Impact of age on NIS2+™ and other non-invasive blood tests for the evaluation of liver disease and detection of at-risk MASH

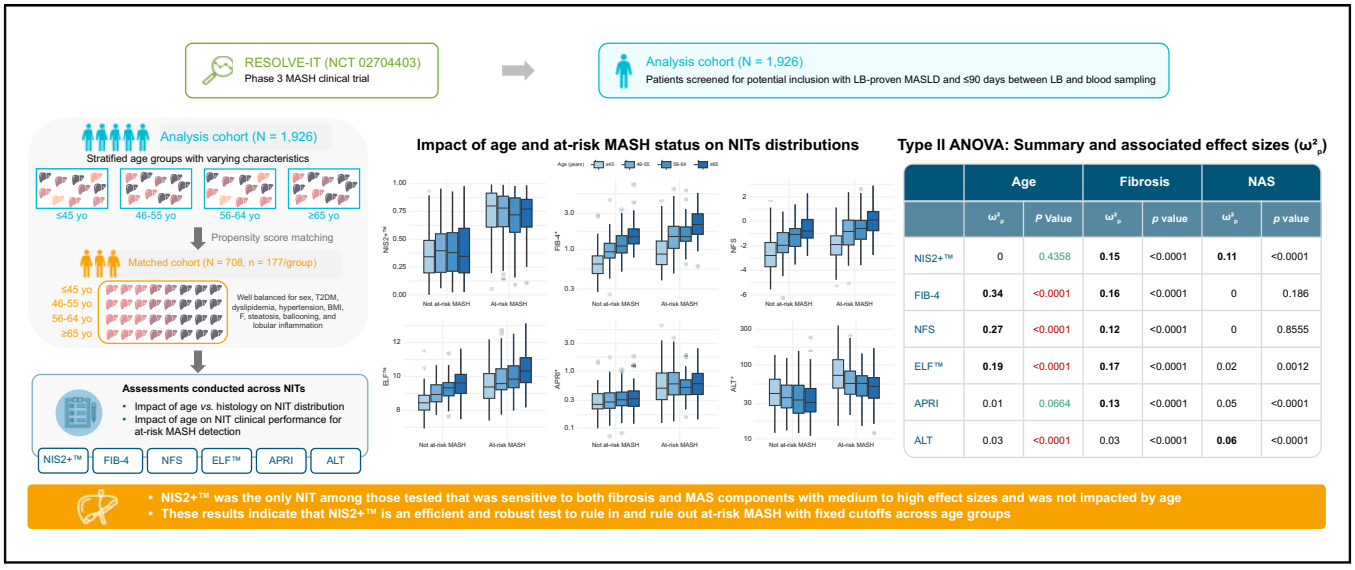
Authors

Quentin M. Anstee, Jeremy Magnanensi, Yacine Hajji, Alexandra Caron, Zouher Majd, Christian Rosenquist, Dean W. Hum, Bart Staels, Margery A. Connelly, Rohit Loomba, Stephen A. Harrison, Vlad Ratziu, Arun J. Sanyal

Correspondence

jeremy.magnanensi@genfit.com (J. Magnanensi).

Graphical abstract



	Age		Fibrosis		NAS	
	ω^2_p	*P* Value	ω^2_p	*p* value	ω^2_p	*p* value
NIS2+™	0	0.4358	**0.15**	<0.0001	**0.11**	<0.0001
FIB-4	**0.34**	<0.0001	**0.16**	<0.0001	0	0.186
NFS	**0.27**	<0.0001	**0.12**	<0.0001	0	0.8555
ELF™	**0.19**	<0.0001	**0.17**	<0.0001	0.02	0.0012
APRI	0.01	0.0664	**0.13**	<0.0001	0.05	<0.0001
ALT	0.03	<0.0001	0.03	<0.0001	**0.06**	<0.0001

Highlights

- Robust non-invasive tests are needed for the evaluation of MASLD across all ages.
- NIS2+™ and APRI were the only panels not significantly impacted by age.
- Age impacted the performance of FIB-4, NFS, and ELF™.
- NIS2+™ was sensitive to both fibrosis and MAS, adapted for at-risk MASH detection.
- NIS2+™ showed robust performance across ages to detect at-risk MASH at fixed cut-offs.

Impact and Implications

While metabolic dysfunction-associated steatotic liver disease (MASLD) can affect individuals of all ages, patient age could represent an important confounding factor when interpreting non-invasive test (NIT) results, highlighting the need for reliable and efficient NITs that are not impacted by age and that could be interpreted with fixed cut-offs, irrespective of patient age. We report the impact of age on different well-established NITs – among those tested, only two panels, NIS2+™ and APRI, were not impacted by age and can be used and interpreted independently of patient age. NIS2+™ was also sensitive to both fibrosis and MAS, further confirming its efficiency for the detection of the composite endpoint of at-risk MASH and its potential as a valuable candidate for large-scale implementation in clinical practice and clinical trials.

https://doi.org/10.1016/j.jhepr.2024.101011

Research article

JHEP|Reports

Impact of age on NIS2+™ and other non-invasive blood tests for the evaluation of liver disease and detection of at-risk MASH

Quentin M. Anstee,[1,2] Jeremy Magnanensi,[3,*] Yacine Hajji,[3] Alexandra Caron,[3] Zouher Majd,[3] Christian Rosenquist,[3] Dean W. Hum,[3] Bart Staels,[4] Margery A. Connelly,[5] Rohit Loomba,[6] Stephen A. Harrison,[7,8,†] Vlad Ratziu,[9,†] Arun J. Sanyal[10,†]

[1]*Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK;* [2]*Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Freeman Hospital, Newcastle upon Tyne, UK;* [3]*Genfit S.A., Loos, France;* [4]*Université de Lille, INSERM, CHU Lille, Institut Pasteur de Lille, Lille, France;* [5]*Labcorp, Morrisville, NC, US;* [6]*NAFLD Research Center, Division of Gastroenterology, University of California at San Diego, La Jolla, CA, US;* [7]*Summit Clinical Research, San Antonio, TX, US;* [8]*Radcliffe Department of Medicine, University of Oxford, Oxford, UK;* [9]*Sorbonne Université, Institute for Cardiometabolism and Nutrition, Hôpital Pitié-Salpêtrière, Paris, France;* [10]*Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University School of Medicine, Richmond, VA, US*

JHEP Reports **2024.** https://doi.org/10.1016/j.jhepr.2024.101011

Background & Aims: Robust performance of non-invasive tests (NITs) across ages is critical to assess liver disease among patients with metabolic dysfunction-associated liver disease (MASLD). We evaluated the impact of age on the performance of NIS2+™ *vs.* other NITs.

Methods: An analysis cohort (N = 1,926) with biopsy-proven MASLD was selected among individuals screened for the phase III RESOLVE-IT clinical trial and divided into ≤45, 46–55, 56–64, and ≥65 years groups. To avoid potential confounding effects, a well-balanced cohort (n = 708; n = 177/age group) was obtained by applying a propensity score-matching algorithm to the analysis cohort. Baseline values of biomarkers and NITs were compared across age groups using one-way ANOVA, and the impact of age and histology were compared through three-way ANOVA. The impact of age on NIT performance for the detection of at-risk metabolic dysfunction-associated steatohepatitis (MASH; MASLD activity score [MAS] ≥4 and fibrosis stage [F] ≥2) was also evaluated.

Results: Age did not affect the distributions of NIS2+™ and APRI (aspartate aminotransferase-to-platelet ratio index), but significantly (p <0.0001) impacted those of NFS (NAFLD fibrosis score), FIB-4 (Fibrosis-4 index), and Enhanced Liver Fibrosis (ELF™) score. NIS2+™ was the only NIT on which fibrosis and MAS exerted a moderate to large effect. While the impact of fibrosis on APRI was moderate, that of MAS was low. The impact of age on FIB-4 and NFS was larger than that of fibrosis. NIS2+™ exhibited the highest AUROC values for detecting at-risk MASH across age groups, with stable performances irrespective of cut-offs.

Conclusions: NIS2+™ was not significantly impacted by age and was sensitive to both fibrosis and MAS grade, demonstrating a robust performance to rule in/out at-risk MASH with fixed cut-offs.

Impact and Implications: While metabolic dysfunction-associated steatotic liver disease (MASLD) can affect individuals of all ages, patient age could represent an important confounding factor when interpreting non-invasive test (NIT) results, highlighting the need for reliable and efficient NITs that are not impacted by age and that could be interpreted with fixed cut-offs, irrespective of patient age. We report the impact of age on different well-established NITs – among those tested, only two panels, NIS2+™ and APRI, were not impacted by age and can be used and interpreted independently of patient age. NIS2+™ was also sensitive to both fibrosis and MAS, further confirming its efficiency for the detection of the composite endpoint of at-risk MASH and its potential as a valuable candidate for large-scale implementation in clinical practice and clinical trials.

Keywords: Non-invasive blood-based tests; NITs; NIS2+™; metabolic dysfunction-associated steatohepatitis; MASH; at-risk MASH; metabolic dysfunction-associated liver disease; MASLD; fibrosis; age; NASH; NAFLD.

Received 29 August 2023; received in revised form 14 December 2023; accepted 8 January 2024; available online 19 January 2024

† Joint senior authors

* Corresponding author. Address: GENFIT - Parc Eurasanté 885, Avenue Eugène Avinée, 59 120 Loos, France; Tel.: +33 (0)3 20 16 40 00.

E-mail address: jeremy.magnanensi@genfit.com (J. Magnanensi).

Introduction

Metabolic dysfunction-associated steatohepatitis (MASH; formerly known as non-alcoholic steatohepatitis [NASH]) is the progressive form of metabolic dysfunction-associated liver disease (MASLD; formerly known as non-alcoholic fatty liver disease [NAFLD]), the leading cause of chronic liver disease.[1–3] MASH is characterized by steatosis with hepatocellular ballooning and lobular inflammation and can progress toward

cirrhosis, potentially leading to severe liver complications, including hepatic decompensation and hepatocellular carcinoma.[2,4] At-risk MASH (MASH with a MASLD activity score [MAS; formerly known as NAFLD activity score, NAS] ≥4 and a fibrosis stage [F] ≥2) is linked to a higher risk of disease progression; therefore, timely diagnosis of this condition is critical.[5] While liver histology is the current clinical reference standard for MASH diagnosis, broad utilization in the clinic is limited by its cost, the invasive nature of the procedure, and variability in interpretation, highlighting an unmet need for simpler/easier-to-implement, accurate diagnostics.[6–8]

NIS4® is a blood-based panel comprising four independent MASH-associated biomarkers (miR-34a-5p, alpha-2 macroglobulin, YKL-40, and glycated hemoglobin) designed to detect at-risk MASH among patients with metabolic risk factors.[5] This non-invasive test (NIT) has been validated in independent cohorts, showing high overall diagnostic performance (AUROC = 0.8).[5] Fixed low and high cut-offs have been derived to achieve high performances for ruling in/out at-risk MASH, with 87.1% specificity, 50.7% sensitivity, and a positive predictive value of 79.2% when ruling in this condition.[5] Recently, the NIMBLE consortium published independent data confirming NIS4® performances for the detection of at-risk MASH, achieving an AUROC of 0.815.[9]

NIS2+™, an optimization of the NIS4® technology, was developed to further improve the robustness of scores across different MASLD subpopulations of interest (*e.g.*, those defined by age, type 2 diabetes mellitus [T2DM] status, sex, and BMI). NIS2+™ comprises only two blood-based biomarkers (miR-34a-5p and YKL-40), as well as a correction for the effect of sex on miR-34a-5p, and returned reliable scores that were not impacted by patients' characteristics, increasing its potential for large-scale use in clinical practice.[10] NIS2+™ exhibited a statistically higher AUROC (0.813) *vs.* NIS4® (0.792; $p = 0.0002$), demonstrating superior clinical performance when ruling in/out at-risk MASH across MASLD subpopulations of interest.[10] In addition, NIS2+™ was shown to have the potential to reduce the screen failure rate associated with liver biopsy in MASH clinical trials, one of their main feasibility aspects, in a cost-effective manner.[11]

Given that MASLD affects individuals of all ages, it is critical that NITs used in the clinic exhibit a robust clinical performance across a wide age range.[12] This is reflected in guidance from the US Food and Drug Administration, which highlights the relevance of considering different age populations for the proper characterization of a device's safety and efficacy.[13] McPherson *et al.* reported that NAFLD fibrosis score (NFS) and Fibrosis-4 (FIB-4) index underperformed for the detection of advanced fibrosis among patients aged ≤35 years, and that specificity decreased as age increased, leading to an unacceptably high false-positive rate in patients aged ≥65 years, for whom higher cut-offs to rule out advanced fibrosis were proposed.[14] Age was also identified as the most relevant factor impacting the enhanced liver fibrosis (ELF™) score by Lichtinghagen *et al.*, with significant increases observed in healthy volunteers and patients with chronic hepatitis C aged <20 years and up to >60 years.[15] In a recent study, Harrison *et al.* reported that NIS2+™ exhibited a robust performance in detecting at-risk MASH; however, additional data are needed to robustly establish performance over a broader age spectrum.[10]

The objective of this analysis was twofold. First, we compared the effect of age with that of histology (fibrosis and MAS) on the distribution of NIS2+™ and other NITs (NFS, FIB-4, ELF™, aspartate aminotransferase-to-platelet ratio index [APRI], and alanine aminotransferase [ALT]). We then evaluated the impact of age on the clinical performances of NIS2+™ and other NITs for the detection of at-risk MASH. Given that NFS, FIB-4, ELF™, and APRI have not been designed for the detection of at-risk MASH, but of advanced fibrosis (F ≥3), the impact of age on the latter parameter was also assessed. ALT was included in the analysis as a surrogate marker of reference for MASH.

Materials and methods

Analysis cohort

The analysis cohort comprised 1,926 patients with biopsy-proven MASLD, aged 18–75 years, selected among those screened for potential inclusion in the phase III RESOLVE-IT clinical trial (NCT02704403), a randomized, double-blind, placebo-controlled, multicenter, international trial that enrolled non-cirrhotic patients with at-risk MASH, and evaluated the safety and efficacy of elafibranor. This global trial screened over 5,000 patients aged 18–75 years in more than 270 centers between March 2016 and March 2020. The patient selection criteria for RESOLVE-IT are shown in Table S1. Patients in the analysis cohort had available biopsy results at baseline and available data for the calculation of NIS2+™, FIB-4, NFS, APRI, and ELF™ scores (formulas for these NITs are depicted in Table S2). To ensure consistency between biomarker levels and histology assessments, the gap between blood sampling and biopsy dates for selected patients was ≤90 days. Patients were stratified into four age groups: ≤45 years (n = 451), 46–55 years (n = 519), 56–64 years (n = 581), and ≥65 years (n = 375). A supplementary analysis was conducted in patients aged ≤35 (n = 152) *vs.* >35 years (n = 1,774).

Matching process

A propensity score–matching (PSM) algorithm was applied to select sets of patients among the four age groups (matched cohort, n = 708; n = 177/age group), which were well balanced for the following confounding factors: sex (male proportion), T2DM, dyslipidemia, hypertension, BMI, and fibrosis stage, as well as steatosis, ballooning, and lobular inflammation by score. The matching process was conducted separately in not at-risk and at-risk MASH populations to allow for an assessment of the impact of age on NIT scores in each of these populations independently. The PSM algorithm was also applied to identify the optimal subset of patients aged >35 years to match those aged ≤35 years (n = 135/group). The *matchit* function of the *MatchIt* R package (version 4.3.2) with genetic method was used to conduct these selections.[16]

Statistical analyses

Descriptive statistics were generated for baseline demographic, clinical, and histologic characteristics of the analysis cohort, as well as for each age group of the analysis or matched cohort. These were compared using the Kruskal-Wallis test (or Student's *t* tests for comparisons of groups aged ≤35 and >35 years) for numerical features, and χ^2 test for proportions comparison.

Baseline distributions of biomarkers and NIT scores (mean ± SD) were compared across age groups using one-way ANOVA tests (or Student's *t* tests for comparisons of groups aged ≤35 and >35 years). To control for type-1 error rate (probability of erroneously rejecting a true null hypothesis), the distribution of all biomarkers and NIT scores was checked for normality through

skewness and kurtosis statistics estimated using the analysis cohort (N = 1,926).[17] Biomarkers and NITs associated with a skewness score >2 and/or kurtosis score >7 were transformed using a $\log_{10}$ transformation (Table S3), while NIS2+™ scores resulting from a logistic regression with values ranging from 0 to 1 were logit-transformed to allow for a linear scale analysis when processed through ANOVA modeling.[18] ANOVA-associated *p* values were therefore derived from original or transformed data.

The impact of age *vs.* histology on the distribution of NIT scores was assessed through three-way type II ANOVA modeling in the matched cohort (n = 708) using age by group, F by stage, and MAS by group (1–2, 3–4, 5–6, 7–8) as factors. While *p* values allow for assessment of the significance of the effect of different factors on the distribution of NIT scores, they do not reflect the magnitude of this effect. Therefore, effect sizes – a quantitative measure that reflects the magnitude of the impact that each factor had on the NIT scores distribution – for each parameter of the ANOVA models were derived and reported utilizing partial ω^2_p estimators.[19] The thresholds used to interpret effect size values (magnitude of impact) were defined as <0.01=very low, ≥0.01 to <0.06=low, ≥0.06 to <0.14=medium, and ≥0.14=large.[20] R^2 was calculated and reported as the proportion of total NIT variance explained by age, F, and MAS. Depending on the intended use of each NIT, estimated marginal means were calculated by age group across F or MAS class.

To further evaluate the impact of age *vs.* histology on NIT scores, a comparison of the overall diagnostic performance of NITs for the detection of histologic endpoints (patients with at-risk MASH or F ≥3) and for the pairwise classification of patients into age groups was conducted using the AUROC values, reported with 95% CIs obtained with 1,000 bootstrap samples.

AUROC values of NITs across age groups were calculated to evaluate the impact of age on their overall clinical performance for the detection of at-risk MASH. Differences in AUROC values were assessed through adapted Delong tests (pROC R package, version 1.18.0). The impact of age on NIT clinical performances for the diagnosis of at-risk MASH was further investigated by calculating sensitivity and specificity across age groups using Youden cut-offs derived for each NIT in the matched cohort (n = 708). Differences in sensitivity and specificity among NITs between the subpopulations of patients aged ≤45 and ≥65 years were also calculated for the entire range of associated meaningful cut-off values to represent the impact of age on clinical performances by cut-off values.

Results

The analysis cohort (N = 1,926) was composed of patients with biopsy-proven MASLD and metabolic risk factors. Demographic and baseline characteristics of this cohort are shown in Table S4. Most patients were male (62%), White or Hispanic (91%), and exhibited a high prevalence of T2DM, dyslipidemia, hypertension, and obesity (43%, 49%, 57%, and 71%, respectively); 72% had MASH, 50% had at-risk MASH, and 36% had F ≥3. Demographic and baseline characteristics of the analysis cohort by age group

Table 1. Demographic and baseline characteristics of the matched cohort by age group.

	≤45 years (n = 177)	46–55 years (n = 177)	56–64 years (n = 177)	≥65 years (n = 177)	*p* values
Demographics					
Sex, male	123 (70%)	120 (68%)	119 (67%)	122 (69%)	0.9671
Age, years	39 (6)	51 (3)	60 (3)	69 (3)	<0.0001
Clinic					
T2DM	65 (37%)	71 (40%)	72 (41%)	70 (40%)	0.8762
Dyslipidemia	85 (48%)	88 (50%)	93 (53%)	90 (51%)	0.8570
Hypertension	102 (58%)	108 (61%)	111 (63%)	108 (61%)	0.7985
BMI, kg/m^2	33.11 (5.96)	33.01 (5.61)	32.93 (4.61)	32.84 (5.22)	0.9714
Obesity	123 (70%)	121 (68%)	128 (72%)	119 (67%)	0.7555
Histology					
MASLD	177 (100%)	177 (100%)	177 (100%)	177 (100%)	n.a.
MASH	123 (70%)	125 (71%)	124 (70%)	124 (70%)	0.9967
At-risk MASH	91 (51%)	91 (51%)	91 (51%)	91 (51%)	1.0
F	1.93 (1.14)	1.89 (1.13)	1.9 (1.12)	1.9 (1.14)	0.9889
F, 0/1/2/3/4	23 (13.0%)/43 (24.3%)/ 45 (25.4%)/56 (31.6%)/ 10 (5.6%)	21 (11.9%)/50 (28.2%)/ 43 (24.3%)/53 (29.9%)/ 10 (5.6%)	19 (10.7%)/53 (29.9%)/ 41 (23.2%)/54 (30.5%)/ 10 (5.6%)	22 (12.4%)/49 (27.7%)/ 41 (23.2%)/55 (31.1%)/ 10 (5.6%)	n.a.
MAS	4.29 (1.74)	4.3 (1.68)	4.29 (1.76)	4.31 (1.71)	0.9991
MAS, 1–2/3–4/ 5–6/7–8	27 (15.3%)/68 (38.4%)/ 62 (35%)/20 (11.3%)	28 (15.8%)/64 (36.2%)/ 68 (38.4%)/17 (9.6%)	30 (16.9%)/62 (35.0%)/ 61 (34.5%)/24 (13.6%)	29 (16.4%)/63 (35.6%)/ 65 (36.7%)/20 (11.3%)	n.a.
Steatosis	2.02 (0.79)	2.04 (0.78)	2.04 (0.78)	2.04 (0.78)	0.9962
Steatosis score, 1/2/3	53 (29.9%)/67 (37.9%)/ 57 (32.2%)	50 (28.2%)/70 (39.5%)/ 57 (32.2%)	50 (28.2%)/70 (39.5%)/ 57 (32.2%)	50 (28.2%)/70 (39.5%)/ 57 (32.2%)	n.a.
Ballooning	1.05 (0.79)	1.04 (0.79)	1.05 (0.79)	1.05 (0.79)	0.9998
Ballooning score, 0/1/2	51 (28.8%)/67 (37.9%)/ 59 (33.3%)	51 (28.8%)/68 (38.4%)/ 58 (32.8%)	51 (28.8%)/67 (37.9%)/ 59 (33.3%)	51 (28.8%)/67 (37.9%)/ 59 (33.3%)	n.a.
Lobular inflammation	1.23 (0.66)	1.22 (0.62)	1.21 (0.63)	1.23 (0.63)	0.9949
Lobular inflammation score, 0/1/2/3	18 (10.2%)/106 (59.9%)/ 48 (27.1%)/5 (2.8%)	16 (9%)/109 (61.6%)/ 49 (27.7%)/3 (1.7%)	17 (9.6%)/109 (61.6%)/ 48 (27.1%)/3 (1.7%)	16 (9.0%)/108 (61.0%)/ 50 (28.2%)/3 (1.7%)	n.a.

Matched cohort, n = 708. At-risk MASH was defined as MASH with MAS ≥4 and F ≥2. Values are expressed as n (%) or mean (SD). *p* values were calculated by using either Chi square tests for proportion comparisons or Kruskal-Wallis tests for numerical features.
MAS, MASLD activity score; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated liver disease; n.a., not applicable; T2DM, type 2 diabetes mellitus.

are depicted in Table S5. While patients were evenly distributed across age groups, the prevalence of T2DM, dyslipidemia, and hypertension, as well as the mean fibrosis score and the prevalence of F ≥3 increased with age, whereas BMI, obesity, MAS, and steatosis scores decreased with age. Given that these clinical and histologic differences could act as potential confounding factors when assessing the performance of NITs, a PSM algorithm was applied to select subpopulations well matched for all the above-mentioned confounders, with the exception of age, the only variable for which distribution differed among groups. The resulting matched cohort (n = 708, n = 177/age group; Table 1) was well balanced for sex, as well as for clinical and histologic characteristics. Across age groups, 37%–41% of patients had T2DM, 48%–53% had dyslipidemia, 58%–63% had hypertension, and 67%–72% were obese in this cohort. The prevalence of MASH and at-risk MASH was 70%–71% and 51%, respectively; mean fibrosis score was 1.9, with ∼36% of patients exhibiting F ≥3, and mean MAS was 4.3.

The distribution of NIT scores and associated biomarkers across age groups in the matched cohort is shown in Table 2. Mean scores for FIB-4, NFS, and ELF™ significantly increased with age (change from ≤45 to ≥65 years: 0.92 to 2.04, −2.28 to −0.27, and 8.94 to 10.06, respectively; p <0.0001 for all comparisons). Aspartate aminotransferase (AST)/ALT means were also impacted, exhibiting increasing values with increasing age (change from ≤45 to ≥65 years: 0.72–0.88; p <0.0001). Neither NIS2+™ nor APRI was significantly impacted by age across groups (p = 0.7111 and p = 0.1891, respectively).

Table 2. Means (SD) of NIT scores and associated biomarker levels in the matched cohort by age group.

	≤45 years		46–55 years		56–64 years		≥65 years		
	n	Value	n	Value	n	Value	n	Value	*p* values
NIS2+™									
NIS2+™*	177	0.56 (0.28)	177	0.56 (0.28)	177	0.55 (0.26)	177	0.57 (0.28)	0.7111
miR34-a, FoldC†	177	1.31 (1.96)	177	1.19 (1.57)	177	0.9 (0.90)	177	0.95 (1.04)	0.1138
YKL-40, ng/ml†	177	63.98 (60.36)	177	90.58 (149.54)	177	95.93 (121.46)	177	117.01 (107.43)	<0.0001
FIB-4/NFS/APRI									
FIB-4†	177	0.92 (0.62)	177	1.38 (0.89)	177	1.52 (0.75)	177	2.04 (1.04)	<0.0001
NFS	177	−2.28 (1.42)	177	−1.36 (1.41)	177	−0.96 (1.16)	177	−0.27 (1.15)	<0.0001
APRI†	177	0.55 (0.63)	177	0.58 (0.54)	177	0.48 (0.29)	177	0.56 (0.39)	0.1891
AST, IU/L†	177	45.61 (37.58)	177	43.49 (33.9)	177	36.97 (17.86)	177	39.92 (23.06)	0.273
ALT, IU/L†	177	68.78 (52.55)	177	58.56 (42.59)	177	50.31 (27.05)	177	48.92 (31.03)	2e-04
Platelets, $10e^9$/L	177	252.84 (65.91)	177	232.08 (67.47)	177	225.31 (59.85)	177	208.87 (58.16)	<0.0001
FPG, mmol/L†	176	5.94 (2.08)	177	6.08 (1.80)	176	6.23 (1.87)	177	6.18 (1.73)	0.1841
Albumin, g/L	177	46.27 (3.32)	177	45.99 (3.06)	177	45.81 (2.87)	177	44.97 (3.02)	7e-04
ELF™									
ELF™	177	8.94 (1.02)	177	9.44 (0.97)	177	9.66 (0.86)	177	10.06 (1.02)	<0.0001
Hyaluronic acid, ng/ml†	177	38.65 (50.93)	177	70.82 (156.35)	177	79.81 (78.45)	177	126.76 (155.42)	<0.0001
P3NP, ng/ml†	177	10.85 (5.54)	177	11.37 (6.36)	177	10.67 (4.91)	177	12.14 (6.37)	0.0774
TIMP1, ng/ml†	177	265.2 (72.52)	177	264.05 (73.86)	177	267.81 (63.35)	177	275.74 (85.77)	0.5791
Other biomarkers									
AST/ALT ratio†	177	0.72 (0.30)	177	0.78 (0.28)	177	0.79 (0.24)	177	0.88 (0.34)	<0.0001
A2M, g/L	176	2.09 (0.85)	170	2.14 (0.79)	172	2.47 (0.92)	176	2.68 (0.86)	<0.0001
HbA1c, %	174	6.12 (1.07)	177	6.16 (0.92)	174	6.22 (0.99)	177	6.2 (0.87)	0.7915
GGT, IU/L†	177	91.88 (131.22)	177	76.4 (100.73)	177	75.36 (88.19)	177	81.98 (110.86)	0.2114
ALP, IU/L†	176	84.64 (38.34)	177	81.73 (31.60)	177	87.35 (26.68)	177	85.74 (29.57)	0.1168
Triglycerides, mmol/L†	177	2.46 (2.31)	177	2.09 (3.14)	177	2.05 (1.31)	177	1.81 (0.81)	0.0214
Total cholesterol, mmol/L	177	4.92 (1.14)	177	4.71 (1.23)	177	4.66 (1.10)	176	4.59 (1.10)	0.0344

Matched cohort, n = 708. p values were calculated using Welch-adapted one-way ANOVA tests based either on raw or transformed (log_{10} or logit) data for biomarkers and NITs that were associated to a Skewness score >2 and/or a Kurtosis score >7.

A2M, alpha-2-macroglobulin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; FPG, fasting plasma glucose; GGT, γ-glutamyl transferase; HbA1c, glycated hemoglobin; MAS, MASLD activity score; MASLD, metabolic dysfunction-associated liver disease; miR34-a, microRNA 34a; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score; NIT, non-invasive test; P3NP, type III procollagen peptide; TIMP1, tissue inhibitor matrix metalloproteinase 1.

* Logit-transformed for computation of p values.

† Log-transformed (base 10) for computation of p values.

Table 3. Type II ANOVA modeling: Summary and associated effect sizes.

	Full model	Age		Fibrosis		MAS	
	R^2	ω^2_p	*p* values	ω^2_p	*p* values	ω^2_p	*p* values
NIS2+™	0.42	0 (0, 0)	0.4358	0.15 (0.1, 0.2)	<0.0001	0.11 (0.07, 0.15)	<0.0001
FIB-4	0.46	0.34 (0.29, 0.39)	<0.0001	0.16 (0.11, 0.2)	<0.0001	0 (0, 0.01)	0.186
NFS	0.35	0.27 (0.21, 0.32)	<0.0001	0.12 (0.08, 0.16)	<0.0001	0 (0, 0)	0.8555
ELF™	0.39	0.19 (0.14, 0.24)	<0.0001	0.17 (0.12, 0.22)	<0.0001	0.02 (0, 0.04)	0.0012
APRI	0.30	0.01 (0, 0.02)	0.0664	0.13 (0.08, 0.17)	<0.0001	0.05 (0.02, 0.08)	<0.0001
ALT	0.22	0.03 (0.01, 0.06)	<0.0001	0.03 (0.01, 0.06)	<0.0001	0.06 (0.03, 0.09)	<0.0001

R^2 was calculated and reported as a measure of total NIT variance explained by age, fibrosis, and NAS.

ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; MAS, MASLD activity score; MASLD, metabolic dysfunction-associated liver disease; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score; NIT, non-invasive test.

The impact of age, fibrosis stage, and MAS on the distribution of NIT scores was further investigated using type II ANOVA modeling (Table 3). This approach allowed for the comparison of the magnitude of the impact of age *vs.* that of histology (fibrosis and MAS) through effect sizes, which were reported as partial ω^2_p estimators as described in the 'Materials and methods' section. Among all NITs evaluated, the magnitude of the impact of age was largest on FIB-4 and NFS (ω^2_p effect size: 0.34 and 0.27, respectively), surpassing that of fibrosis, their theoretical main driver (ω^2_p effect size: 0.16 and 0.12, respectively). While ELF™ was strongly impacted by age (ω^2_p effect size: 0.19), the magnitude of this effect was similar to that of fibrosis (ω^2_p effect size: 0.17). Neither FIB-4 nor NFS were impacted by MAS (p = 0.186 and p = 0.8555, respectively). Even though MAS exerted an effect on ELF™, the magnitude of this effect was low (ω^2_p effect size: 0.02). The impact of age and fibrosis on ALT was low (ω^2_p effect size for both parameters: 0.03), while that of MAS was moderate (ω^2_p effect size: 0.06), consistent with its role as a surrogate for disease activity. Neither NIS2+™ nor APRI were significantly impacted by age (p = 0.4358 and p = 0.0664, respectively). MAS and fibrosis exhibited a low and moderate impact on APRI scores, respectively (ω^2_p effect size: 0.05 and ω^2_p: 0.13). Among all NITs, including the disease activity surrogate ALT, the effect of MAS was largest on NIS2+™ (ω^2_p effect size: 0.11); the magnitude of the impact of fibrosis on this NIT was also large (ω^2_p effect size: 0.15), comparable to those of FIB-4 or ELF™. NIS2+™ and FIB-4 achieved the highest R^2 across NITs, with 42% and 46% of their total variability, respectively, being associated with age, F, and MAS; however, while FIB-4 variability was mainly related to age, that of NIS2+™ was mostly associated with histology, as this NIT was not impacted by age. The effect size of age *vs.* histology on the distribution of NIT scores observed in the linear modeling is depicted using estimated marginal means in Figs S1 and S2.

To further compare the effects of age and histology on NIS2+™ and other NITs, ROC curves and associated AUROC values were derived for the detection of their respective histologic endpoints (F ≥3 for FIB-4, NFS, ELF™, and APRI; MASH for ALT; at-risk MASH for NIS2+™) and used as benchmarks (Fig. 1). This analysis included six pairs of age-based subgroups (≤45 *vs.* 46–55 years, ≤45 *vs.* 56–64 years, ≤45 *vs.* ≥65 years, 46–55 *vs.* 56–64 years, 46–55 *vs.* ≥65 years, 56–64 *vs.* ≥65 years), which were used to evaluate the ability of NITs to accurately classify patients within their corresponding age subgroup through ROC curves and AUROC values. Given that the four age-based groups originally established in this analysis were matched for different factors (including histology), the distribution of NITs across the six age subgroups is expected to be similar. Therefore, AUROC values for the classification of patients by age subgroup should be ∼0.5 for NITs that are not impacted by age, and >0.5 for those

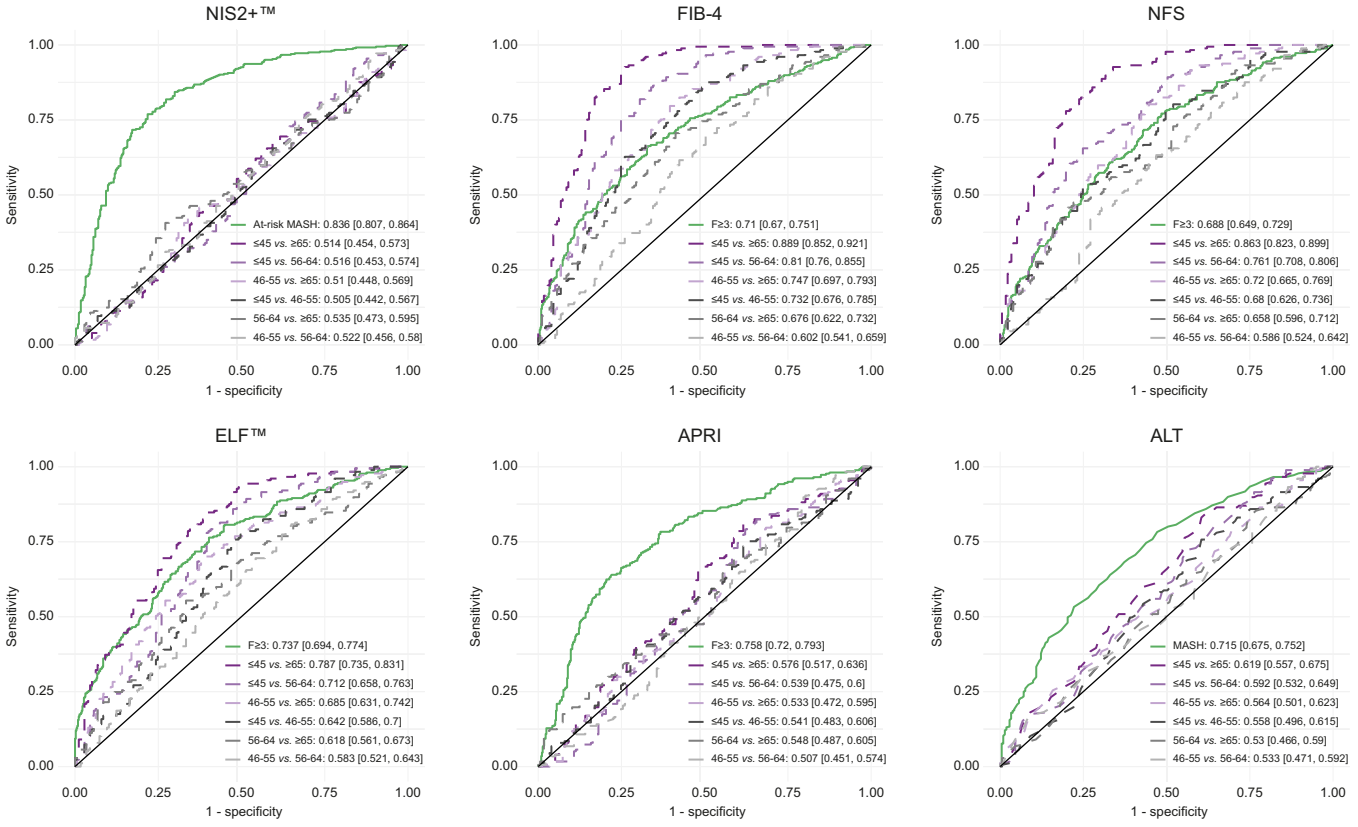


Fig. 1. ROC curves and AUROC values (95% CIs) of NITs for the detection of subpopulations characterized by age or intended histologic endpoints. Matched cohort, N = 708. ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score; NIT, non-invasive test.

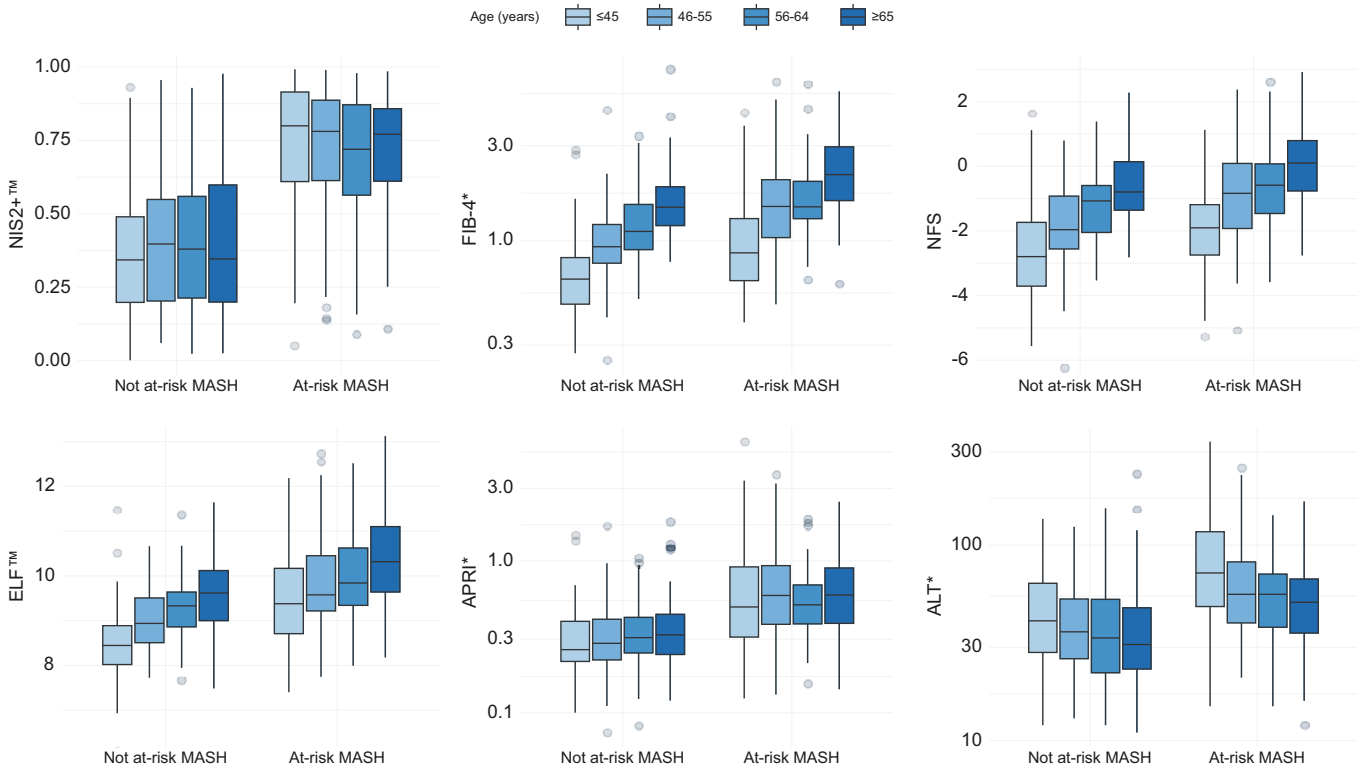


Fig. 2. NIT scores distributions by age and at-risk MASH status. *y-axis on a log10 scale for improved representations. Matched cohort, n = 708. At-risk MASH was defined as MASH with MAS ≥4 and F ≥2. ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; MAS, MASLD activity score; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated liver disease; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score; NIT, non-invasive test.

that are significantly impacted by age. FIB-4 exhibited the highest AUROC values (reaching up to 0.889 for the classification of patients in age subgroups ≤45 *vs.* ≥65 years), followed by NFS (0.823) and ELF™ (0.787). These values were significantly higher than 0.5 independently of the paired age groups considered, corroborating the significant impact of age, even across populations with a mean difference of only 10 years. FIB-4 and NFS exhibited higher AUROC values for the classification of patients into age subgroups characterized by a gap ≥20 years (≤45 *vs.* ≥65 years, ≤45 *vs.* 56–64 years, and 46–55 *vs.* ≥65 years) than for the detection of advanced fibrosis. AUROC values for ELF™ were only higher for the classification of patients into age subgroups than for the detection of advanced fibrosis for the ≤45 *vs.* ≥65 years subgroup. For ALT and APRI, AUROC values were lower for the classification of patients into age subgroups than for the detection of their respective intended histologic endpoints. AUROC values for ALT were significantly higher than 0.5 for the classification of patients into age subgroups with a gap >10 years (≤45 *vs.* ≥65 years, ≤45 *vs.* 56–64 years, and 46–55 *vs.* ≥65 years), while those for APRI were significantly higher than 0.5 only for the classification of patients into the ≤45 *vs.* ≥65 years subgroup. NIS2+™ was the only NIT that exhibited AUROC values that did not significantly differ from 0.5 for the classification of patients into age subgroups (range: 0.505–0.535), demonstrating that age did not impact this NIT regardless of the mean age difference. Furthermore, NIS2+™ achieved an AUROC of 0.836 for the detection of at-risk MASH.

The distribution of NIT scores by age groups and at-risk MASH status is shown in Fig. 2. FIB-4, NFS, and ELF™ returned lower scores in biopsy-proven at-risk MASH patients aged ≤45 years than in patients aged ≥65 years without at-risk MASH, corroborating the significant impact of age on these NITs, which could lead to a substantial bias when using fixed cut-offs to detect advanced disease stage.

To evaluate the impact of age on the clinical performances of NITs at fixed cut-offs, Youden cut-offs for the detection of at-risk MASH were derived using the matched cohort (n = 708) (Table S6, Fig. S3). FIB-4 exhibited major differences in clinical performances, characterized by a strong decrease in specificity (from 0.94 to 0.36) and increase in sensitivity (from 0.26 to 0.90) with increasing age. Age exhibited a similar effect on the clinical performances of NFS and ELF™, while those of NIS2+™ and APRI were stable across age groups.

To evaluate the impact of age on NIT clinical performances, irrespective of selected fixed cut-off values, differences in the sensitivity and specificity of NITs between patients aged ≤45 and ≥65 years were calculated across the entire spectrum of meaningful cut-off values (Fig. 3). Depending on the cut-off value, FIB-4, NFS, and ELF™ exhibited differences higher than 50% in sensitivity and specificity (absolute values) between these two age groups. Specifically, the differences in sensitivity and specificity for FIB-4 with cut-off values of 0.88–1.43 were >50%, reaching a maximum of 74% difference in specificity. NIS2+™ showed the most stable performance among all NITs across the entire range of cut-off values, with maximal differences in sensitivity and specificity <14%.

The impact of age on the clinical performance of NITs in the detection of patients with F ≥3 was also evaluated and shown to

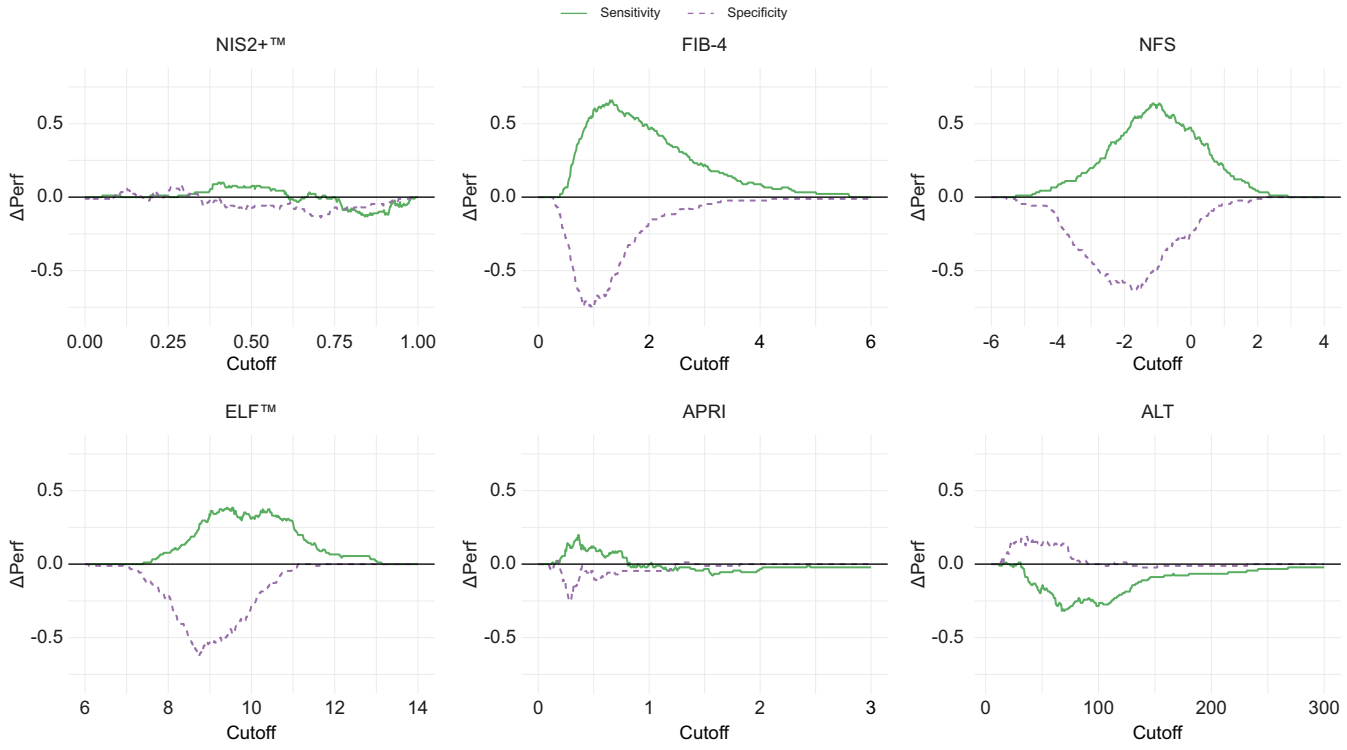


Fig. 3. Differences in clinical performance (sensitivity and specificity) for the detection of at-risk MASH between ≤45 and ≥65 years age groups by cut-off. Matched cohort, n = 708. At-risk MASH was defined as MASH with MAS ≥4 and F ≥2. APRI, aspartate aminotransferase-to-platelet ratio index; ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; MAS, MASLD activity score; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated liver disease; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score.

be similar to that observed in the detection of at-risk MASH (Table S7, Figs S4 and S5).

AUROC values of NITs across age groups were calculated to assess the impact of age on their overall clinical performance for the detection of at-risk MASH (Fig. 4, Table S8). Despite a small variability among AUROC values (up to 5 points difference), none of the comparisons achieved statistical significance, and no trend was observed for any of the NITs, with the exception of ELF™. NIS2+™ achieved the highest AUROC values across age groups (0.821–0.870).

Based on the deficient performances previously reported for NFS and FIB-4 for the population ≤35 years, a supplementary analysis to assess the impact of age on NITs in patients ≤35 *vs.* >35 years was conducted.[12] Of note, the main analysis did not include a group of patients aged ≤35 years, as the low number of patients would have led to a loss of power in statistical analyses, as well as unreliable estimations of AUROC values, sensitivity, and specificity, among other parameters. A well-balanced cohort was selected by applying the PSM algorithm (matched cohort, n = 270; n = 135/age group; Table S9). Overall, the impact of age in patients ≤35 years was similar to that observed in the main analysis, with an extension/continuation of trends previously observed in the distribution of NITs and biomarkers across age groups (Tables S10 and S11). Key outcomes of this supplementary analysis are included in the 'Supplementary Results' section.

Discussion

The burden of MASLD, a condition that can progress to MASH, affects patients of all ages. It is therefore critical for NITs used in the evaluation of liver disease (MASH, at-risk MASH, advanced fibrosis) to exhibit robust clinical performances across age groups. Age has been shown to impact FIB-4, NFS, and ELF™, suggesting that they should be used with age-adapted cut-offs, which would need to be further derived and validated, representing a potential challenge.[14,21] While NIS2+™, an optimization of NIS4® technology for the detection of at-risk MASH in patients with metabolic risk factors, exhibited robust clinical performances in patients aged ≤50 *vs.* ≥60 years, additional data were needed to further characterize its performance across age groups.[5,10] This report thoroughly evaluated the potential confounding role of age across different NITs commonly used in clinical practice for the evaluation of liver disease, mainly among patients with MASLD/MASH.

Neither NIS2+™ nor APRI were impacted by age, which allows for their implementation with fixed cut-offs, irrespective of the patient's age. Overall, for NIS2+™ (Table S2), the mean levels of YKL-40 increased with age, while those of miR34-a decreased with age, leading to steady mean scores across age groups and confirming preliminary observations.[10] The mean scores of APRI, which positively correlated with AST levels and inversely correlated with platelet levels (Table S2), remained stable across groups due to the decrease in mean AST and platelet levels with age.

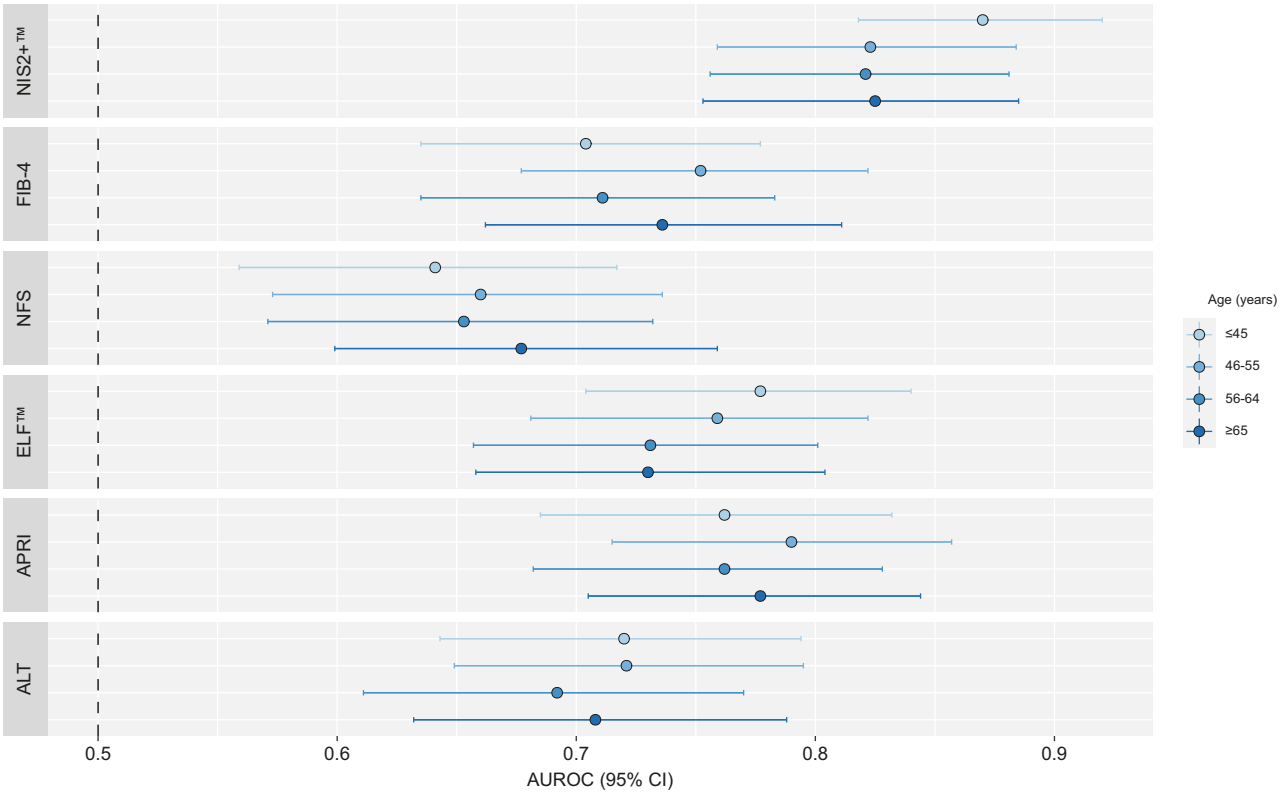


Fig. 4. AUROC values for at-risk MASH detection across age groups. Matched cohort, n = 708. At-risk MASH was defined as MASH with MAS ≥4 and F ≥2. ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; ELF™, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; MAS, MASLD activity score; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated liver disease; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score.

Age significantly impacted the performance of FIB-4, NFS, and ELF™. These results confirm previously reported outcomes on FIB-4 and NFS, supporting the use of age-adapted cut-off values for these NITs.[14]

The effect of age on the mean scores of FIB-4 and NFS can be explained by the positive correlation between these scores (Table S2) and age, as well as by their negative correlation with ALT and platelet levels, which generally decrease with age. Furthermore, NFS negatively correlated with albumin levels, which generally decrease with age. While the mean levels of the three biomarkers associated with ELF™ (Table S2) reached their highest values in patients aged ≥65 years, hyaluronic acid exhibited a significant and consistent increase (from 38.65 ng/ml to 126.76 ng/ml; p <0.0001; Table 2), acting as the driving factor behind the impact of age on the mean scores of this NIT.

Using well-designed multi-way type II ANOVA, the effect size of age on FIB-4 and NFS was higher than that of fibrosis, their theoretical main driver. Furthermore, most AUROC values for age subgroup classifications were higher for these NITs than for the detection of their histologic endpoint (F ≥3).

Overall, among the NITs evaluated, NIS2+™ showed the strongest association with histology and was the only test on which fibrosis and MAS exerted a moderate to large effect, supporting its role as the most suitable panel to detect the composite endpoint of at-risk MASH.

Age also impacted clinical performances of FIB-4, NFS, and ELF™ for the detection of at-risk MASH or F ≥3, which could result in up to a 50% difference in sensitivity/specificity between the ≤45 and ≥65 years age groups. On the other hand, the clinical performances of APRI and NIS2+™ were not impacted by age.

While this study was not powered to allow for the inclusion of a stand-alone age group comprising patients aged ≤35 years in the main analysis, a supplementary analysis that evaluated the impact of age in patients ≤35 *vs.* >35 years corroborated the outcomes of the main analysis (similar trends/significance were observed on NITs and biomarkers), and showed that neither NFS, FIB-4, nor ELF™ should be used/interpreted using their published cut-off values in patients aged ≤35 years.

Unlike NIS2+™, neither FIB-4, NFS, nor ELF™ were specifically designed to detect at-risk MASH, which represents a limitation of this study. Of note, however, the impact of age on the clinical performances of NFS, FIB-4, and ELF™ for the detection of advanced fibrosis was similar to that observed in the detection of at-risk MASH. While evaluation of the impact of age on NIT scores in patients aged <18 years would be of interest, the analysis cohort did not include this subpopulation. Lastly, the matched subpopulations selected through the implementation of the PSM algorithm within each age category are not representative of real-world subpopulations of the same age range. Age usually correlates positively with fibrosis in most MASLD clinical databases. Age also positively correlates with FIB-4 and NFS based on their formulas (Table S2); furthermore, ELF™ scores increase with age due to the impact of this parameter on its biomarkers (mainly hyaluronic acid). Therefore, age

contributes to the ability of these NITs to differentiate patients with F ≥3 from those with F0–F2, improving their estimated clinical performances by acting as a confounding factor. While it is likely that the implementation of the PSM in the analysis reported here led to lower AUROC values for the detection of F ≥3 compared with values that would have been obtained without using this algorithm, it allows for a more precise estimation of their ability to exclusively detect histologic endpoints. The goal of this analysis was to quantify the impact of age on the distribution of NITs in an independent manner and compare these effects with those of histology (fibrosis and MAS). These thorough assessments require that the age-based subpopulations be similar with respect to selected factors, except for age, to control for potential confounding factors.

In summary, NIS2+^{TM} was the only NIT among those tested that was not impacted by age and was sensitive to both fibrosis and MAS, further confirming that this NIT constitutes an efficient and robust test for the detection of the composite endpoint of at-risk MASH. Given the unmet need for NITs with performances that are not impacted by age, the demonstrated robust performance of NIS2+^{TM} to rule in and rule out at-risk MASH with fixed cut-offs across age groups highlights its potential to improve management of patients and interpretation of results, facilitating the large-scale use of this test.

Abbreviations

ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; ELF^{TM}, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated liver disease; MAS, MASLD activity score; NFS, NAFLD fibrosis score; NITs, non-invasive tests; PSM, propensity score-matching; T2DM, type 2 diabetes mellitus.

Financial support

The funder of this study, Genfit S.A., had a role in the study design, data collection, data analysis, data interpretation, and writing of the report. All authors and the funder had full access to all aggregate data in the study and had full responsibility for the decision to submit the manuscript.

Conflicts of interest

QMA: research support from LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) consortium funded by the Innovative Medicines Initiative Program of the European Union under Grant Agreement 777377 (this multistakeholder consortium includes industry partners and received funding from EFPIA); grants or contracts from AstraZeneca, Boehringer Ingelheim, and Intercept Pharmaceuticals; royalties or licenses from Elsevier Ltd.; consulting fees (on behalf of Newcastle University) from Alimentiv, Akero Therapeutics, Inc., AstraZeneca, Axcella Health, Inc., 89bio, Inc., Boehringer Ingelheim, Bristol Myers Squibb, Galmed Pharmaceuticals, Genfit S.A., Genentech, Gilead Sciences Inc., GlaxoSmithKline, Hanmi, HistoIndex Pte Ltd., Intercept Pharmaceuticals, Inventiva, Ionis, IQVIA, Janssen, Madrigal Pharmaceuticals, Medpace, Merck, NGM Biopharmaceuticals, Novartis Pharmaceuticals, Novo Nordisk, PathAI, Pfizer, Prosciento, Poxel S.A., Resolution Therapeutics, Roche, Ridgeline Therapeutics, RTI, Shionogi, and Terns Pharmaceuticals; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal Pharmaceuticals, Medscape, and Springer Healthcare; served on advisory boards or data safety monitoring boards (on behalf of Newcastle University) for Medpace (NorthSea Therapeutics B.V., DSMB). BS: consulting fees from Genfit S.A. MAC: Labcorp employee. RL: grants/funding support from the National Center for Advancing Translational Sciences, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Heart, Lung, and Blood Institute, Arrowhead Pharmaceuticals, AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, Galectin Therapeutics, Inc., Galmed Pharmaceuticals, Gilead Sciences, Inc., Hanmi, Intercept Pharmaceuticals, Inventiva, Ionis, Janssen, Madrigal Pharmaceuticals, Merck, NGM Biopharmaceuticals, Novo Nordisk, Pfizer, Sonic Incytes, and Terns Pharmaceuticals; consulting fees from Aardvark Therapeutics, Altimmune, Anylam/Regeneron, Amgen, Arrowhead Pharmaceuticals, AstraZeneca, Bristol Myers Squibb, CohBar, Inc., Eli Lilly, Galmed Pharmaceuticals, Gilead Sciences, Inc., Glympse Bio, HighTide Therapeutics, Ini Pharma, Intercept Pharmaceuticals, Intercept Pharmaceuticals, Ionis, Janssen, Madrigal Pharmaceuticals, Metacrine, NGM Biopharmaceuticals, Novartis Pharmaceuticals, Novo Nordisk, Merck, Pfizer, Sagimet Biosciences, Theratechnologies, Inc., 89bio, Inc., Terns Pharmaceuticals, and Viking Therapeutics; other financial interests: LipoNexus, Inc. (cofounder). SAH: grants or contracts from Akero Therapeutics, Alentis Therapeutics, Altimmune, B. Riley FBR, ChronWell, Corcept Therapeutics, Echosens, Axcella Health, Cirius Therapeutics, CiVi Biopharma, Cymabay Therapeutics, Inc., Enyo Pharma S.A., Galectin Therapeutics, Inc., Galmed Research & Development, Ltd., Genfit S.A., Gilead Sciences, Inc., Hepion Pharmaceuticals, Inc., Hepta Bio, HighTide Therapeutics, Inc, HistoIndex, Intercept Pharmaceuticals, Ionis, Madrigal Pharmaceuticals, Medpace, NGM Biopharmaceuticals, Inc., NeuroBo, NorthSea Therapeutics B.V., Novartis Pharmaceuticals, Novo Nordisk, Path AI, Perspectum, Poxel S.A., Sagimet Biosciences, Sonic Incytes, Terns Pharmaceuticals, and Viking Therapeutics; stock or stock options for Akero Therapeutics, ChronWell, Cirius Therapeutics, Galectin Therapeutics, Inc., Genfit S.A., Hepion Pharmaceuticals, Inc., HistoIndex Pte Ltd., Metacrine, NGM Biopharmaceuticals, Inc., NorthSea Therapeutics B.V. VR: grants or contracts from Intercept Pharmaceuticals, and Gilead Sciences, Inc.; consulting fees from Boehringer Ingelheim, Novo Nordisk, Poxel S.A., Enyo Pharma S.A., Madrigal Pharmaceuticals, Terns Pharmaceuticals, Intercept Pharmaceuticals, NGM Biopharmaceuticals Inc., and Pfizer. AJS: stock options in Genfit S.A., Tiziana, Indalo, Durect, Inversago, and Galmed Pharmaceuticals; consultant to AstraZeneca, Salix, Tobira, Takeda, Jannsen, Gilead Sciences, Inc., Terns Pharmaceuticals, Merck, Madrigal Pharmaceuticals, NGM Biopharmaceuticals, Inc., Sagimet Biosciences, Valeant, Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, Hemoshear, Novartis Pharmaceuticals, Inventiva, Enyo, Akero Therapeutics, 89bio, Inc., Novo Nordisk, Pfizer, Amgen, Genentech, Regeneron, Alnylam, Hanmi, LG Chem, Histoindex, Thera Technologies, Intercept Pharmaceuticals, Target-RWE, Surrozen, Zydus, Path AI, Exhalenz, and Genfit S.A.; his institution has received grant support from Gilead Sciences, Inc., Salix, Tobira, Bristol Myers Squibb, Pfizer, Intercept Pharmaceuticals, Merck, AstraZeneca, Mallinckrodt, and Novartis Pharmaceuticals; royalties received from Elsevier and UpToDate. JM, YH, AC, ZM, CR, and DWH: stock or stock options from Genfit S.A. and serve as Genfit S.A. employees.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

JM conceived and designed the study and performed data acquisition and analyses. QMA, JM, YH, AC, CR, ZM, DWH, and BS contributed to data interpretation and editing. All authors critically reviewed the manuscript and approved the final manuscript prior to submission.

Data availability statement

Data analyzed in this study are available upon reasonable request.

Acknowledgments

Medical writing assistance was provided by Syneos Health and supported by Genfit S.A.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2024.101011.

References

Author names in bold designate shared co-first authorship

[1] Younossi ZM. Non-alcoholic fatty liver disease – a global public health perspective. J Hepatol 2019;70(3):531–544.

[2] European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016;64(6):1388–1402.

[3] Rinella ME, Lazarus JV, Ratziu V, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol 2023; S0168-8278(23)00418-X. https://doi.org/10.1016/j.jhep.2023.06.003. Online ahead of print.

[4] Wong T, Wong RJ, Gish RG. Diagnostic and treatment implications of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Gastroenterol Hepatol (NY) 2019;15(2):83–89.

[5] **Harrison SA**, **Ratziu V**, Boursier J, et al. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020;5(11):970–985.

[6] Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol 2014;20(2):475–485.

[7] Brunt EM, Clouston AD, Goodman Z, et al. Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. J Hepatol 2022;76:1030–1041.

[8] **Vali Y**, **Lee J**, Boursier J, et al. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. Lancet Gastroenterol Hepatol 2023. https://doi.org/10.1016/S2468-1253(23)00017-1. published online March 20.

[9] Sanyal AJ, Shankar SS, Calle RA, et al. Non-invasive biomarkers of nonalcoholic steatohepatitis: the FNIH NIMBLE project. Nat Med 2022;28(3):430–432.

[10] **Harrison SA**, **Ratziu V**, Magnanensi J, et al. NIS2+TM, an optimization of the blood-based biomarker NIS4® technology for the detection of at-risk NASH: a prospective derivation and validation study. J Hepatol 2023;79(3):758–767.

[11] **Ratziu V**, **Harrison SA**, Hajji Y, et al. NIS2+™ as a screening tool to optimize patient selection in metabolic dysfunction-associated steatohepatitis clinical trials. J Hepatol 2024;80(2):209–219.

[12] National Library of Medicine (US). Non-alcoholic fatty liver disease. Bethesda (MD): MedlinePlus; 2016. Updated November 1, https://medlineplus.gov/genetics/condition/non-alcoholic-fatty-liver-disease/. [Accessed 21 December 2022].

[13] U.S. Food & Drug Administration. Evaluation and reporting of age-, race-, and ethnicity-specific data in medical device clinical studies. Guidance Industry Food Drug Adm Staff September 2017; https://www.fda.gov/regulatory-information/search-fda-guidance-documents/evaluation-and-reporting-age-race-and-ethnicity-specific-data-medical-device-clinical-studies. [Accessed 21 December 2022].

[14] McPherson S, Hardy T, Dufour J-F, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol 2017;112(5):740–751.

[15] 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(2):236–242.

[16] Ho D, Imai K, King G, et al. MatchIt: nonparametric preprocessing for parametric causal inference. J Stat Softw 2011;42(8):1–28.

[17] Lix LM, Keselman JC, Keselman HJ. Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance *F* test. Rev Educ Res 1996;66(4):579–619.

[18] Curran PJ, West SG, Finish JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychol Methods 1996;1(1):16–29.

[19] Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013;4:863.

[20] Kirk RE. Practical significance: a concept whose time has come. Educ Psychol Meas 1996;56(5):746–759.

[21] Ishiba H, Sumida Y, Tanaka S, et al. The novel cutoff points for the FIB4 index categorized by age increase the diagnostic accuracy in NAFLD: a multi-center study. J Gastroenterol 2018;53(11):1216–1224.