0.2)	0.2 (0.1–0.2)	7e-04	< .0001
Platelets, $10^3/\mu L$	201 (157–256.0)	191 (144–245.5)	194 (147–249)	181 (135–227)	190 (143–243)	< .0001	.0051
Hemoglobin A_{1c}, %	6.2 (5.6–7.4)	6.2 (5.6–7.2)	6.2 (5.6–7.3)	6.2 (5.6–7.1)	6.2 (5.6–7.2)	.8439	< .0001
Histology							
Biopsy length ≥2 cm	319 (59)	994 (57)	1316 (58)	281 (37)	1597 (52)	< .0001	.4837
NAS ≥4	384 (71)	1353 (78)	1743 (76)	556 (73)	2299 (76)	.0969	.0091
Steatosis grade ≥2	35 (6)	114 (7)	150 (7)	37 (5)	187 (6)	.0946	1
Lobular inflammation grade 3	198 (37)	705 (41)	907 (40)	283 (37)	1190 (39)	.2968	.1857
Hepatocellular ballooning grade 2	314 (58)	1080 (62)	1399 (61)	456 (60)	1855 (61)	.7224	.2976
Fibrosis stage							
F0	85 (16)	86 (5)	171 (7)	52 (7)	223 (7)	.0022	< .0001
F1	57 (11)	122 (7)	180 (8)	83 (11)	263 (9)	.0022	< .0001
F2	85 (16)	194 (11)	281 (12)	113 (15)	394 (13)	.0022	< .0001
F3	137 (25)	551 (32)	691 (30)	254 (33)	945 (31)	.0022	< .0001
F4	175 (32)	781 (45)	958 (42)	260 (34)	1218 (40)	.0022	< .0001

Note: Data are presented as median (interquartile range) or number (%).

Note: The Cochran-Mantel-Haenszel test with diabetes as stratification factor was used to evaluate association between categorical variables. The Wilcoxon rank-sum test with diabetes as stratification factor was used compare baseline laboratory tests in Asian vs white and Hispanic vs non-Hispanic. Among the 3043 white and Asian patients with evaluable histology, there were 20 with missing diabetes status who were removed from the *P*-value calculation.

ALT, Alanine aminotransferase; *AST*, aspartate aminotransferase; *GGT*, gamma-glutamyl transpeptidase; *NAFLD*, nonalcoholic fatty liver disease; *NAS*, NAFLD activity score.

Table 2. Results of Noninvasive Tests in White and Asian Patients

	Hispanic (n = 539)	White		Asian (n = 762)	White and Asian (N = 3043)	White vs Asian *P* value	Hispanic vs non-Hispanic *P* value
		Non-Hispanic (n = 1734)	All white (n = 2281)				
NAFLD fibrosis score	−0.21 (−1.23 to 0.69)	0.26 (−0.72 to 1.15)	0.18 (−0.86 to 1.09)	−0.15 (−1.04 to 0.76)	0.1 (−0.92 to 0.98)	< .0001	< .0001
F0	−1.04 (−2.28 to −0.4)	−1.12 (−1.52 to −0.29)	−1.07 (−1.86 to −0.29)	−1.87 (−2.62 to −0.93)	−1.22 (−2.11 to −0.45)	.129	.1201
F1	−1.05 (−2.08 to 0.29)	−0.71 (−1.65 to 0.09)	−0.75 (−1.82 to 0.09)	−0.99 (−2.23 to −0.33)	−0.88 (−2.01 to 0.02)	.0532	.2632
F2	−0.73 (−1.77 to 0.23)	−0.47 (−1.7 to 0.39)	−0.61 (−1.73 to 0.3)	−0.98 (−1.79 to −0.1)	−0.73 (−1.74 to 0.26)	.5721	.357
F3	−0.44 (−1.18 to 0.34)	−0.06 (−1 to 0.77)	−0.11 (−1.04 to 0.69)	−0.3 (−1 to 0.43)	−0.18 (−1.01 to 0.6)	.0521	.0337
F4	0.7 (−0.13 to 1.69)	0.72 (−0.16 to 1.61)	0.72 (−0.15 to 1.63)	0.57 (−0.3 to 1.3)	0.68 (−0.18 to 1.53)	.0417	.7883
FIB-4	1.58 (1.03–2.44)	1.84 (1.23–2.78)	1.78 (1.19–2.69)	2.11 (1.33–3.31)	1.85 (1.22–2.86)	< .0001	< .0001
F0	1.03 (0.82–1.34)	1 (0.75–1.34)	1.02 (0.8–1.34)	0.97 (0.68–1.47)	1.01 (0.77–1.36)	.8487	.3447
F1	1.05 (0.8–1.49)	1.18 (0.85–1.56)	1.12 (0.82–1.56)	1.25 (0.8–1.89)	1.18 (0.82–1.59)	.2873	.2666
F2	1.24 (0.92–1.73)	1.31 (0.96–1.94)	1.29 (0.94–1.87)	1.46 (1.11–1.99)	1.37 (0.99–1.91)	.0151	.0983
F3	1.66 (1.22–2.35)	1.65 (1.23–2.41)	1.65 (1.23–2.41)	2.26 (1.47–3.08)	1.76 (1.29–2.61)	< .0001	.9003
F4	2.54 (1.82–3.83)	2.41 (1.73–3.61)	2.43 (1.75–3.66)	3.17 (2.07–4.42)	2.55 (1.8–3.84)	< .0001	.1955
ELF	9.92 (9.07–10.8)	10.11 (9.36–10.88)	10.07 (9.3–10.86)	10.13 (9.36–10.89)	10.08 (9.32–10.87)	.1306	.0037
F0	9.07 (8.39–9.51)	8.95 (8.32–9.45)	8.99 (8.37–9.5)	8.91 (8.35–9.34)	8.97 (8.37–9.48)	.5805	.5725
F1	9.23 (8.77–9.62)	9.14 (8.68–9.79)	9.15 (8.7–9.73)	9.18 (8.76–9.83)	9.16 (8.72–9.75)	.7284	.7307
F2	9.35 (8.7–9.85)	9.5 (8.94–10.15)	9.49 (8.86–10.11)	9.6 (8.94–10.18)	9.51 (8.89–10.13)	.1422	.0588
F3	10.05 (9.62–10.63)	9.93 (9.33–10.53)	9.97 (9.36–10.56)	10.18 (9.57–10.84)	10.02 (9.42–10.63)	.0022	.1227
F4	10.8 (10.17–11.41)	10.6 (10–11.3)	10.64 (10.04–11.34)	10.72 (10.21–11.68)	10.66 (10.07–11.39)	.0234	.0361
Transient elastography							
Patients with available liver stiffness by transient elastography, n	238	1006	1247	431	1678		
Liver stiffness by transient elastography, *kPa*	14.4 (8.9–20.7)	15.4 (10.5–23.9)	15.3 (10.4–23.2)	13.6 (9.4–20.8)	14.8 (10.1–22.3)	4e-04	.0346
Interquartile range-to-median ratio of liver stiffness measurement	0.13 (0.09–0.18)	0.15 (0.1–0.21)	0.14 (0.1–0.2)	0.13 (0.08–0.18)	0.14 (0.1–0.2)	< .0001	.0033
XL probe	150 (63)	683 (68)	835 (67)	109 (25)	944 (56)	< .0001	.3546

Table 2. Continued

	Hispanic (n = 539)	White		Asian (n = 762)	White and Asian (N = 3043)	White vs Asian *P* value	Hispanic vs non-Hispanic *P* value
		Non-Hispanic (n = 1734)	All white (n = 2281)				
Reliable liver stiffness measurement[a]	223 (94)	918 (91)	1143 (92)	351 (81)	1494 (89)	.232	.3116
F0	5.9 (4.55–7.2)	7.85 (6.2–11.3)	6.7 (4.9–9.2)	6.5 (4.8–8.8)	6.6 (4.9–9.2)	.8148	.0572
F1	6.8 (5.6–9.4)	9.8 (7.3–11.6)	9.1 (6.8–11.4)	9.05 (7.4–10.4)	9.05 (6.85–10.65)	.486	.0327
F2	8.94 (6.05–12.9)	10.25 (7.2–14.15)	9.8 (6.85–13.75)	8.95 (6.9–12)	9.5 (6.9–13.4)	.1767	.2243
F3	13.3 (9.9–18)	12.8 (9.7–17.6)	12.9 (9.7–17.7)	12.45 (9.6–17.1)	12.8 (9.7–17.3)	.4404	.3277
F4	20.7 (15.3–31.6)	21 (14.1–29.7)	21.05 (14.3–29.9)	21.3 (15.9–29.9)	21.25 (14.4–29.9)	.4629	.6375

Note: Data are presented as median (interquartile range) or number (%).
Note: Percentage of patients with XL probe and reliable liver stiffness measurement are based on patients with non-missing values of liver stiffness by transient elastography (kPa) in each sub-population.
ELF, Enhanced liver fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease.
[a]Reliable liver stiffness measurement was defined as 10 valid acquisitions and an interquartile range-to-median ratio of 0.3 or less among 1678 patients who underwent transient elastography.

negative predictive value of liver stiffness measurement was higher in Asians, particularly in those with BMI ≥30 kg/m^2 (51%, compared with 37% in Asians with BMI <30 kg/m^2 and 26% in whites).

Using dual cutoffs, the NAFLD fibrosis score included more patients with BMI <30 kg/m^2 in the gray zone (48%–58% among racial and BMI groups) (Table 4). Only 6% to 12% of patients were in the gray zone by liver stiffness measurement.

Noninvasive Tests in White and Asian Patients by Age

Sensitivities were low in white patients <40 years for the NAFLD fibrosis score (53%), FIB-4 (34%), and ELF (40%) (Table 5, Supplementary Table 4). In Asian patients <40 years, the sensitivities were 57%, 42%, and 50%, respectively. The sensitivities of these tests were 71% to 98% in patients aged 40 to 64 and ≥65 years. The sensitivity of VCTE was ≥80% in all age groups in both racial groups, with the highest sensitivity of 94% in Asian patients <40 years.

By contrast, the NAFLD fibrosis score, FIB-4, and ELF had high specificities of 91% to 100% in both white and Asian patients <65 years (Table 5). However, the specificity of the NAFLD fibrosis score was low at 69% in white patients aged ≥65 years, although the specificities remained high for FIB-4 (88%) and ELF (96%). Although the specificities were 81% for the NAFLD fibrosis score and 94% for ELF in Asian patients aged ≥65 years, the specificity of FIB-4 dropped to 56%.

McPherson and colleagues proposed higher cutoffs for the NAFLD fibrosis score (0.12) and FIB-4 (2.0) for patients aged ≥65 years.[21] Using these cutoffs, the sensitivity and specificity of the NAFLD fibrosis score were 71% and 55% in white patients aged ≥65 years, and 67% and 56% in Asian patients aged ≥65 years. The sensitivity and specificity of FIB-4 were 70% and 73% in white patients aged ≥65 years, and 83% and 38% in Asian patients aged ≥65 years.

Two-step Approach

Supplementary Table 5 shows the 2-step approach of different noninvasive tests, with the second test performed when the first (FIB-4 or NAFLD fibrosis score) yielded indiscriminate results. In general, this approach allowed a drastic reduction in the proportion of patients classified in the gray zone with preserved positive predictive values and modest reduction in negative predictive values in all racial groups evaluated.

Discussion

In this large international study using high-quality clinical trial data, routinely available noninvasive tests including NAFLD fibrosis score, FIB-4, ELF, and liver

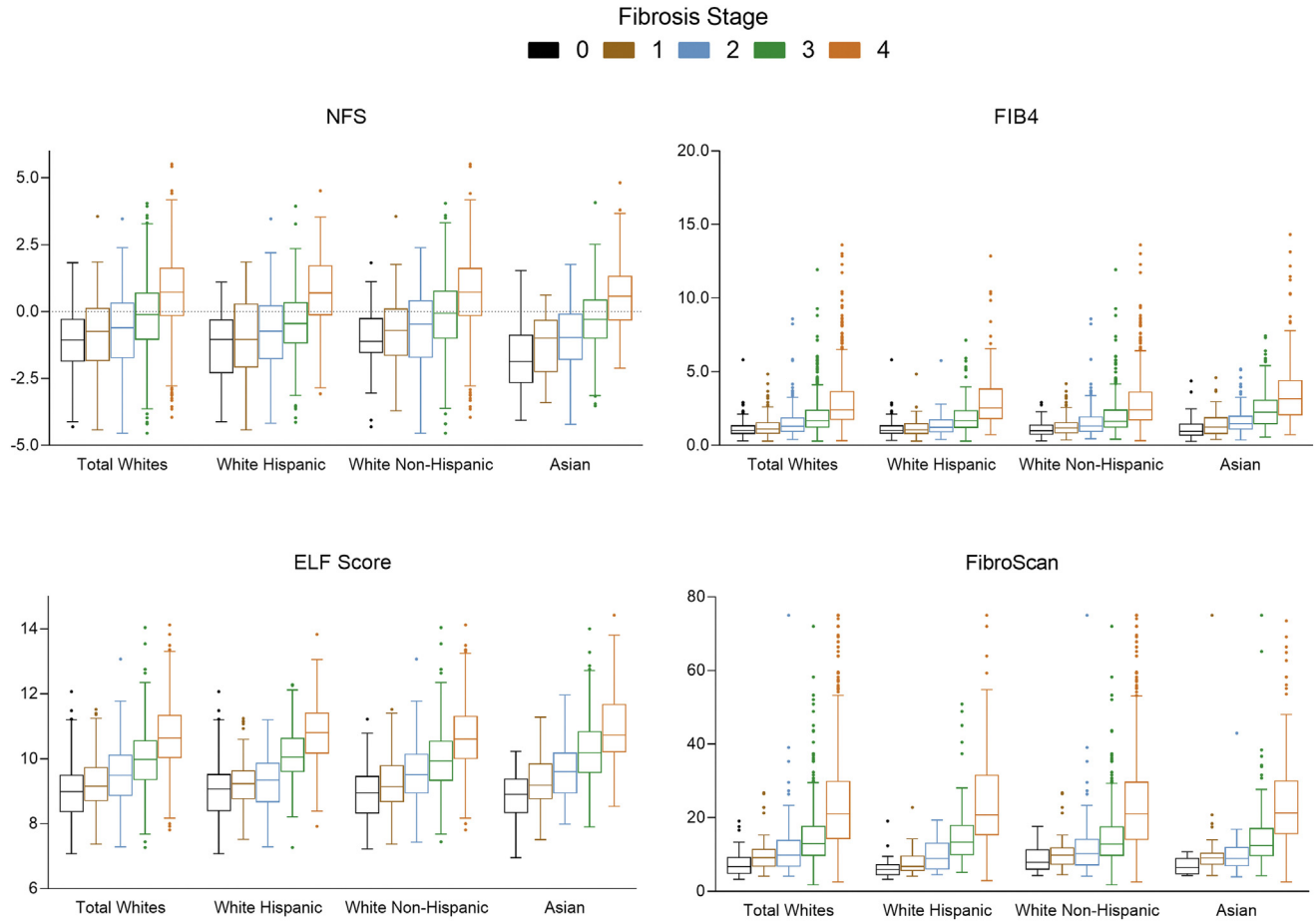


Figure 1. Distribution of noninvasive tests in white and Asian patients. *NFS*, Nonalcoholic fatty liver disease fibrosis score.

stiffness measurement had similar performance in white and Asian patients with NAFLD. Importantly, the sensitivities and specificities of the tests were reasonable when the original published cutoffs were applied to Asian patients, suggesting that currently available noninvasive tests can be used in Asians without further calibration or adjustment of cutoffs. In addition, the noninvasive tests also performed similarly in Hispanic and non-Hispanic whites.

Among the noninvasive tests evaluated in this study, VCTE is the most extensively evaluated in both white and Asian patients.[26–29] Previous studies suggested that liver stiffness measurement might lead to false-positive diagnosis of advanced fibrosis in patients with extreme BMI, likely due to technical difficulties.[17–19] Obesity also increases the risk of failure to obtain liver stiffness measurements.[30] Nonetheless, the availability of the XL probe allows reliable liver stiffness measurement in the majority of obese patients.[31] In the current study, VCTE had similarly high accuracy in whites and Asians with BMI below and above 30 kg/m^2. Compared with other noninvasive tests, VCTE classified the fewest patients in the gray zone. However, the lower specificity of VCTE in whites than in Asians, especially among whites with BMI above 30 kg/m^2, suggests that the confounding effect of obesity on liver stiffness measurement was still at play.[32]

Among the blood tests, ELF had the highest AUROC in white and Asian patients, although its superiority over FIB-4 was marginal. Although BMI is a component of the NAFLD fibrosis score, the diagnostic accuracy of this score was similar in the racial groups (Table 3). Likewise, the score performed similarly in patients with BMI below and above 30 kg/m^2. This is probably due to the relative low weighting assigned to BMI in the NAFLD fibrosis score.[14] Recently, a cross-sectional study showed that the NAFLD fibrosis score and FIB-4 might perform even better in patients with BMI below 25 kg/m^2.[33]

Because the noninvasive tests are imperfect, investigators usually propose dual cutoffs with a low cutoff to rule out significant fibrosis and a high cutoff to rule in advanced fibrosis.[9] Although it is not difficult to determine cutoffs with high sensitivities and specificities, the important issue is the proportion of patients left in the gray zone with indeterminate results. Based on published cutoffs, around 40% to 50% of white and Asian patients in this study had indeterminate results with the blood biomarkers (Table 3). Although VCTE classified much fewer patients in the gray zone (8%–10%), the tradeoff was a slightly lower overall accuracy.

McPherson and colleagues first reported lower diagnostic accuracy of the aspartate aminotransferase/alanine aminotransferase ratio, NAFLD fibrosis score,

Table 3. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3–F4) in White and Asian Patients (2-cutoff Model)

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)					
				Sensitivity	Specificity	PPV	NPV	Gray zone[a]	Accuracy
White Hispanic patients									
NAFLD fibrosis score (n = 372)	66	0.73 (0.68–0.78)	<−1.455 to ≥0.676	90 (86–94)	90 (84–95)	95 (91–97)	83 (75–89)	55 (50–60)	90 (87–93)
FIB-4 (n = 528)	58	0.81 (0.78–0.85)	<1.3 to ≥2.67	81 (76–85)	97 (94–99)	97 (94–99)	78 (73–83)	42 (37–46)	88 (84–90)
ELF (n = 539)	58	0.83 (0.79–0.86)	<9.8 to ≥11.3	78 (73–83)	99 (97–100)	99 (97–100)	77 (72–82)	41 (37–45)	87 (84–90)
Liver stiffness measurement (n = 238)	76	0.87 (0.82–0.92)	<9.9 to ≥11.4 kPa	84 (78–89)	75 (62–86)	92 (86–95)	59 (47–71)	5 (2–8)	82 (76–87)
White non-Hispanic patients									
NAFLD fibrosis score (n = 1348)	86	0.72 (0.68–0.76)	<−1.455 to ≥0.676	89 (87–91)	85 (79–90)	97 (96–98)	56 (50–62)	49 (46–52)	88 (87–90)
FIB-4 (n = 1697)	77	0.76 (0.73–0.78)	<1.3 to ≥2.67	81 (78–83)	91 (88–94)	97 (96–98)	58 (54–62)	45 (43–48)	83 (81–85)
ELF (n = 1720)	77	0.77 (0.75–0.8)	<9.8 to ≥11.3	71 (69–74)	97 (95–99)	99 (98–99)	50 (47–54)	47 (45–50)	77 (75–79)
Liver stiffness measurement (n = 1006)	89	0.76 (0.72–0.81)	<9.9 to ≥11.4 kPa	84 (82–87)	65 (55–74)	95 (93–97)	33 (27–40)	8 (7–10)	82 (80–84)
All white patients									
NAFLD fibrosis score (n = 1725)	82	0.73 (0.7–0.76)	<−1.455 to ≥0.676	89 (87–91)	87 (83–91)	97 (96–98)	65 (60–69)	50 (48–53)	89 (87–90)
FIB-4 (n = 2233)	73	0.78 (0.76–0.8)	<1.3 to ≥2.67	81 (79–83)	93 (91–95)	97 (96–98)	64 (61–68)	44 (42–47)	84 (83–86)
ELF (n = 2267)	72	0.79 (0.77–0.81)	<9.8 to ≥11.3	73 (70–75)	98 (96–99)	99 (98–99)	58 (55–61)	46 (44–48)	80 (78–81)
Liver stiffness measurement (n = 1247)	87	0.8 (0.76–0.84)	<9.9 to ≥11.4 kPa	84 (82–86)	68 (61–75)	95 (93–96)	40 (34–46)	8 (6–9)	82 (80–84)
Asian patients									
NAFLD fibrosis score (n = 586)	78	0.75 (0.7–0.8)	<−1.455 to ≥0.676	88 (85–91)	92 (86–96)	98 (96–99)	68 (61–75)	54 (50–58)	89 (86–91)
FIB-4 (n = 735)	69	0.8 (0.76–0.83)	<1.3 to ≥2.67	88 (85–90)	90 (85–93)	95 (92–97)	77 (71–82)	40 (37–44)	88 (86–91)
ELF (n = 748)	68	0.81 (0.77–0.84)	<9.8 to ≥11.3	76 (72–80)	96 (93–98)	98 (96–99)	66 (60–70)	44 (40–47)	83 (80–85)
Liver stiffness measurement (n = 431)	76	0.83 (0.78–0.87)	<9.9 to ≥11.4 kPa	83 (78–87)	78 (68–85)	92 (88–95)	58 (50–67)	10 (7–13)	81 (77–85)

AUROC, Area under the receiver operating characteristics curve; *CI*, confidence interval; *ELF*, enhanced liver fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

[a]"Gray zone" introduced by utilizing dual cutoffs is defined as indeterminate results where the tests are nondiagnostic.

Table 4. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3-F4) in White and Asian Patients by BMI (2-cutoff Model)

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)					
				Sensitivity	Specificity	PPV	NPV	Gray zone[a]	Accuracy
White patients, BMI <30 kg/m^2									
NAFLD fibrosis score (n = 428)	81	0.76 (0.71–0.81)	<−1.455 to ≥0.676	80 (76–84)	99 (93–100)	100 (98–100)	55 (46–63)	58 (53–62)	84 (80–87)
FIB-4 (n = 429)	81	0.75 (0.69–0.8)	<1.3 to ≥2.67	82 (78–86)	87 (78–93)	96 (93–98)	54 (45–63)	44 (39–49)	83 (79–87)
ELF (n = 436)	80	0.80 (0.75–0.85)	<9.8 to ≥11.3	70 (65–75)	100 (96–100)	100 (98–100)	45 (38–53)	48 (43–53)	76 (72–80)
Liver stiffness measurement (n = 263)	91	0.79 (0.68–0.89)	<9.9 to ≥11.4 kPa	78 (73–83)	78 (56–93)	97 (94–99)	26 (16–38)	12 (8–16)	78 (73–83)
White patients, BMI ≥30 kg/m^2									
NAFLD fibrosis score (n = 1297)	82	0.73 (0.69–0.76)	<−1.455 to ≥0.676	92 (90–94)	83 (78–88)	96 (95–97)	70 (64–75)	48 (45–50)	91 (89–92)
FIB-4 (n = 1300)	82	0.78 (0.75–0.81)	<1.3 to ≥2.67	80 (77–82)	94 (90–97)	98 (97–99)	50 (45–55)	46 (44–49)	82 (80–84)
ELF (n = 1310)	82	0.8 (0.76–0.83)	<9.8 to ≥11.3	73 (70–76)	97 (95–99)	99 (98–100)	45 (40–49)	49 (46–52)	78 (75–80)
Liver stiffness measurement (n = 846)	93	0.77 (0.71–0.83)	<9.9 to ≥11.4 kPa	86 (83–88)	61 (48–73)	97 (95–98)	26 (19–33)	6 (5–8)	84 (81–86)
Asian patients, BMI <30 kg/m^2									
NAFLD fibrosis score (n = 376)	78	0.75 (0.69–0.81)	<−1.455 to ≥0.676	87 (83–91)	93 (85–97)	98 (95–99)	68 (58–76)	57 (51–62)	89 (85–92)
FIB-4 (n = 377)	78	0.75 (0.69–0.81)	<1.3 to ≥2.67	93 (89–95)	77 (67–86)	93 (90–96)	74 (64–83)	37 (32–42)	89 (86–92)
ELF (n = 381)	77	0.78 (0.72–0.83)	<9.8 to ≥11.3	78 (73–83)	97 (90–99)	99 (96–100)	56 (48–64)	49 (44–54)	82 (78–86)
Liver stiffness measurement (n = 220)	87	0.8 (0.72–0.88)	<9.9 to ≥11.4 kPa	80 (74–86)	79 (59–92)	96 (92–99)	37 (25–50)	10 (7–15)	80 (74–85)
Asian patients, BMI ≥30 kg/m^2									
NAFLD fibrosis score (n = 210)	79	0.75 (0.68–0.83)	<−1.455 to ≥0.676	90 (84–94)	91 (78–97)	97 (93–99)	70 (57–82)	50 (43–56)	90 (85–94)
FIB-4 (n = 210)	79	0.75 (0.68–0.83)	<1.3 to ≥2.67	78 (71–84)	93 (81–99)	98 (94–100)	53 (42–65)	45 (38–52)	81 (75–86)
ELF (n = 211)	79	0.78 (0.71–0.86)	<9.8 to ≥11.3	71 (64–78)	93 (81–99)	98 (93–99)	46 (35–57)	44 (37–51)	76 (69–81)
Liver stiffness measurement (n = 137)	85	0.85 (0.76–0.94)	<9.9 to ≥11.4 kPa	85 (78–91)	86 (64–97)	97 (92–99)	51 (34–69)	10 (6–17)	85 (78–91)

AUROC, Area under the receiver operating characteristics curve; *BMI*, body mass index; *CI*, confidence interval; *ELF*, enhanced liver fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

[a]"Gray zone" introduced by utilizing dual cutoffs is defined as indeterminate results where the tests are non-diagnostic.

Table 5. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3–F4) in White and Asian Patients by Age (2-cutoff Model)

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)					
				Sensitivity	Specificity	PPV	NPV	Gray zone[a]	Accuracy
White patients, age <40 years									
NAFLD fibrosis score (n = 68)	69	0.7 (0.57–0.82)	<−1.455 to ≥0.676	53 (38–68)	95 (76–100)	96 (80–100)	48 (32–64)	34 (23–46)	66 (54–77)
FIB-4 (n = 105)	53	0.71 (0.61–0.81)	<1.3 to ≥2.67	34 (22–48)	100 (93–100)	100 (82–100)	57 (46–68)	18 (11–27)	65 (55–74)
ELF (n = 111)	51	0.75 (0.66–0.84)	<9.8 to ≥11.3	40 (28–54)	100 (93–100)	100 (85–100)	61 (50–72)	23 (16–32)	69 (60–78)
Liver stiffness measurement (n = 51)	78	0.71 (0.55–0.86)	<9.9 to ≥11.4 kPa	80 (64–91)	64 (31–89)	89 (74–97)	47 (21–73)	6 (1–16)	76 (63–87)
White patients, age 40–64 years									
NAFLD fibrosis score (n = 1261)	81	0.74 (0.7–0.77)	<−1.455 to ≥0.676	89 (86–90)	91 (87–94)	98 (96–99)	65 (60–70)	52 (49–55)	89 (87–91)
FIB-4 (n = 1637)	72	0.77 (0.75–0.8)	<1.3 to ≥2.67	79 (77–81)	94 (91–96)	97 (96–98)	64 (60–67)	45 (43–47)	83 (81–85)
ELF (n = 1655)	72	0.79 (0.76–0.81)	<9.8 to ≥11.3	71 (69–74)	98 (96–99)	99 (98–100)	58 (54–61)	45 (43–48)	79 (77–81)
Liver stiffness measurement (n = 923)	86	0.82 (0.78–0.86)	<9.9 to ≥11.4 kPa	85 (82–87)	70 (62–78)	95 (93–96)	44 (37–51)	8 (6–10)	83 (80–85)
White patients, age ≥65 years									
NAFLD fibrosis score (n = 396)	86	0.68 (0.6–0.75)	<−1.455 to ≥0.676	96 (94–98)	69 (55–81)	95 (92–97)	75 (60–86)	47 (42–52)	92 (89–95)
FIB-4 (n = 491)	80	0.78 (0.73–0.83)	<1.3 to ≥2.67	92 (89–94)	88 (79–93)	97 (94–98)	73 (64–80)	48 (44–53)	91 (88–93)
ELF (n = 501)	80	0.77 (0.72–0.82)	<9.8 to ≥11.3	81 (77–85)	96 (90–99)	99 (97–100)	57 (49–64)	53 (49–58)	84 (81–87)
Liver stiffness measurement (n = 273)	92	0.72 (0.62–0.83)	<9.9 to ≥11.4 kPa	83 (77–87)	57 (34–78)	96 (92–98)	21 (12–34)	7 (4–11)	81 (75–85)
Asian patients, age <40 years									
NAFLD fibrosis score (n = 39)	59	0.83 (0.69–0.97)	<−1.455 to ≥0.676	57 (34–77)	94 (70–100)	93 (66–100)	60 (39–79)	28 (15–45)	72 (55–85)
FIB-4 (n = 64)	38	0.76 (0.64–0.88)	<1.3 to ≥2.67	42 (22–63)	100 (91–100)	100 (69–100)	74 (60–85)	19 (10–30)	78 (66–87)
ELF (n = 66)	36	0.76 (0.64–0.88)	<9.8 to ≥11.3	50 (29–71)	100 (92–100)	100 (74–100)	78 (64–88)	23 (13–35)	82 (70–90)
Liver stiffness measurement (n = 35)	51	0.82 (0.69–0.96)	<9.9 to ≥11.4 kPa	94 (73–100)	65 (38–86)	74 (52–90)	92 (62–100)	17 (7–34)	80 (63–92)
Asian patients, age 40-64 years									
NAFLD fibrosis score (n = 392)	79	0.76 (0.71–0.82)	<−1.455 to ≥0.676	87 (82–90)	95 (88–99)	99 (96–100)	66 (57–74)	58 (53–63)	89 (85–92)
FIB-4 (n = 489)	68	0.8 (0.76–0.84)	<1.3 to ≥2.67	87 (83–91)	94 (89–97)	97 (94–99)	77 (71–83)	45 (41–50)	89 (86–92)
ELF (n = 500)	67	0.8 (0.76–0.84)	<9.8 to ≥11.3	74 (69–78)	96 (91–98)	97 (94–99)	64 (57–70)	43 (38–47)	81 (77–84)

Table 5. Continued

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)					
				Sensitivity	Specificity	PPV	NPV	Gray zone[a]	Accuracy
Liver stiffness measurement (n = 291)	75	0.85 (0.79–0.9)	<9.9 to ≥11.4 kPa	82 (76–87)	82 (71–90)	93 (89–96)	60 (49–69)	10 (7–14)	82 (77–86)
Asian patients, age ≥65 years									
NAFLD fibrosis score (n = 155)	83	0.68 (0.57–0.79)	<−1.455 to ≥0.676	98 (93–100)	81 (62–94)	96 (91–99)	88 (69–97)	52 (43–60)	95 (90–98)
FIB-4 (n = 182)	81	0.64 (0.54–0.74)	<1.3 to ≥2.67	97 (92–99)	56 (38–73)	91 (85–95)	79 (58–93)	35 (28–42)	89 (84–93)
ELF (n = 182)	81	0.76 (0.67–0.84)	<9.8 to ≥11.3	86 (80–92)	94 (80–99)	98 (95–100)	62 (47–75)	54 (47–62)	88 (82–92)
Liver stiffness measurement (n = 105)	87	0.78 (0.68–0.89)	<9.9 to ≥11.4 kPa	82 (73–90)	71 (42–92)	95 (88–99)	38 (20–59)	8 (3–14)	81 (72–88)

AUROC, Area under the receiver operating characteristics curve; *CI*, confidence interval; *ELF*, enhanced liver fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

[a]"Gray zone" introduced by utilizing dual cutoffs is defined as indeterminate results where the tests are nondiagnostic.

and FIB-4 in patients aged below 35 years, and low specificity of the NAFLD fibrosis score and FIB-4 in those aged above 65 years.[21] In the current study, although the AUROCs of the noninvasive tests were not as low as what was reported by McPherson, NAFLD fibrosis score, FIB-4, and ELF had rather low sensitivities and negative predictive values among patients younger than 40 years. Although the low sensitivities can be due to their inclusion of age in the calculation, this cannot explain the similar observation for ELF. Among apparently healthy individuals, ELF is higher in men and older subjects.[34] Although ELF considers specific biomarkers for liver fibrosis, it is noteworthy that type III collagen is also present in other organs.[32] Amino-terminal propeptide of procollagen type III may also be elevated in bone fracture, other fibrotic disease, and kidney disease. Tissue inhibitor of matrix metalloproteinases-1 is elevated in cancer and systemic inflammation. Because these inflammatory and fibrotic conditions are likely to be less prominent in younger patients, using the same ELF cutoffs in this population would result in reduced sensitivity.

Conversely, we confirmed the low specificity of NAFLD fibrosis score and FIB-4 in patients >65 years. Using the original low cutoffs (−1.455 for NAFLD fibrosis score and 1.3 for FIB-4), the specificities for advanced fibrosis were 7% and 35% in whites, and 19% and 9% in Asians, respectively (Supplementary Table 4). The use of age-specific cutoffs would reduce the number of patients unnecessarily selected for further evaluation.

In a recent individual patient data meta-analysis of 37 studies from Europe and the Middle East and Asia-Pacific regions, Mózes et al reported a pooled AUROC for advanced fibrosis of 0.85 for VCTE, 0.76 for FIB-4, and 0.73 for NAFLD fibrosis score.[35] The current study adds to the existing literature by using a central pathologist, a central laboratory for blood tests, and clinical trial data as well as direct comparison of the test performance in different racial groups.

Our study has the strengths of a multicenter design, high-quality clinical trial data, large sample size, and the use of a central expert pathologist. Nevertheless, it also has a few limitations. First, because of the inclusion criteria of the STELLAR trials, patients had a high pretest probability of advanced fibrosis, as reflected by the high positive predictive values (>90%) of all 4 noninvasive tests. Second, liver histology is an imperfect reference standard with inherent sampling, intraobserver, and interobserver variability. In recent years, artificial intelligence-based assessments of histology have demonstrated good reliability and accuracy, and may be used in the future for the evaluation of noninvasive tests.[36] Finally, although this study included a large number of white and Asian patients and compared between Hispanic and non-Hispanic whites, future studies are needed to evaluate the performance of noninvasive tests in the smaller subgroups and populations not covered in this study (notably blacks).

Conclusions

In conclusion, in the large global phase III STELLAR trials, the diagnostic performance of routinely available noninvasive tests for the detection of advanced fibrosis due to NASH was acceptable and similar between white and Asian patients.

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 http://doi.org/10.1016/j.cgh.2022.01.015

References

1. Younossi Z, Tacke F, Arrese M, et al. Global perspectives on nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Hepatology 2019;69:2672–2682.
2. Fan JG, Kim SU, Wong VW. New trends on obesity and NAFLD in Asia. J Hepatol 2017;67:862–873.
3. Li J, Zou B, Yeo YH, et al. Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999-2019: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2019;4:389–398.
4. Zhou J, Zhou F, Wang W, et al. Epidemiological features of NAFLD from 1999 to 2018 in China. Hepatology 2020; 71:1851–1864.
5. Wu Y, Zheng Q, Zou B, et al. The epidemiology of NAFLD in Mainland China with analysis by adjusted gross regional domestic product: a meta-analysis. Hepatol Int 2020;14:259–269.
6. Lee HW, Wong VW. Changing NAFLD epidemiology in China. Hepatology 2019;70:1095–1098.
7. Estes C, Anstee QM, Arias-Loste MT, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. J Hepatol 2018;69:896–904.
8. Estes C, Chan HLY, Chien RN, et al. Modelling NAFLD disease burden in four Asian regions-2019-2030. Aliment Pharmacol Ther 2020;51:801–811.
9. Wong VW, Adams LA, de Ledinghen V, et al. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol 2018;15:461–478.
10. Wong VW, Chitturi S, Wong GL, et al. Pathogenesis and novel treatment options for non-alcoholic steatohepatitis. Lancet Gastroenterol Hepatol 2016;1:56–67.
11. Wei JL, Leung JC, Loong TC, et al. Prevalence and severity of nonalcoholic fatty liver disease in non-obese patients: a population study using proton-magnetic resonance spectroscopy. Am J Gastroenterol 2015;110:1306–1314; quiz: 1315.
12. Leung JC, Loong TC, Wei JL, et al. Histological severity and clinical outcomes of nonalcoholic fatty liver disease in nonobese patients. Hepatology 2017;65:54–64.
13. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363:157–163.
14. 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.
15. Wong VW, Wong GL, Chim AM, et al. Validation of the NAFLD fibrosis score in a Chinese population with low prevalence of advanced fibrosis. Am J Gastroenterol 2008;103:1682–1688.
16. Ye Q, Zou B, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2020;5:739–752.
17. Das K, Sarkar R, Ahmed SM, et al. "Normal" liver stiffness measure (LSM) values are higher in both lean and obese individuals: a population-based study from a developing country. Hepatology 2012;55:584–593.
18. Wong GL, Chan HL, Choi PC, et al. Association between anthropometric parameters and measurements of liver stiffness by transient elastography. Clin Gastroenterol Hepatol 2013; 11:295–302.e1–3.
19. Petta S, Wai-Sun Wong V, Bugianesi E, et al. Impact of obesity and alanine aminotransferase levels on the diagnostic accuracy for advanced liver fibrosis of noninvasive tools in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2019; 114:916–928.
20. Kallwitz ER, Daviglus ML, Allison MA, et al. Prevalence of suspected nonalcoholic fatty liver disease in Hispanic/Latino individuals differs by heritage. Clin Gastroenterol Hepatol 2015; 13:569–576.
21. McPherson S, Hardy T, Dufour JF, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol 2017;112:740–751.
22. Harrison SA, Wong VW, Okanoue T, et al. STELLAR-3 and STELLAR-4 Investigators. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: results from randomized phase III STELLAR trials. J Hepatol 2020;73:26–39.
23. 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.
24. Kleiner DE, Brunt EM, Van Natta M, et al. , Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313–1321.
25. LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. Electron J Stat 2015;9:1583–1607.
26. Kwok R, Tse YK, Wong GL, et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease–the role of transient elastography and plasma cytokeratin-18 fragments. Aliment Pharmacol Ther 2014;39:254–269.
27. Cassinotto C, Boursier J, de Ledinghen V, et al. Liver stiffness in nonalcoholic fatty liver disease: a comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. Hepatology 2016;63:1817–1827.
28. Tapper EB, Challies T, Nasser I, et al. The performance of vibration controlled transient elastography in a US cohort of patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2016;111:677–684.
29. 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.
30. Wong GL, Wong VW, Chim AM, et al. Factors associated with unreliable liver stiffness measurement and its failure with transient elastography in the Chinese population. J Gastroenterol Hepatol 2011;26:300–305.

31. Wong VW, Vergniol J, Wong GL, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862–1871.
32. Wai JW, Fu C, Wong VW. Confounding factors of non-invasive tests for nonalcoholic fatty liver disease. J Gastroenterol 2020; 55:731–741.
33. Fu C, Wai JW, Nik Mustapha NR, et al. Performance of simple fibrosis scores in nonobese patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2020;18:2843–2845.e2.
34. Lichtinghagen R, Pietsch D, Bantel H, et al. The Enhanced Liver Fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. J Hepatol 2013;59:236–242.
35. Mozes FE, Lee JA, Selvaraj EA, et al. LITMUS Investigators. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. Gut 2021 May 17: Online ahead of print.
36. Wong GL, Yuen PC, Ma AJ, et al. Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. J Gastroenterol Hepatol 2021;36:543–550.

Reprint requests
Address requests for reprints to: Vincent Wong, MD, Department of Medicine and Therapeutics, 9/F, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, Hong Kong. e-mail: wongv@cuhk.edu.hk; tel: +852 35051205; fax: +852 26373852.

Acknowledgement
The authors thank Dr. Bryan Downie, PhD, of Gilead Sciences, for his assistance in the genetic analysis.

CRediT Authorship Contributions
Vincent Wai-Sun Wong, M.D. (Conceptualization: Lead; Investigation: Equal; Methodology: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)
Won Young Tak (Investigation: Equal; Writing – review & editing: Equal)
George Boon Bee Goh (Investigation: Equal; Writing – review & editing: Equal)
Pin-Nan Cheng (Investigation: Equal; Writing – review & editing: Equal)
Eric J. Lawitz (Investigation: Equal; Writing – review & editing: Equal)
Zobair M. Younossi (Investigation: Equal; Writing – review & editing: Equal)
Raj Vuppalanchi (Investigation: Equal; Writing – review & editing: Equal)
Ziad Younes (Investigation: Equal; Writing – review & editing: Equal)
Naim Alkhouri (Investigation: Equal; Writing – review & editing: Equal)
Lulu Wang (Formal analysis: Equal; Methodology: Lead; Writing – review & editing: Equal)
Jiaoluo Liu (Formal analysis: Lead; Methodology: Lead; Writing – review & editing: Equal)
Kathryn Kersey (Conceptualization: Equal; Writing – review & editing: Equal)
Robert P. Myers (Conceptualization: Lead; Writing – review & editing: Equal)
Stephen A. Harrison (Investigation: Equal; Writing – review & editing: Equal)
Zachary Goodman (Data curation: Lead; Writing – review & editing: Equal)
Michael Trauner (Investigation: Equal; Writing – review & editing: Equal)
Manuel Romero-Gomez (Investigation: Equal; Writing – review & editing: Equal)
Quentin M. Anstee (Investigation: Equal; Writing – review & editing: Equal)
Mindie H. Nguyen (Investigation: Equal; Writing – review & editing: Equal)
Takeshi Okanoue (Investigation: Equal; Writing – review & editing: Equal)
This paper was presented in part at the Annual Meeting of the American Association for the Study of Liver Diseases held in Boston, Massachusetts, on November 8–12, 2019.

Conflicts of interest
These authors disclose the following: Vincent Wong has served as a consultant or advisory board member for 3V-BIO, AbbVie, Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens, Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk, Perspectum Diagnostics, Pfizer, ProSciento, Sagimet Biosciences, and Terns; is a speaker for AbbVie, Bristol-Myers Squibb, Echosens, and Gilead Sciences; and received an unrestricted grant from Gilead Sciences for fatty liver research. George Boon Bee Goh has served as consultant or advisory board member for Boehringer Ingelheim, Gilead Sciences, and Ionis Pharmaceuticals; and as speaker for Echosens and Bayer. Eric J. Lawitz has received research grants from 89Bio, Allergan Inc, Akero Therapeutics, Astrazeneca, Axcella Health, Bristol-Myers Squibb, Boehringer Ingelheim, Celgene, Durect Corporation, Eli Lilly and Company, Enanta Pharmaceuticals, Enyo, Galmed Pharmaceuticals, Genfit, Gilead Sciences, Hanmi Pharmaceuticals, Intercept Pharmaceuticals, Madrigal Pharmaceuticals, Merck, Metacrine, Novartis, Novo Nordisk, Octeta Therapeutics, Poxel, Roche Pharmaceuticals, Viking, and Zydus; and has served as a speaker for Intercept. Zobair M. Younossi has received research funds and/or consultant fees from Abbott, Allergan, Bristol-Myers Squibb, Echosens, Genfit, Gilead Sciences, Intercept, Madrigal, Merck, Novartis, Novo Nordisk, Quest Diagnostics, Siemens, Shionogi, Terns, and Viking. Raj Vuppalanchi received consulting fees for serving on the Data Safety Monitoring Boards for Covance, Enyio, and Enanta; has received research grant support from Gilead Sciences, Zydus Discovery, Cara Therapeutics, Novo Nordisk, and Intercept, where his institution gets the funding. Ziad Younes has received research support from Gilead, Intercept, BMS, Madrigal, Cymabay, BMS, Axcella, HighTide, NST, NovoNordisk, Novartis, and Zydus; has served as a consultant or on advisory boards for Gilead and Intercept; and has served as a speaker for Gilead, Intercept, and AbbVie. Naim Alkhouri has served on advisory boards for Allergan, Gilead, Intercept, and Pfizer; has served as a speaker for AbbVie, Alexion, Gilead, Intercept, and Simply Speaking; and has received research support from Akero, Allergan, Axcella, BMS, Gilead, Galectin, Genfit, Enanta, Enyo, Inventiva, Madrigal, Merck, NGM, Novartis, Novo Nordisk, Pfizer, Poxel, and Zydus. Lulu Wang, Jialuo Liu, Kathryn Kersey, and Robert P. Myers are employees of and own stock in Gilead Sciences. Stephen A. Harrison has served on advisory boards for Akero, AltImmune, Arrowhead, Axcella, Blade Therapeutics, Cirius, Civi, CLDF, CymaBay, Echosens, Foresite, Galectin, Galmed, Gelesis, Genfit, Gilead, Hepion, Hightide, Hisoindex, Indalo, Innovate, Intercept, Madrigal, MedPace, MSD, Metcrine, NGM, Northsea, Novartis, Novo Nordisk, Perspectum, Poxel, Prometic, Ridgeline, Sagimet, Terns, and Viking; has served as a consultant for Akero, AltImmune, Axcella, Blade, BVF Partners, Cirius, Civi, CLDF, CymaBay, Echosens, Enyo, Foresite, Fortress, Galectin, Galmed, Gelesis, Genfit, Gilead, Hepion, Hightide, Hisoindex, Indalo, Inipharm, Innovate, Intercept, Kowa, Madrigal, MedPace, MSD, Metcrine, NGM, Northsea, Novartis, Novo Nordisk, Perspectum, Poxel, Prometic, Ridgeline, Sagimet, Silverback, Terns, and Viking; has served as the principal investigator of research grants for Axcella, BMS, Cirius, Clvi, Conatus, Cymabay, Enyo, Galectin, Galmed, Genentech, Genfit, Gilead, Hepion, Hightide, Immuron, Intercept, Madrigal, NGM, Northsea, Novartis, Novo Nordisk, Pfizer, Sagimet, Second Genome, Tobira/Allergan, and Viking; and owns stock in Akero, Cirius, Galectin, Genfit, Histoindex, Madrigal, Metacrine, NGM, and Northsea. Zachary Goodman has received research grant funding from Gilead Sciences, Allergan, Novartis, Bristol-Myers Squibb, Intercept, Genfit, and NGM Biopharmaceuticals. Michael Trauner has received speaker fees from Bristol-Myers Squibb, Falk Foundation, Gilead, Intercept, and Merck Sharp & Dohme; has received advisory board fees from Albireo, Boehringer Ingelheim, BiomX, Falk Pharma GmbH, GENFIT, Gilead, Intercept, MSD, Novartis, Phenex, and Regulus; has received travel grants from AbbVie, Falk, Gilead, and Intercept; has received research grants from Albireo, CymaBay, Falk, Gilead, Intercept, MSD, and Takeda; and is also coinventor of patents on the medical use of norUDCA filed by the Medical University of Graz. Manuel Romero-Gomez has received research grants from Intercept and Gilead; has served as a consultant for Shionogi, Prosciento, Kaleido, MSD, BMS, Allergan, Boehringer-Ingelheim, Zydus, Intercept, and Gilead; and has served as a speaker for Alfa-Wasserman, Novonordisk, Intercept, and Gilead. Quentin M. Anstee has received research grant funding from Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex; has active research collaborations (including research collaborations supported through the EU IMI2 LITMUS Consortium*) with Abbvie, Antaros Medical,* Allergan/Tobira,* AstraZeneca,* BMS,* Boehringer Ingelheim International GMBH,* Echosens,* Ellegaard Gottingen Minipigs AS,* Eli Lilly & Company Ltd,* Exalenz Bioscience Ltd,* Genfit SA,* Glympse Bio, GlaxoSmithKline, HistoIndex,* Intercept Pharma Europe Ltd,* iXscient Ltd,* Nordic Bioscience,* Novartis Pharma AG,* Novo Nordisk A/S,* One Way Liver Genomics SL,* Perspectum Diagnostics,* Pfizer Ltd,* Resoundant,* Sanofi-Aventis Deutschland GMBH,* SomaLogic Inc,* and Takeda Pharmaceuticals International SA*; performs consultancy for 89Bio, Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, BMS, BNN Cardio, Celgene, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd, Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd, Inventiva, IQVIA, Janssen, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd, Poxel, ProSciento, Raptor Pharma, Servier, Terns, and Viking Therapeutics; is a speaker for Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, and MedScape; and receives royalties from Elsevier Ltd. Mindie H. Nguyen has received research grants from Enanta, Gilead Sciences, and Pfizer; and has served as a consultant or advisory board member for Gilead Sciences and Intercept. The remaining authors disclose no conflicts.

Funding
This study was funded by Gilead Sciences, Inc.

Supplementary Methods

Patients

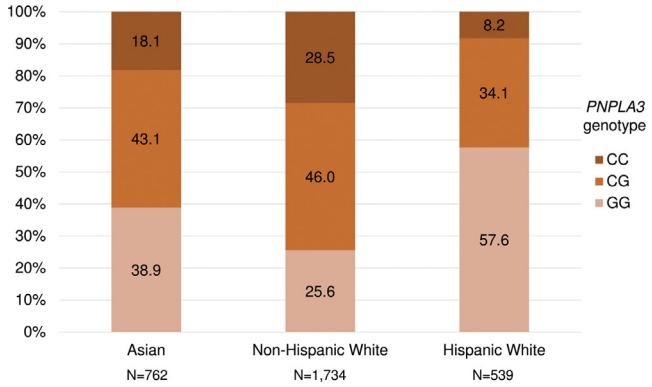
The STELLAR trials included patients aged 18 to 70 years from North and South America, Europe, Australia, New Zealand, and Asia.[1] Liver biopsies were performed at screening or ≤6 months of screening for STELLAR-3 and ≤12 months of screening for STELLAR-4. We excluded patients with other liver diseases (eg, chronic hepatitis B, alcohol-related liver disease, and autoimmune liver disease), platelet count $<100 \times 10^3/\mu L$ or a history of liver transplantation, hepatic decompensation, or hepatocellular carcinoma.

Noninvasive Tests

The nonalcoholic fatty liver disease fibrosis score was calculated as $-1.675 + 0.037 \times$ age (years) $+ 0.094 \times$ body mass index (kg/m^2) $+ 1.13 \times$ impaired fasting glucose or diabetes (yes = 1; no = 0) $+ 0.99 \times$ aspartate aminotransferase-to-alanine aminotransferase ratio – $0.013 \times$ platelet ($\times 10^9/L$) – $0.66 \times$ albumin (g/dL).[2] Fibrosis-4 index was calculated as age (years) × aspartate aminotransferase (U/L) / platelet ($\times 10^9/L$) /√ alanine aminotransferase (U/L).[3] The Enhanced Liver Fibrosis panel (Siemens Healthcare GmbH, Eriangen, Germany) was a proprietary formula to predict fibrosis based on 3 specific fibrosis biomarkers: hyaluronic acid, tissue inhibitor of matrix metalloproteinases-1 and amino-terminal propeptide of procollagen type III.[4] Liver stiffness was measured using vibration-controlled transient elastography (FibroScan, Echosens, Paris, France) by experienced and trained operators. The choice of M and XL probes was in accordance to the machine's probe selection tool, and measurements by XL probe were prioritized for analysis if available.[5,6] The cutoffs of the noninvasive tests were adopted from published reports.

Supplementary References

1. Harrison SA, Wong VW, Okanoue T, et al. STELLAR-3 and STELLAR-4 Investigators.. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: results from randomized phase III STELLAR trials. J Hepatol 2020;73:26–39.
2. 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.
3. 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.
4. Lichtinghagen R, Pietsch D, Bantel H, et al. The Enhanced Liver Fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. J Hepatol 2013;59:236–242.
5. Tapper EB, Challies T, Nasser I, et al. The performance of vibration controlled transient elastography in a US cohort of patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2016;111:677–684.
6. Wong VW, Irles M, Wong GL, et al. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. Gut 2019;68:2057–2064.



Supplementary Figure 1. PNPLA3 rs738409 genotype frequencies in the STELLAR population.

Supplementary Table 1. Countries of Origin and Reported Ethnicities of the Included Patients

Country/region	Asian, n	White (total), n[a]	Hispanic white, n	Non-Hispanic white, n
Argentina	0	13	13	0
Australia	8	84	0	81
Austria	0	8	0	8
Belgium	0	28	1	27
Brazil	0	54	50	1
Canada	20	85	1	83
France	2	112	4	108
Germany	0	38	0	38
Hong Kong	68	0	0	0
India	128	0	0	0
Israel	0	47	0	47
Italy	1	19	0	19
Japan	239	0	0	0
Malaysia	19	0	0	0
Mexico	0	31	31	0
The Netherlands	0	1	0	1
New Zealand	3	6	1	5
Poland	0	44	0	44
Portugal	0	2	1	1
Puerto Rico	0	19	19	0
Singapore	37	0	0	0
South Korea	104	0	0	0
Spain	0	86	3	83
Switzerland	0	10	0	10
Taiwan	77	0	0	0
Turkey	0	3	0	3
United Kingdom	10	40	2	38
United States	46	1551	413	1137

[a]Some patients preferred not to report whether they were Hispanics. Thus, the total number might be larger than the sum of Hispanics and non-Hispanics in some countries.

Supplementary Table 2. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3–F4) in White and Asian Patients (1-cutoff Model)

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)			
				Sensitivity	Specificity	PPV	NPV
White Hispanic patients							
NAFLD fibrosis score (n = 372)	66	0.73 (0.68–0.78)	−1.455	90 (86–94)	40 (31–49)	74 (69–79)	68 (56–78)
			0.676	33 (27–40)	90 (84–95)	87 (79–93)	41 (35–47)
FIB-4 (n = 528)	58	0.81 (0.78–0.85)	1.3	81 (76–85)	62 (56–69)	75 (70–80)	70 (63–76)
			2.67	34 (29–39)	97 (94–99)	94 (87–97)	51 (46–56)
ELF (n = 539)	58	0.83 (0.79–0.86)	9.8	78 (73–83)	78 (72–84)	83 (79–87)	72 (66–78)
			11.3	22 (18–27)	99 (97–100)	97 (90–100)	48 (44–53)
Liver stiffness measurement (n = 238)	76	0.87 (0.82–0.92)	9.9 kPa	84 (78–89)	71 (58–83)	91 (85–94)	58 (45–70)
			11.4 kPa	79 (72–85)	75 (62–86)	91 (86–95)	52 (41–64)
White non-Hispanic patients							
NAFLD fibrosis score (n = 1348)	86	0.72 (0.68–0.76)	−1.455	89 (87–91)	30 (24–37)	89 (87–90)	31 (24–38)
			0.676	41 (38–44)	85 (79–90)	94 (92–96)	19 (17–22)
FIB-4 (n = 1697)	77	0.76 (0.73–0.78)	1.3	81 (78–83)	57 (52–62)	87 (84–88)	46 (42–51)
			2.67	32 (30–35)	91 (88–94)	93 (90–95)	28 (26–31)
ELF (n = 1720)	77	0.77 (0.75–0.8)	9.8	71 (69–74)	71 (66–75)	89 (87–91)	42 (39–46)
			11.3	18 (16–20)	97 (95–99)	96 (92–98)	26 (24–28)
Liver stiffness measurement (n = 1006)	89	0.76 (0.72–0.81)	9.9 kPa	84 (82–87)	52 (42–62)	94 (92–95)	28 (22–35)
			11.4 kPa	76 (73–79)	65 (55–74)	95 (93–96)	25 (20–30)
All white patients							
NAFLD fibrosis score (n = 1725)	82	0.73 (0.7–0.76)	−1.455	89 (87–91)	34 (29–39)	86 (84–87)	41 (35–48)
			0.676	40 (37–42)	87 (83–91)	93 (91–95)	25 (22–27)
FIB-4 (n = 2233)	73	0.78 (0.76–0.8)	1.3	81 (79–83)	59 (55–63)	84 (82–86)	53 (49–57)
			2.67	33 (30–35)	93 (91–95)	93 (91–95)	34 (32–36)
ELF (n = 2267)	72	0.79 (0.77–0.81)	9.8	73 (70–75)	73 (70–77)	88 (86–89)	51 (47–54)
			11.3	19 (17–21)	98 (96–99)	96 (93–98)	32 (29–34)
Liver stiffness measurement (n = 1247)	87	0.8 (0.76–0.84)	9.9 kPa	84 (82–86)	59 (51–66)	93 (91–95)	36 (30–42)
			11.4 kPa	77 (74, 79)	68 (61, 75)	94 (92, 96)	31 (26, 36)
Asian patients							
NAFLD fibrosis score (n = 586)	78	0.75 (0.7–0.8)	−1.455	88 (85–91)	43 (34–52)	85 (81–88)	50 (40–60)
			0.676	33 (29–37)	92 (86–96)	94 (89–97)	28 (23–32)
FIB-4 (n = 735)	69	0.8 (0.76–0.83)	1.3	88 (85–90)	48 (41–55)	79 (75–82)	64 (56–71)
			2.67	48 (43–52)	90 (85–93)	91 (87–94)	44 (39–48)
ELF (n = 748)	68	0.81 (0.77–0.84)	9.8	76 (72–80)	71 (65–77)	85 (81–88)	58 (53–64)
			11.3	24 (20–28)	96 (93–98)	93 (87–97)	37 (33–41)
Liver stiffness measurement (n = 431)	76	0.83 (0.78–0.87)	9.9 kPa	83 (78–87)	65 (55–74)	88 (84–92)	54 (45–63)
			11.4 kPa	73 (68–78)	78 (68–85)	91 (87–94)	48 (40–56)

AUROC, Area under the receiver operating characteristics curve; *CI*, confidence interval; *ELF*, Enhanced Liver Fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

Supplementary Table 3. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3–F4) in White and Asian Patients With Different BMI (1-cutoff Model)

	Percentage of patients with F3-F4	AUROC (95% CI)	Cutoff	% (95% CI)			
				Sensitivity	Specificity	PPV	NPV
White patients, BMI <30 kg/m^2							
NAFLD fibrosis score (n = 428)	81	0.76 (0.71–0.81)	−1.455	80 (76–84)	52 (41–63)	87 (83–91)	39 (30–48)
			0.676	20 (16–25)	99 (93–100)	99 (92–100)	23 (19–28)
FIB-4 (n = 429)	81	0.75 (0.69–0.8)	1.3	82 (78–86)	52 (41–63)	88 (84–91)	41 (32–51)
			2.67	36 (31–41)	87 (78–93)	92 (86–96)	25 (20–30)
ELF (n = 436)	80	0.8 (0.75–0.85)	9.8	70 (65–75)	77 (67–85)	92 (89–95)	39 (32–47)
			11.3	16 (12–20)	100 (96–100)	100 (94–100)	23 (19–27)
Liver stiffness measurement (n = 263)	91	0.79 (0.68–0.89)	9.9 kPa	78 (73–83)	74 (52–90)	97 (93–99)	25 (15–36)
			11.4 kPa	66 (59–72)	78 (56–93)	97 (93–99)	18 (11–27)
White patients, BMI ≥30 kg/m^2							
NAFLD fibrosis score (n = 1297)	82	0.73 (0.69–0.76)	−1.455	92 (90–94)	27 (22–34)	85 (83–87)	43 (35–52)
			0.676	46 (43–49)	83 (78–88)	93 (90–95)	25 (22–29)
FIB-4 (n = 1300)	82	0.78 (0.75–0.81)	1.3	80 (77–82)	61 (54–67)	90 (88–92)	39 (34–45)
			2.67	30 (27–33)	94 (90–97)	96 (93–98)	23 (20–26)
ELF (n = 1310)	82	0.8 (0.76–0.83)	9.8	73 (70–76)	75 (69–80)	93 (91–95)	38 (34–43)
			11.3	19 (16–21)	97 (95–99)	97 (94–99)	21 (19–24)
Liver stiffness measurement (n = 846)	93	0.77 (0.71–0.83)	9.9 kPa	86 (83–88)	50 (37–63)	96 (94–97)	22 (15–30)
			11.4 kPa	80 (77–83)	61 (48–73)	96 (95–98)	19 (14–26)
Asian patients, BMI <30 kg/m^2							
NAFLD fibrosis score (n = 376)	78	0.75 (0.69–0.81)	−1.455	87 (83–91)	43 (33–55)	84 (80–88)	49 (37–61)
			0.676	29 (24–34)	93 (85–97)	93 (86–98)	27 (22–32)
FIB-4 (n = 377)	78	0.75 (0.69–0.81)	1.3	93 (89–95)	39 (28–50)	84 (80–88)	59 (45–72)
			2.67	56 (50–62)	77 (67–86)	90 (84–94)	33 (27–40)
ELF (n = 381)	77	0.78 (0.72–0.83)	9.8	78 (73–83)	67 (56–77)	89 (85–93)	47 (38–56)
			11.3	23 (19–29)	97 (90–99)	96 (88–99)	27 (22–32)
Liver stiffness measurement (n = 220)	87	0.8 (0.72–0.88)	9.9 kPa	80 (74–86)	64 (44–81)	94 (89–97)	32 (20–46)
			11.4 kPa	70 (63–77)	79 (59–92)	96 (91–98)	28 (18–39)
Asian patients, BMI ≥30 kg/m^2							
NAFLD fibrosis score (n = 210)	79	0.75 (0.68–0.83)	−1.455	90 (84–94)	41 (26–57)	85 (79–90)	51 (34–69)
			0.676	40 (33–48)	91 (78–97)	94 (86–98)	29 (21–37)
FIB-4 (n = 210)	79	0.75 (0.68–0.83)	1.3	78 (71–84)	50 (35–65)	86 (79–91)	38 (26–52)
			2.67	33 (26–41)	93 (81–99)	95 (86–99)	27 (20–35)
ELF (n = 211)	79	0.78 (0.71–0.86)	9.8	71 (64–78)	70 (55–83)	90 (84–95)	39 (28–51)
			11.3	22 (16–29)	93 (81–99)	92 (79–98)	24 (18–31)
Liver stiffness measurement (n = 137)	85	0.85 (0.76–0.94)	9.9 kPa	85 (78–91)	71 (48–89)	94 (88–98)	47 (29–65)
			11.4 kPa	76 (67–83)	86 (64–97)	97 (91–99)	39 (25–55)

AUROC, Area under the receiver operating characteristics curve; *BMI*, body mass index; *CI*, confidence interval; *ELF*, Enhanced Liver Fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

Supplementary Table 4. Performance of Noninvasive Tests in Diagnosing Advanced Fibrosis (F3–F4) in White and Asian Patients by Age (1-cutoff Model)

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)			
				Sensitivity	Specificity	PPV	NPV
White patients, age <40 years							
NAFLD fibrosis score (n = 68)	69	0.7 (0.57–0.82)	−1.455	53 (38–68)	76 (53–92)	83 (65–94)	42 (26–59)
			0.676	13 (5–26)	95 (76–100)	86 (42–100)	33 (21–46)
FIB-4 (n = 105)	53	0.71 (0.61–0.81)	1.3	34 (22–48)	94 (83–99)	86 (65–97)	55 (44–66)
			2.67	5 (1–15)	100 (93–100)	100 (29–100)	48 (38–58)
ELF (n = 111)	51	0.75 (0.66–0.84)	9.8	40 (28–54)	91 (80–97)	82 (63–94)	59 (48–70)
			11.3	4 (0–12)	100 (93–100)	100 (16–100)	50 (40–59)
Liver stiffness measurement (n = 51)	78	0.71 (0.55–0.86)	9.9 kPa	80 (64–91)	55 (23–83)	86 (71–95)	43 (18–71)
			11.4 kPa	75 (59–87)	64 (31–89)	88 (73–97)	41 (18–67)
White patients, age 40–64 years							
NAFLD fibrosis score (n = 1261)	81	0.74 (0.7–0.77)	−1.455	89 (86–90)	36 (30–43)	85 (83–88)	43 (36–50)
			0.676	37 (34–40)	91 (87–94)	94 (92–97)	25 (23–28)
FIB-4 (n = 1637)	72	0.77 (0.75–0.8)	1.3	79 (77–81)	60 (56–65)	84 (81–86)	53 (49–57)
			2.67	30 (27–32)	94 (91–96)	93 (89–95)	34 (32–37)
ELF (n = 1655)	72	0.79 (0.76–0.81)	9.8	71 (69–74)	74 (70–78)	87 (85–89)	51 (47–54)
			11.3	18 (16–20)	98 (96–99)	96 (92–98)	32 (30–35)
Liver stiffness measurement (n = 923)	86	0.82 (0.78–0.86)	9.9 kPa	85 (82–87)	61 (52–70)	93 (91–95)	40 (34–48)
			11.4 kPa	77 (74–80)	70 (62–78)	94 (92–96)	34 (28–40)
White patients, age ≥65 years							
NAFLD fibrosis score (n = 396)	86	0.68 (0.6–0.75)	−1.455	96 (94–98)	7 (2–18)	87 (83–90)	24 (7–50)
			0.676	52 (46–57)	69 (55–81)	91 (86–95)	19 (14–25)
			0.12	71 (66–75)	55 (41–68)	91 (86–94)	23 (16–31)
FIB-4 (n = 491)	80	0.78 (0.73–0.83)	1.3	92 (89–94)	35 (26–45)	85 (81–88)	52 (39–64)
			2.67	45 (40–50)	88 (79–93)	94 (89–97)	28 (23–33)
			2	70 (66–75)	73 (63–82)	91 (88–94)	38 (31–45)
ELF (n = 501)	80	0.77 (0.72–0.82)	9.8	81 (77–85)	62 (52–71)	89 (86–92)	46 (37–55)
			11.3	24 (19–28)	96 (90–99)	96 (90–99)	24 (20–29)
Liver stiffness measurement (n = 273)	92	0.72 (0.62–0.83)	9.9 kPa	83 (77–87)	43 (22–66)	95 (91–97)	17 (8–30)
			11.4 kPa	76 (70–81)	57 (34–78)	96 (92–98)	17 (9–27)
Asian patients, age <40 years							
NAFLD fibrosis score (n = 39)	59	0.83 (0.69–0.97)	−1.455	57 (34–77)	88 (62–98)	87 (60–98)	58 (37–78)
			0.676	13 (3–34)	94 (70–100)	75 (19–99)	43 (26–61)
FIB-4 (n = 64)	38	0.76 (0.64–0.88)	1.3	42 (22–63)	88 (73–96)	67 (38–88)	71 (57–83)
			2.67	12 (3–32)	100 (91–100)	100 (29–100)	66 (52–77)
ELF (n = 66)	36	0.76 (0.64–0.88)	9.8	50 (29–71)	90 (77–97)	75 (48–93)	76 (62–87)
			11.3	4 (0–21)	100 (92–100)	100 (3–100)	65 (52–76)
Liver stiffness measurement (n = 35)	51	0.82 (0.69–0.96)	9.9 kPa	94 (73–100)	47 (23–72)	65 (44–83)	89 (52–100)
			11.4 kPa	78 (52–94)	65 (38–86)	70 (46–88)	73 (45–92)
Asian patients, age 40–64 years							
NAFLD fibrosis score (n = 392)	79	0.76 (0.71–0.82)	−1.455	87 (82–90)	42 (31–53)	84 (80–88)	46 (35–58)
			0.676	28 (23–33)	95 (88–99)	96 (89–99)	26 (22–32)
FIB-4 (n = 489)	68	0.8 (0.76–0.84)	1.3	87 (83–91)	46 (38–55)	78 (73–82)	63 (53–71)
			2.67	43 (37–48)	94 (89–97)	94 (89–97)	43 (38–49)
ELF (n = 500)	67	0.8 (0.76–0.84)	9.8	74 (69–78)	69 (62–76)	83 (78–87)	56 (49–63)
			11.3	23 (19–28)	96 (91–98)	92 (84–97)	38 (33–42)

Supplementary Table 4. Continued

	Percentage of patients with F3–F4	AUROC (95% CI)	Cutoff	% (95% CI)			
				Sensitivity	Specificity	PPV	NPV
Liver stiffness measurement (n = 291)	75	0.85 (0.79–0.9)	9.9 kPa	82 (76–87)	69 (57–80)	89 (84–93)	56 (45–66)
			11.4 kPa	73 (66–78)	82 (71–90)	92 (87–96)	50 (40–59)
Asian patients, age ≥65 years							
NAFLD fibrosis score (n = 155)	83	0.68 (0.57–0.79)	−1.455	98 (93–100)	19 (6–38)	85 (78–90)	62 (24–91)
			0.676	48 (40–57)	81 (62–94)	93 (83–98)	25 (16–35)
			0.12	67 (58–75)	56 (35–75)	88 (80–94)	26 (16–40)
FIB-4 (n = 182)	81	0.64 (0.54–0.74)	1.3	97 (92–99)	9 (2–24)	82 (76–88)	38 (9–76)
			2.67	65 (57–73)	56 (38–73)	86 (79–92)	27 (17–39)
			2	83 (76–89)	38 (22–56)	85 (79–91)	34 (20–51)
ELF (n = 182)	81	0.76 (0.67–0.84)	9.8	86 (80–92)	56 (38–73)	90 (83–94)	49 (32–65)
			11.3	28 (21–36)	94 (80–99)	95 (85–99)	23 (16–31)
Liver stiffness measurement (n = 105)	87	0.78 (0.68–0.89)	9.9 kPa	82 (73–90)	64 (35–87)	94 (86–98)	36 (18–57)
			11.4 kPa	75 (65–83)	71 (42–92)	94 (86–98)	30 (16–49)

AUROC, Area under the receiver operating characteristics curve; *CI*, confidence interval; *ELF*, Enhanced Liver Fibrosis panel; *FIB-4*, Fibrosis-4 index; *NAFLD*, nonalcoholic fatty liver disease; *NPV*, negative predictive value; *PPV*, positive predictive value.

Supplementary Table 5. Two-step Approach for the Diagnosis of Advanced Fibrosis (F3–4) in White and Asian Patients

Test sequence	% (95% CI)					
	Sensitivity	Specificity	PPV	NPV	Gray zone	Accuracy
White patients						
FIB-4 → ELF (n = 2273)	67 (64–69)	93 (91–95)	96 (95–97)	52 (49–55)	25 (24–27)	74 (72–76)
FIB-4 → LSM (n = 2245)	75 (73–77)	90 (87–92)	95 (94–96)	57 (54–60)	21 (19–22)	79 (77–81)
NFS → ELF (n = 2271)	72 (70–74)	92 (90–94)	96 (95–97)	56 (53–59)	28 (26–30)	78 (76–79)
NFS → LSM (n = 1872)	82 (80–84)	80 (76–84)	94 (92–95)	55 (51–59)	19 (17–21)	82 (80–84)
Asian patients						
FIB-4 → ELF (n = 748)	76 (72–79)	89 (84–92)	93 (91–96)	63 (58–68)	22 (19–25)	80 (77–83)
FIB-4 → LSM (n = 740)	83 (80–87)	86 (81–90)	93 (90–95)	70 (65–76)	19 (17–22)	84 (81–87)
NFS → ELF (n = 748)	74 (70–78)	94 (90–96)	96 (94–98)	63 (58–68)	31 (28–35)	80 (77–83)
NFS → LSM (n = 662)	81 (77–84)	84 (78–89)	93 (90–95)	62 (56–68)	21 (18–24)	82 (78–84)

Note: In the first step, patients with results below the lower cutoff and above the higher cutoff were considered to have advanced fibrosis excluded and confirmed, respectively. Those with results between the 2 cutoffs (ie, gray zone) would undergo the second test. The cutoffs of the noninvasive tests are as follows: FIB-4, 1.3–2.67; NFS, −1.455 to 0.676; ELF, 9.8–11.3; LSM, 9.9–11.4.

ELF, Enhanced Liver Fibrosis panel; *FIB-4*, Fibrosis-4 index; *LSM*, liver stiffness measurement; *NFS*, nonalcoholic fatty liver disease fibrosis score; *NPV*, negative predictive value; *PPV*, positive predictive value.