

FIB-4 is not one size fits all: real-life accuracy and misclassification in a multicentre italian hepatology cohort

Felice Cinque

University of Milan: Università degli Studi di Milano

Matteo Mureddu

University of Milan: Università degli Studi di Milano

Annalisa Cespiati

Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico

Anna Mantovani

Azienda Ospedaliera Universitaria Integrata Verona

Nicola Pugliese

Humanitas University

Grazia Pennisi

University of Palermo: Università degli Studi di Palermo

Rosanna Villani

University of Foggia: Università degli Studi di Foggia

Mario Masarone

University of Salerno: Università degli Studi di Salerno

Pietro Torre

University of Salerno: Università degli Studi di Salerno

Federico Ravaoli

University of Bologna: Università di Bologna

Silvia Ferri

Università degli Studi di Bologna Azienda Ospedaliera di Bologna Policlinico S Orsola-Malpighi: IRCCS Azienda Ospedaliero-Universitaria di Bologna Policlinico di Sant'Orsola

Paolo Gallo

Campus Bio-Medico University Hospital: Policlinico Universitario Campus Bio-Medico

Francesca Terracciani

Campus Bio-Medico University Hospital: Policlinico Universitario Campus Bio-Medico

Elena Buzzetti

University of Modena and Reggio Emilia: Università degli Studi di Modena e Reggio Emilia

Alessandro Mantovani

Università degli Studi di Verona: Università degli Studi di Verona

Salvatore Petta

University of Palermo: Università degli Studi di Palermo

Gaetano Serviddio

University of Foggia: Università degli Studi di Foggia

Alessio Aghemo

Humanitas University

Andrea Dalbeni

Azienda Ospedaliera Universitaria Integrata Verona

Fabio Piscaglia

Università di Bologna: Università di Bologna

Umberto Vespasiani-Gentilucci

Campus Bio-Medico University Hospital: Policlinico Universitario Campus Bio-Medico

Marcello Persico

University of Salerno: Università degli Studi di Salerno

Anna Ludovica Fracanzani

Università degli Studi di Milano: Università degli Studi di Milano

Rosa Lombardi

rosa.lombardi@unimi.it

Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico <https://orcid.org/0000-0001-6958-927X>

Research Article

Keywords: liver stiffness measurement, diagnostic accuracy, fibrosis screening, risk stratification, hepatology referral

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10566104/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Detection of fibrosis in metabolic dysfunction–associated steatotic liver disease (MASLD) is crucial for referral to hepatology care. Current guidelines recommend a two-step strategy based on FIB-4 (cut-offs 1.3 or 2.0 if > 65 years) followed by liver stiffness measurement (LSM $\geq$ 8 kPa). We assessed the real-world FIB-4 performance in a large Italian MASLD cohort and identified factors determining misclassification.

Methods

We enrolled 2638 ultrasound-proven MASLD patients from 9 Italian hepatological centers at their first evaluation. All subjects underwent simultaneous Fibroscan® and FIB-4 calculation within 6 months from referral.

Results

LSM $\geq$ 8 kPa in 21.9% of patients. Among those with FIB-4 < 1.3/2.0, 13.2% were false negatives (LSM $\geq$ 8 kPa), 58.7% of patients with FIB-4 $\geq$ 1.3/2.0 were false positives (LSM < 8 kPa). Diabetes, hypertension, overweight and obesity, elevated ALT, and grade 3 steatosis independently predicted false-negative results, whereas older age, diabetes, and elevated ALT reduced the likelihood of false-positive classification. FIB-4 showed 59% sensitivity, 77% specificity, 87% negative predictive value, and an AUROC of 0.72 (95% CI 0.69–0.74). Diagnostic accuracy improved with age (AUROC 0.59 in patients < 45 years vs 0.77 in $\geq$ 65 years), declined with increasing BMI (0.82 in lean vs 0.71 in obesity), and in severe steatosis (0.64 vs 0.76). Lower rule-out thresholds were required in younger, obese, and severe steatosis patients.

Conclusions

FIB-4 showed only moderate performance for detecting elevated LSM in MASLD. Patients with multiple cardiometabolic factors were frequently misclassified, supporting the need for a direct fibrosis assessment in high-risk individuals and for individualized FIB-4 thresholds.

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD), defined by hepatic steatosis and at least one cardiometabolic risk factor, is the leading cause of chronic liver disease worldwide. It affects approximately 30% of the general population and up to 60% and 90% of individuals with type 2 diabetes (T2DM) and obesity¹. MASLD encompasses a spectrum ranging from simple steatosis to metabolic dysfunction-associated steatohepatitis (MASH) and advanced fibrosis, the main determinant of liver-related morbidity and mortality².

Although liver biopsy remains the reference standard for fibrosis staging, its invasiveness, cost and potential complications limit widespread use. Several non-invasive tests have therefore been developed, including serum biomarkers and imaging-based techniques such as vibration-controlled transient elastography (VCTE; FibroScan®)³. The Fibrosis-4 (FIB-4) score⁴ is the most widely used serum-based test because it is simple, inexpensive and readily available. Based on age, platelet count, AST and ALT levels, it applies cut-offs of 1.3 (2.0 in individuals aged > 65 years) and 2.67 to rule out and rule in advanced fibrosis, respectively, achieving AUROCs of 0.77–0.815 in large multicentre studies^{5,6}.

Since September 2024 international guidelines^{3,7} recommend the use of FIB-4 in primary care to improve case findings of advanced liver fibrosis, especially in high-risk individuals, such as those with T2DM, obesity or elevated liver enzymes. Briefly, current guidelines propose a two-step algorithm based on FIB-4 followed by FibroScan®: individuals FIB-4 < 1.3 do not require hepatology referral and should undergo periodic reassessment of the score, whereas those with FIB-4 ≥ 1.3 should undergo intensified management of comorbidities and/or second-line non-invasive testing with FibroScan®. Those with increased liver stiffness measurement (LSM ≥ 8 kPa) or with FIB-4 ≥ 2.67 are referred to a hepatology centre^{8,9}. However, the predictive values of FIB-4 depend on fibrosis prevalence and therefore vary across clinical settings¹⁰. Emerging evidence suggests that standard thresholds may be suboptimal in high-risk metabolic populations, potentially missing clinically relevant fibrosis, particularly in patients with MASLD and T2DM^{11–13}.

With the advent of pharmacological treatments for non-cirrhotic MASH, accurate fibrosis detection is increasingly important. Treatment eligibility has recently been proposed for patients with LSM ≥ 10 kPa¹⁴, while optimization of obesity and T2DM remains central to MASLD management⁷.

This multicentre study therefore aimed to evaluate the performance of the guideline-recommended two-step algorithm in a real-world Italian tertiary hepatology setting, identify factors associated with misclassification, and assess FIB-4 performance across clinically relevant subgroups.

Materials and Methods

Subject selection

All consecutive patients with US-proven MASLD at their first evaluation at 9 Italian tertiary hepatology centres were retrospectively enrolled over the last 5 years up preceding January 2025. Patients underwent simultaneous Fibroscan® and FIB-4 calculation within 6 months from referral.

Definition of MASLD and other metabolic comorbidities

MASLD was defined by the presence of hepatic steatosis plus at least one metabolic comorbidity³. Subjects with other causes of steatotic liver disease and with alcohol consumption > 30/20

g/day in men/women were excluded.

Hepatic steatosis was diagnosed and graded by ultrasonography (US) using a convex 3.5 MHz probe.

Insulin resistance was diagnosed by a Homeostasis Model Assessment of Insulin Resistance (HOMA-IR) > 2.5. Impaired renal function was defined by an estimating the glomerular filtration rate (GFR) < 60 mL/min/1.73 m² using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI)

Fibrosis assessment

FIB-4 was calculated using available blood tests at the time of inclusion in the study according to the validated formula⁴. FibroScan® was performed by one expert physician for each centre for the determination of the LSM using the M probe or the XL probe in case of unsuccessful measurements. The final value was obtained using standardized procedures and LSM ≥8kPa defined the presence of significant fibrosis³, whereas LSM ≥15kPa was considered highly suggestive of compensated advanced chronic liver disease (cACLD)¹⁵.

Statistical analysis

Continuous and categorical variables were summarized using standard descriptive statistics and compared with appropriate parametric or non-parametric tests. Missing data were handled by complete-case analysis. FIB-4 diagnostic performance was assessed using sensitivity, specificity, PPV, NPV and AUROC. Misclassification was defined as false negatives (LSM ≥ 8 kPa and FIB-4 < 1.3 or 2.0 if > 65 years) and false positives (LSM < 8 kPa and FIB-4 ≥ 1.3 or 2.0 if > 65 years). Factors associated with misclassification were investigated using centre-adjusted logistic regression.

Diagnostic performance was also evaluated across clinically relevant subgroups, with AUROCs compared using the DeLong test according to: (1) age, sex and metabolic comorbidities; (2) severity of liver disease; (3) enrollment from September 2024 onward and to the presence of risk factors for advanced liver disease^{3,7}. A two-sided p-value < 0.05 was considered significant. Analyses were performed using SPSS 30.0 and R.

Results

Demographic and metabolic features

A total of 2638 subjects with US-defined MASLD were enrolled from nine Italian tertiary hepatology centres (Supplementary table 1 and 2): Milan (n = 562), Bologna (n = 140), Verona (n = 409), Rome (n = 638), Foggia (n = 158), Humanitas (n = 186), Modena (n = 268), Palermo (n = 145) and Salerno (n = 132). Overall, 712 (27%) patients were enrolled from September 2024 onwards.

Mean age was 55 ± 13 years and the 61.2% were males. According to BMI (mean 30.7 ± 5.3 kg/m²) 13.3% of patients were lean, 35.5% overweight and 51.2% obese. Increased WC was present in 88.9% of patients. Hypertension, dyslipidaemia and T2DM affected 53.7%, 65.6%, and 27.1% of the cohort. According to current guidelines, 2007 (76.1%) patients fulfilled the criteria for strict algorithm applicability.

Significant heterogeneity across centres was observed for most demographic and clinical characteristics (Supplementary Table 2), including age, metabolic comorbidities (BMI, hypertension, dyslipidaemia, T2DM) and reduced GFR (< 60 ml/min).

Liver disease features

Liver disease characteristics of the study population are summarized in Supplementary Table 1.

Severe steatosis (grade 3) was present in 25.5% of patients and 21.9% ($n = 579$) showed $\text{LSM} \geq 8$ kPa (median LSM 5.5 kPa (IQR 4.4–7.5)). $\text{LSM} \geq 10$ kPa and ≥ 15 kPa was in 14% and 7.3% of the cohort. Overall, 1815 (68.8%) patients had $\text{FIB-4} < 1.3/2.0$, 823 (31.2%) $\geq 1.3/2.0$ and 217 (8.2%) had $\text{FIB-4} > 2.67$.

Elevated AST, ALT, and GGT were observed in 25.6%, 42.9%, and 34.2% of the cohort, respectively, while 45.6% had at least one elevated aminotransferase.

Liver disease severity varied significantly across centres (Supplementary Table 2). The prevalence of $\text{FIB-4} \geq 1.3/2.0$ ranged from 16.4% in Milan to 53.1% in Palermo ($p < 0.001$) and $\text{LSM} \geq 8$ kPa from 7.5% to 73.8% in the same centres (both $p < 0.001$).

Performance of FIB4 as first line strategy to detect fibrosis

FIB-4 classified 1815 (68.8%) of patients as low risk ($< 1.3/2.0$) and among them, 239 (13.2%) were false negatives ($\text{LSM} \geq 8$ kPa). Conversely, FIB-4 was $\geq 1.3/2$ in 823 (31.2%) patients, but 483 (58.7%) were false positives ($\text{LSM} < 8$ kPa). Overall, FIB-4 showed good specificity (77%) and NPV (87%), but suboptimal sensitivity (59%), with an AUROC of 0.72 (95% CI 0.69–0.74) for the identification of $\text{LSM} \geq 8$ kPa (Fig. 1A).

When considering $\text{LSM} \geq 10$ kPa, FIB-4 presented a superimposable performance, with a good specificity (74.8%) and NPV (93.4%), suboptimal sensitivity (67.8%), with an AUROC of 0.77 (95% CI 0.74–0.79). (Supplementary Fig. 1A). Interestingly, in patients with $\text{FIB-4} < 1.3/2.0$, 119 (6.6%) had $\text{LSM} \geq 10$ kPa. When considering $\text{LSM} \geq 15$ kPa, FIB-4 presented a better performance, with very good specificity (86%) and negative predictive value (98%), but a very low sensitivity (26%), with an AUROC of 0.84 (95% CI 0.81–0.87) (Supplementary Fig. 1B).

Factors affecting false negative referrals

Among patients with FIB-4 < 1.3/2 (n = 1815), 239 (13.2%) showed LSM $\geq$ 8 kPa, representing false negatives.

Among patients with FIB-4 < 1.3/2, no relevant difference in age distribution was observed between those with LSM < 8 kPa and those with LSM $\geq$ 8 kPa (p = 0.154). Conversely, a more adverse metabolic phenotype was observed in patients with LSM $\geq$ 8 kPa, including a higher prevalence of obesity, increased WC, hypertension, T2DM and insulin resistance. No difference in the prevalence of dyslipidemia was observed. (Table 1).

Table 1

Baseline characteristics of patients classified as low risk by FIB-4 (< 1.3, or < 2.0 if aged > 65 years), stratified according to LSM < 8 kPa versus LSM ≥ 8 kPa.

Variable	LSM < 8 kPa (n = 1576)	LSM ≥ 8 kPa (n = 239)	p-value
Male sex, n (%)	959 (60.9)	141 (59.0)	0.585
Age, years	52.6 ± 13.3	54.3 ± 13.5	0.073
< 45 years, n(%)	440 (27.9)	58(24.3)	0.154
45–55 years, n(%)	452(28.7)	62(25.9)	
56–65 years, n(%)	382(24.2)	59(24.7)	
> 65 years, n(%)	302(19.2)	60(25.1)	
BMI, kg/m ²	30.1 ± 5.1	33.7 ± 6.0	< 0.001
< 25 kg/m ² , n(%)	243(15.4)	6(2.6)	< 0.001
≥ 30 kg/m ² , n(%)	750 (47.6)	163 (68.2)	< 0.001
Waist circumference, cm	104.1 ± 12.1	112.1 ± 12.8	< 0.001
> 80/94 cm M/F, n(%)	1373 (87.1)	234 (97.9)	< 0.001
Hypertension, n (%)	730 (46.3)	154 (64.4)	< 0.001
Dyslipidemia, n (%)	1040 (66.0)	159 (66.5)	0.879
T2DM, n (%)	293 (18.6)	102 (42.7)	< 0.001
HbA1c ≥ 53 mmol/mol, n (%)	107 (6.8)	67 (28.1)	< 0.001
HOMA-IR > 2.5, n (%)	898 (57.0)	181 (75.6)	< 0.001
Increased AST, n (%)	235 (14.9)	62 (25.9)	< 0.001
Increased ALT, n (%)	625 (39.7)	127 (53.1)	< 0.001
Increased GGT, n (%)	430 (27.3)	92 (38.6)	< 0.001
Elevated transaminases, n (%)	630 (40.0)	128 (53.6)	< 0.001
eGFR < 60 mL/min/1.73m ² , n (%)	52 (3.3)	10 (4.3)	0.482
Severe steatosis (grade 3), n (%)	348 (22.1)	108 (45.2)	< 0.001
Abbreviations:ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; eGFR, estimated glomerular filtration rate; FIB-4, Fibrosis-4 index; GGT, gamma-glutamyl transferase; HbA1c, glycated hemoglobin; HOMA-IR, homeostatic model assessment of insulin resistance; LSM, liver stiffness measurement; T2DM, type 2 diabetes mellitus; WC, waist circumference.			

Table 2

Baseline characteristics of patients classified as at increased risk by FIB-4 (≥ 1.3 , or ≥ 2.0 if aged > 65 years), stratified according to LSM < 8 kPa versus LSM ≥ 8 kPa.

Variable	LSM < 8 kPa (n = 483)	LSM ≥ 8 kPa (n = 340)	p-value
Male sex, n (%)	181 (37.5)	127 (37.4)	0.972
Age, years	60.6 $\pm$ 9.3	63.1 $\pm$ 9.7	< 0.001
< 45 years, n(%)	25(5.2)	14(4.1)	< 0.001
45–55 years, n(%)	97(20.1)	58(17.1)	
56–65 years, n(%)	247(51.1)	134(39.4)	
> 65 years, n(%)	114(23.6)	134(39.4)	
BMI, kg/m ²	29.9 $\pm$ 4.8	32.3 $\pm$ 5.6	< 0.001
< 25 kg/m ² , n(%)	71 (14.8)	31 (9.2)	< 0.001
≥ 30 kg/m ² , n(%)	221 (45.8)	219 (64.3)	< 0.001
Waist circumference, cm	103.7 $\pm$ 12.0	110.6 $\pm$ 13.4	< 0.001
> 80/94 cm M/F, n(%)	424 (87.8)	326 (95.8)	0.007
Hypertension, n (%)	296 (61.3)	238 (70.0)	0.010
Dyslipidemia, n (%)	311 (64.5)	232 (68.2)	0.285
T2DM, n (%)	126 (26.0)	194 (57.1)	< 0.001
HbA1c ≥ 53 mmol/mol, n (%)	24 (4.9)	78 (23.1)	< 0.001
HOMA-IR > 2.5 , n (%)	258 (53.5)	278 (81.8)	< 0.001
Increased AST, n (%)	162 (33.5)	216 (63.5)	< 0.001
Increased ALT, n (%)	179 (37.1)	200 (58.8)	< 0.001
Increased GGT, n (%)	157 (32.6)	223 (65.7)	< 0.001
Elevated transaminases, n (%)	209 (43.3)	235 (69.1)	< 0.001
eGFR < 60 mL/min/1.73m ² , n (%)	24 (4.9)	34 (10.0)	0.015
Severe steatosis (grade 3), n (%)	108 (22.3)	115 (33.9)	0.002
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; eGFR, estimated glomerular filtration rate; FIB-4, Fibrosis-4 index; GGT, gamma-glutamyl transferase; HbA1c, glycated hemoglobin; HOMA-IR, homeostatic model assessment of insulin resistance; LSM, liver stiffness measurement; T2DM, type 2 diabetes mellitus; WC, waist circumference.			

Markers of liver injury were significantly more pronounced in patients with LSM $\geq$ 8 kPa, including higher prevalence of elevated AST, ALT, GGT and severe steatosis.

At multivariate analysis (Fig. 2), T2DM (OR 2.75; 95% CI: 1.95–3.87), hypertension (OR 1.59; 95% CI: 1.12–2.25), overweight (OR 4.4, 95% CI 1.84–10.53), obesity (OR 6.65, 95% CI 2.83–15.62), increased ALT (OR 1.81, 95% CI 1.30–2.51) and steatosis grade 3 (OR 1.73; 95% CI: 1.17–2.54) were independently associated with misclassification.

Factors affecting false positive referrals

Among patients with FIB-4 $\geq$ 1.3/2 (n = 823), 483 (58.6%) had LSM < 8 kPa, representing false-positive cases.

Among patients with FIB-4 $\geq$ 1.3/2, a significant difference in age distribution was observed between those with LSM < 8 kPa and those with LSM $\geq$ 8 kPa, as well as a more favourable metabolic profile, with lower prevalence of obesity, hypertension, T2DM, insulin resistance and impaired renal function. (Table 1).

Similarly, markers of liver injury were significantly less pronounced in patients with LSM < 8 kPa, including lower prevalence of elevated AST, ALT and GGT and severe steatosis.

At multivariate analysis (Fig. 3), age (OR 0.97; 95% CI: 0.95–0.99), T2DM (OR 0.27; 95% CI: 0.19–0.40) and increased ALT (OR 0.43, 95% CI 0.29–0.64) were independently inversely associated with LSM<8kPa.

Stratified analysis of the diagnostic performance of FIB4

As shown in supplementary table 3, both the diagnostic performance of FIB-4 and the cut-offs required to achieve 80% sensitivity and 95% specificity varied across clinically relevant subgroups.

• Age and sex

Diagnostic accuracy increased progressively with age, with AUROC values of 0.59 (95% CI 0.51–0.66) in patients < 45 years, 0.65 (95% CI 0.59–0.72) in those aged 45–55 years, 0.72 (95% CI 0.68–0.76) in those aged 56–65 years, and 0.77 (95% CI 0.73–0.81) in those aged $\geq$ 65 years (Fig. 1B). (De Long test: patients < 45 vs 45–55 years (p = 0.002) and vs $\geq$ 65 years (p < 0.001); patients 45–55 years vs $\geq$ 65 years (p < 0.003)) (Fig. 1B). FIB4 cut-off for sensitivity at 80% and specificity at 95% in patients aged < 45 years were 0.47 and 1.3, 45–55 years 0.79 and 1.68, 56–65 years 1.17 and 2.29 and $\geq$ 65 years 1.75 and 3.34.

Conversely, performance was comparable between sexes (female: AUROC 0.72; 95% CI 0.69–0.75; male: AUROC 0.71; 95% CI 0.67–0.76; DeLong test p = 0.829) (Fig. 1C).

• Metabolic comorbidities

The diagnostic performance of FIB-4 progressively declined with increasing BMI, with the highest AUROC observed in lean individuals (AUROC: 0.82 95% CI 0.74–0.91), intermediate values in overweight patients (AUROC: 0.73 95% CI 0.68–0.78), and the lowest accuracy in obese individuals (AUROC 0.71, 95% CI 0.67–0.74) (Fig. 4A). (DeLong test lean vs overweight patients ($p = 0.045$) and vs obese patients ($p = 0.009$)). FIB4 cut-off for sensitivity at 80% and specificity at 95% in lean were 1.38 and 2.34, overweight 1.04 and 2.34 and obese 0.94 and 2.30.

No significant differences in AUROC were observed according to increased vs non increased WC (AUROC 0.68, 95% CI 0.64–0.72 vs 0.81; 95% CI 0.65–0.96; DeLong test $p = 0.108$) (Fig. 4B), presence vs absence of T2DM (AUROC 0.71, 95% CI 0.66–0.74 vs 0.68, 95% CI 0.64–0.71; DeLong test $p = 0.404$) (Fig. 4C), presence vs absence of hypertension (AUROC 0.69, CI 95% CI 0.66–0.73 vs 0.71, 95% CI 0.67–0.76); DeLong test $p = 0.496$) (Fig. 4D), presence vs absence of dyslipidemia (AUROC 0.72, CI 95% CI 0.69–0.75 vs 0.72, 95% CI 0.68–0.77); DeLong test $p = 0.496$) (Fig. 4E) presence vs absence of or insulin-resistance (AUROC 0.70, CI 95% CI 0.65–0.76 vs 0.66, CI 95% CI 0.56–0.77, DeLong test $p = 0.459$) (Fig. 4F).

FIB-4 seemed to perform better in patients with vs without $GFR < 60$ ml/min (AUROC 0.83, 95% CI 0.74–0.91 vs 0.71, 95% CI 0.67–0.73; DeLong test $p = 0.010$). FIB4 cut-off for sensitivity at 80% and specificity at 95% in patients with $GFR \geq 60$ ml/min were 0.93 and 2.26, whereas in those with $GFR < 60$ ml/min 1.56 and 2.54.

• Severity of liver disease

The diagnostic performance of FIB4 was affected by the presence of increased AST and/or ALT, showing better AUROC in those presenting with vs without a rise in transaminases (AUROC 0.75, 95% CI 0.72–0.78 vs 0.68, 95% CI 0.63–0.72; DeLong test $p < 0.001$) (Supplementary Fig. 2A). FIB4 cut-off for sensitivity at 80% and specificity at 95% in patients without increased transaminases were 0.93 and 2.24, whereas in those with increased transaminases 1.04 and 2.65. On the contrary, FIB-4 performance was significantly reduced according to the presence vs absence of severe steatosis (AUROC 0.64, 95% CI 0.58–0.69 vs 0.76; 95% CI 0.73–0.80; DeLong test $p < 0.001$) (Supplementary Fig. 2B). Lower cut-offs were required in patients with vs without severe steatosis for sensitivity at 80% (0.77 vs 1.09), with superimposable thresholds for specificity (2.24 vs 2.34).

• Applying the EASL algorithm by September 2024 and to selected populations

When analyzing the temporal referral, FIB-4 did not show any difference in performance in patients referred from September 2024 onwards vs previous years (AUROC 0.74, 95% CI 0.70–0.78 vs 0.71; 95% CI 0.67–0.74; DeLong test $p = 0.207$), nor between patients meeting vs not meeting EASL-recommended high-risk criteria (T2DM, obesity with metabolic comorbidities, and/or elevated transaminases) (AUROC 0.72, 95% CI 0.69–0.75 vs 0.66, 95% CI 0.57–0.75; DeLong test $p = 0.198$).

Discussion

In this large multicentre Italian cohort of 2638 patients with MASLD referred to tertiary hepatology centres, FIB-4 showed only moderate accuracy (AUROC 0.72, 95% CI 0.69–0.74) for identifying clinically significant fibrosis according to current guideline-recommended referral pathways. Using the recommended rule-out threshold (1.3 or 2.0 in individuals ≥ 65 years), 13% of patients with LSM ≥ 8 kPa, including 7% with LSM ≥ 10 kPa, a proposed threshold for disease-modifying therapies¹⁴, would not have been referred for specialist evaluation. Conversely, nearly 60% of patients above the FIB-4 referral threshold had LSM < 8 kPa, suggesting a substantial burden of unnecessary hepatology referrals and healthcare costs.

Notably, the demographic and metabolic profile of our cohort closely reflected the contemporary epidemiology of MASLD¹. Despite a relatively young mean age (55 years), more than half of patients were obese, nearly 90% had increased WC, over one quarter had T2DM, and a subset had chronic kidney disease, further supporting the concept of MASLD as a key component of the cardio-renal-metabolic syndrome¹⁶. Importantly, the prevalence of hepatic fibrosis (by LSM) was relatively high (21.9%) compared with estimates from population-based studies (around 5–8%)^{1,3}. However, this finding likely reflects the tertiary-care setting of our cohort, which is enriched for patients with more advanced liver disease. Consistently, a recent multinational study including more than 18750 biopsy-proven MASLD patients managed in tertiary hepatology centres reported a fibrosis prevalence of 36%^{10,13}. As expected in a real-world multicentre setting, relevant heterogeneity was observed across centres in terms of age distribution, metabolic comorbidities and severity of liver disease. This likely reflects different regional referral patterns and the heterogeneous epidemiology of metabolic disease across Italy, an issue currently being investigated by the nationwide ITA-MASLD study¹⁷; however, adjusting all analyses for the recruiting centers enhanced the robustness of the results.

A major finding of this study is that FIB-4 misclassification was not random but clustered in patients with an adverse cardiometabolic phenotype. False-negative patients were more likely to have obesity, T2DM, elevated liver enzymes, and severe steatosis, all established drivers of MASLD progression³. Consequently, the patients most likely to be missed by FIB-4 are paradoxically those at greatest risk of clinically significant liver disease and those most likely to benefit from a direct fibrosis assessment. Conversely, among patients with a positive FIB-4, the absence of T2DM and normal aminotransferases identified individuals with a lower likelihood of significant fibrosis. Consistent with this observation, the moderate performance of FIB-4 in identifying clinically significant fibrosis among individuals with T2DM in our cohort mirrors findings from previous studies using the same sequential screening strategy^{11,12}.

In addition, FIB-4 accuracy increased progressively with age, from an AUROC of 0.59 (95% CI 0.51–0.66) in patients < 45 years to 0.77 (95% CI 0.73–0.81) in those ≥ 65 years, whereas no sex-related differences were observed. Accordingly, the FIB-4 thresholds required to maintain fixed sensitivity and specificity targets increased across age categories, consistent with recent multinational studies reporting age-

dependent variability in optimal thresholds^{13,18}. Interestingly, although overall test performance improved with age, age was not independently associated with false-negative classification below the referral threshold. The apparent discrepancy between these findings likely reflects the different aims of the analyses. While increasing age improved the overall discriminatory performance of FIB-4, false-negative classification was evaluated only among patients below the referral threshold. Within this selected subgroup, metabolic risk factors, rather than age itself, emerged as the main determinants of clinically relevant missed fibrosis.

Indeed, also BMI highly influenced FIB4 performance, that progressively declined from lean (AUROC: 0.82 95% CI 0.74–0.91) to overweight (AUROC: 0.73 95% CI 0.68–0.78) and obese individuals (AUROC 0.71, 95% CI 0.67–0.74), with consensual lower rule-out thresholds to maintain high sensitivity. Obesity is associated with a more complex metabolic phenotype in which fibrosis may develop despite only modest alterations in aminotransferases or platelet count. In addition, excess adiposity may affect the accuracy of transient elastography, particularly in patients with increased skin-to-liver distance^{19,20}. Similarly, FIB-4 performed less accurately in patients with severe steatosis, who showed significantly lower AUROC values and a higher likelihood of false-negative classification. Severe steatosis is a marker of metabolic dysfunction and fibrogenesis and may also influence liver stiffness independently of fibrosis stage, contributing to discordance between FIB-4 and LSM²¹.

As expected, FIB-4 performed better in patients with elevated aminotransferases (AUROC 0.75 vs 0.68), likely because AST and ALT are incorporated into the score and identify patients with more active liver disease. By contrast, the numerically higher performance observed in patients with impaired renal function should be interpreted cautiously given the small sample size and probably reflects a greater burden of advanced metabolic disease rather than improved test performance^{22,23}.

Unexpectedly, FIB-4 performance was unaffected by referral timing or fulfilment of EASL high-risk criteria. This probably reflects the characteristics of our tertiary-care cohort, in which most patients already fulfilled guideline criteria and referral patterns had anticipated formal implementation of the guidelines' pathway.

The clinical implications of these findings are several. For sure FIB-4 remains a useful and scalable first-line tool for MASLD risk stratification, particularly because of its high negative predictive value and low cost. However, it should be used in the right setting, where the prevalence of advanced liver disease and metabolic alterations are lower. In fact, as demonstrated by our results, low FIB-4 results should be interpreted cautiously. Moreover, current algorithms may benefit from further refinement by integrating simple anthropometric and metabolic variables, particularly BMI and diabetes status, to identify patients who should undergo second-line assessment despite low FIB-4.

The strengths of this study include the large sample size, the multicentre design, the real-world setting and the systematic availability of liver stiffness and FIB-4 calculation. Moreover, the study provides

clinically relevant information on the performance of FIB-4 across subgroups commonly encountered in specialist practice.

Several limitations should also be acknowledged. First, liver biopsy was not available and LSM ≥ 8 kPa was used as the reference standard. Although transient elastography has known limitations, particularly in obese individuals, examinations were performed using the most appropriate probe (M or XL) and only reliable measurements were included. Second, the tertiary-care setting may have introduced referral bias, limiting the generalizability of fibrosis prevalence estimates to primary-care populations. However, this was also the target population of the study, allowing evaluation of FIB-4 under real-world specialist practice. Although differences in referral pathways likely contributed to inter-centre heterogeneity, all multivariable analyses were adjusted for recruiting centre. Finally, although subgroup-specific FIB-4 thresholds were explored, they were derived within the same cohort and therefore require external validation before any clinical application.

In conclusion, in this large multicentre Italian cohort of patients with MASLD referred to tertiary hepatology centres, FIB-4 showed only moderate performance for identifying patients with elevated liver stiffness. While its high negative predictive value supports its role as a first-line triage tool, a substantial proportion of patients with significant fibrosis and an adverse cardiometabolic profile were missed by the current referral pathway. These findings challenge the use of a universal fibrosis screening strategy and support a more individualized approach. In particular, selected high-risk individuals, as those with T2DM or obesity, may benefit from direct liver stiffness assessment rather than a sequential approach. In addition, future referral algorithms may benefit from incorporating metabolic risk profiling and from evaluating tailored FIB-4 thresholds according to age and metabolic characteristics in prospective validation studies to optimize the trade-off between missed cases and unnecessary referrals.

Abbreviations

AASLD: American Association for the Study of Liver Disease ; AUROC: area under the curve; BMI: body mass index; CI: confidence intervals; EASL: European Association for the Study of the Liver; FIB-4: Fibrosis-4; GFR: glomerular filtration rate; HbA1c: glycated hemoglobin; HOMA-IR: Homeostasis Model Assessment of Insulin Resistance; LSM: liver stiffness measurement; MASLD: Metabolic Dysfunction Associated Steatotic Liver Disease;; MASH: metabolic dysfunction steatohepatitis; NPV: negative predictive value; OR: odds ratios; PPV: positive predictive value; T2DM: type 2 diabetes; US: ultrasound/ultrasonography; VCTE: vibration-controlled transient elastography; WC: waist circumference

Declarations

Competing interests:

A.D received speaking fees from Madrigal, Lilly, Novonordisk ; A.A has received speaking fees from MSD, Gilead, AlfaSigma, Intercept, AbbVie, and Sobi; has served as a consultant for Gilead, Intercept, Sobi,

AbbVie, and MSD and has received research grants from Gilead and AbbVie; N.P served as a speaker or advisor for Gilead, Gore, Astra Zeneca, AlfaSigma, Madrigal and Novo Nordisk and has received travel grants from Gilead, AlfaSigma and AbbVie; G.P had a role in speaking, consulting, teaching and advisory board for Echosense, Madrigal, Novo Nordisk, Advanz; F.P had a role on consulting or advisory board with AstraZeneca, Bracco, BMS, ESAOTE, Eisai, GE, Gilead, Ipsen, MSD, Nerviano, Roche, Samsung, Signant Health, and Siemens Healthineer; F.R had a role on consulting or advisory board with Madrigal, Eli Lilly, Errekappa, Novo Nordisk ; F.T: served as consultant for Echosense, Ipsen, Gilead, Travel Mirum; P.G served as consultant for AlfaSigma, AbbVie, Gilead, AstraZeneca, Madrigal; S.P served as a speaker or advisor for Boeringher Ingelheim, Madrigal, Novo Nordisk; U.V.G had a role as advisor for Ipsen, Gilead, Novo-Nordisk, Madrigal, Advanz Pharma, Mirum, GSK; A.L.F had a role as PI in pharmacological trials from Boeringher Ingelheim, Novo Nordisk, MSD, AstraZeneca, GSK; R.L had a role as consultant for Boeringher Ingelheim and Business Integration Partner S.P.A and has received research grants from Gilead; F.C, A.C, M.M, M.P, P.T, S.F, A.M, E.B, A.M, R.V, G.S declares no conflicts of interests.

Informed consent in studies with human subjects: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Informed consent was obtained from all patients for being included in the study.

Funding:

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript

Authors' contribution:

R.L. conceived and designed the study, coordinated the multicentre collaboration, supervised all phases of the project, performed the statistical analyses, interpreted the data, and drafted the manuscript. F.C. contributed to the study design, coordinated data collection, curated the database, participated in the statistical analyses and interpretation of the results, and drafted the manuscript. M.M. (Matteo Mureddu), A.C., N.P., R.V., M.M. (Mario Masarone), P.T., S.F., P.G., F.T., and A.M. (Anna Mantovani) contributed to patient recruitment, data collection, database curation, data validation, and critical revision of the manuscript. E.B., F.R., A.D., A.M. (Alessandro Mantovani), F.P., G.S., M.P., A.A., U.V.G., S.P., and A.L.F. contributed to the study methodology, interpretation of the findings, scientific supervision, and critical revision of the manuscript for important intellectual content. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Acknowledgments:

References

1. Younossi ZM, Kalligeros M, Henry L. Epidemiology of metabolic dysfunction-associated steatotic liver disease. *Clin Mol Hepatol*. 2025;31(Suppl):S32–50.
2. Vilar-Gomez E, Calzadilla-Bertot L, Wai-Sun Wong V, et al. Fibrosis Severity as a Determinant of Cause-Specific Mortality in Patients With Advanced Nonalcoholic Fatty Liver Disease: A Multi-National Cohort Study. *Gastroenterology*. 2018;155(2):443–57. e417.
3. European Association for the Study of the L. European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J Hepatol*. 2024;81(3):492–542.
4. Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology*. 2007;46(1):32–6.
5. 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;8(8):714–25.
6. Yip TC-F, Lee HW, Lin H, et al. Prognostic performance of the two-step clinical care pathway in metabolic dysfunction-associated steatotic liver disease. *J Hepatol*. 2025;83(2):304–14.
7. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology*. 2023;77(5):1797–835.
8. Kanwal F, Shubbrook JH, Adams LA, et al. Clinical Care Pathway for the Risk Stratification and Management of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2021;161(5):1657–69.
9. Younossi ZM, Zelber-Sagi S, Lazarus JV, et al. Global Consensus Recommendations for Metabolic Dysfunction-Associated Steatotic Liver Disease and Steatohepatitis. *Gastroenterology*. 2025;169(5):1017–32. e1012.
10. Semmler G, Thone P, Embacher J et al. Prevalence of Fibrosis and Applicability of Lab-Based Noninvasive Tests From Primary to Tertiary Care. *Clin Gastroenterol Hepatol*. 2025.
11. Caviglia GP, Ferro A, D'Ambrosio R, et al. Effectiveness of a Model of Care Based on Fibrosis-4 and Liver Stiffness Measurement for the Screening of Patients With Type 2 Diabetes Mellitus at Risk of Advanced Liver Disease: Results From an Italian Prospective Multicenter Study. *Am J Gastroenterol*. 2026;121(2):375–82.
12. Mantovani A, Lombardi R, Dalbeni A, et al. Two-tier screening approach for liver fibrosis stratification in outpatients with type 2 diabetes mellitus: A multicenter cross-sectional study. *Diabetes Obes Metab*. 2026;28(1):644–53.

13. Younossi ZM, de Avila L, Petta S et al. Diagnostic Accuracy of Noninvasive Tests for Metabolic Dysfunction-associated Steatotic Liver Disease Across Age, Type 2 Diabetes, and Obesity Subgroups: A Multinational Study. *Clin Gastroenterol Hepatol*. 2026.
14. Noureddin M, Charlton MR, Harrison SA, et al. Expert Panel Recommendations: Practical Clinical Applications for Initiating and Monitoring Resmetirom in Patients With MASH/NASH and Moderate to Noncirrhotic Advanced Fibrosis. *Clin Gastroenterol Hepatol*. 2024;22(12):2367–77.
15. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno VIIF. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol*. 2022;76(4):959–74.
16. Marassi M, Fadini GP. The cardio-renal-metabolic connection: a review of the evidence. *Cardiovasc Diabetol*. 2023;22(1):195.
17. Bugianesi E, Buscemi S, Burra P, et al. ITA-MASLD: A national observational study to characterize the profile of patients with MASLD in specialistic care in Italy. *Dig Liver Dis*. 2025;57(11):2083–6.
18. 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(5):740–51.
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(6):916–28.
20. Weiss J, Rau M, Meertens J, et al. Feasibility of liver stiffness measurement in morbidly obese patients undergoing bariatric surgery using XL probe. *Scand J Gastroenterol*. 2016;51(10):1263–8.
21. Petta S, Maida M, Macaluso FS, et al. The severity of steatosis influences liver stiffness measurement in patients with nonalcoholic fatty liver disease. *Hepatology*. 2015;62(4):1101–10.
22. Mantovani A, Lombardi R, Cattazzo F, Zusi C, Cappelli D, Dalbeni A. MAFLD and CKD: An Updated Narrative Review. *Int J Mol Sci* 2022;23(13).
23. Mantovani A, Petracca G, Beatrice G, et al. Non-alcoholic fatty liver disease and risk of incident chronic kidney disease: an updated meta-analysis. *Gut*. 2022;71(1):156–62.

Figures

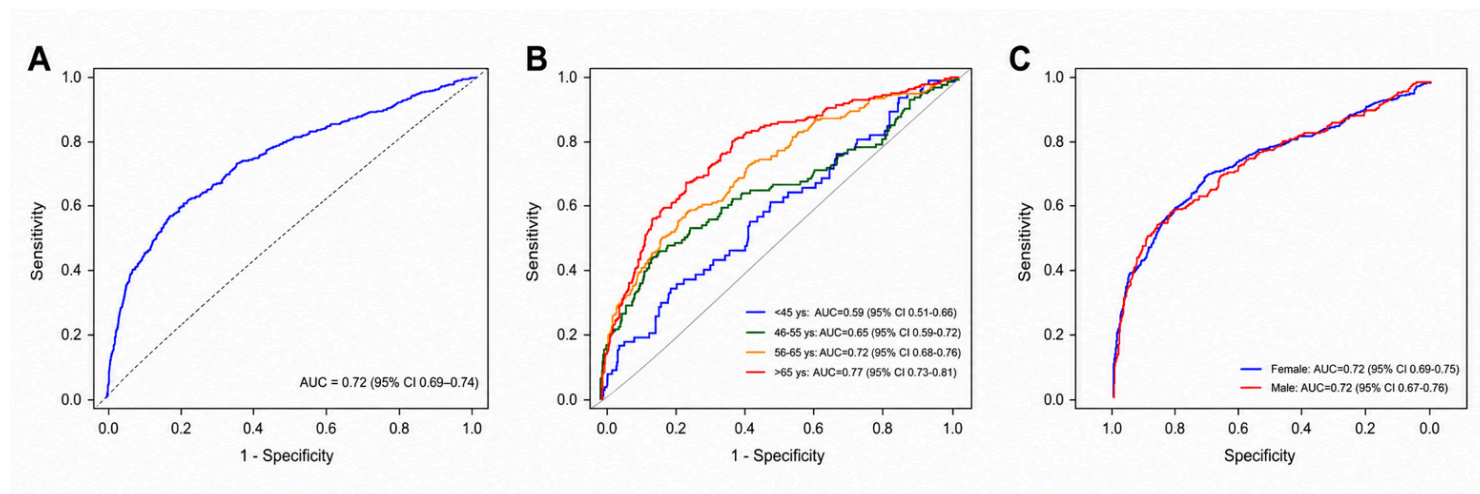


Figure 1

Diagnostic performance of FIB-4 for predicting LSM ≥ 8 kPa in the overall cohort (A), by age category (<45, 46–55, 56–65, and >65 years) (B), and by sex (C).

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; FIB-4, Fibrosis-4 index; LSM, liver stiffness measurement

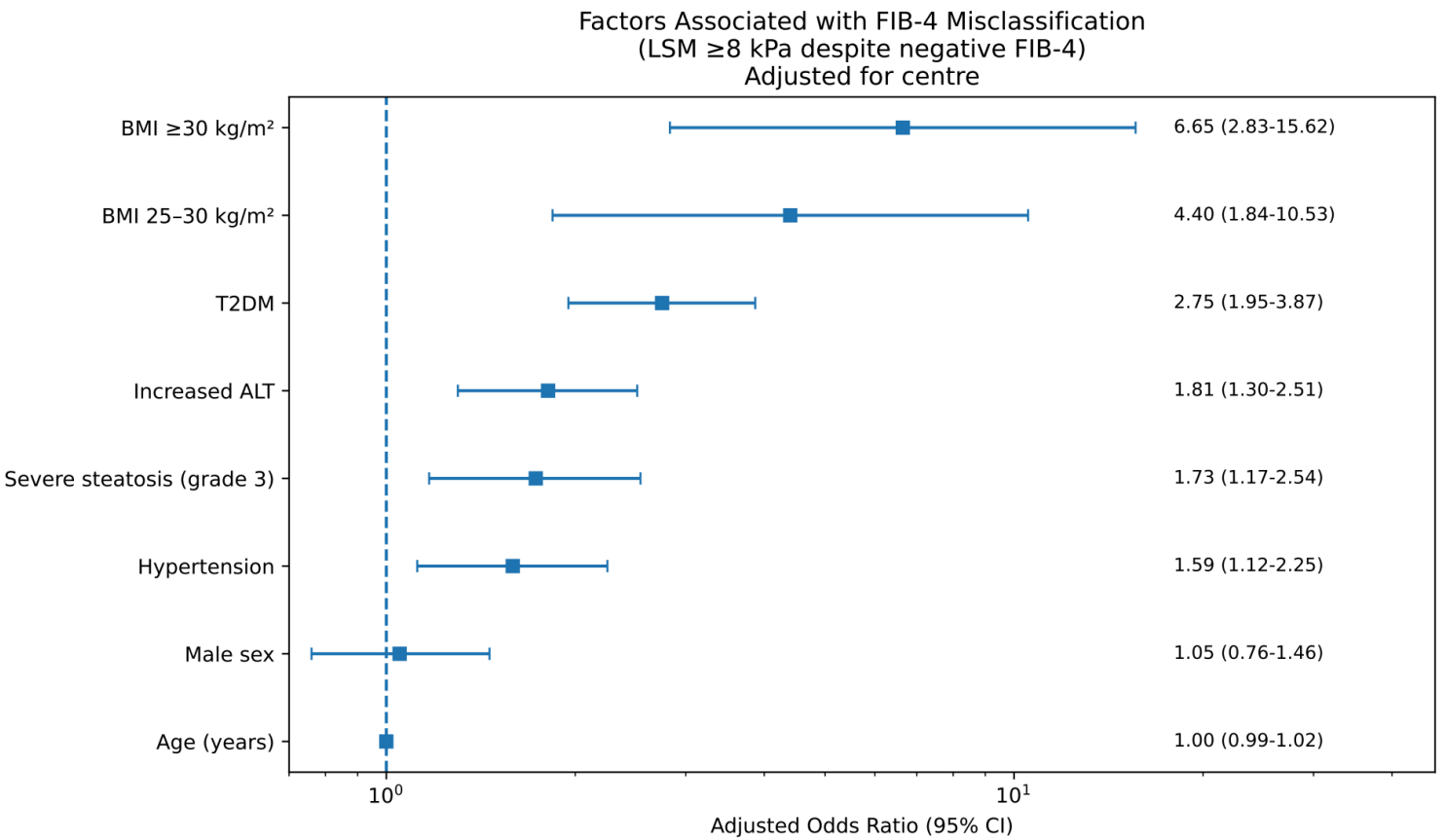


Figure 2

Multivariate analysis of factors affecting misclassification (LSM ≥ 8 kPa) in patients with negative FIB-4 (<1.3/2).

Abbreviations: ALT, alanine aminotransferase; BMI, body mass index; T2DM: type 2 diabetes mellitus CI, confidence interval; LSM, liver stiffness measurement; OR, odds ratio.

*adjusted for centre

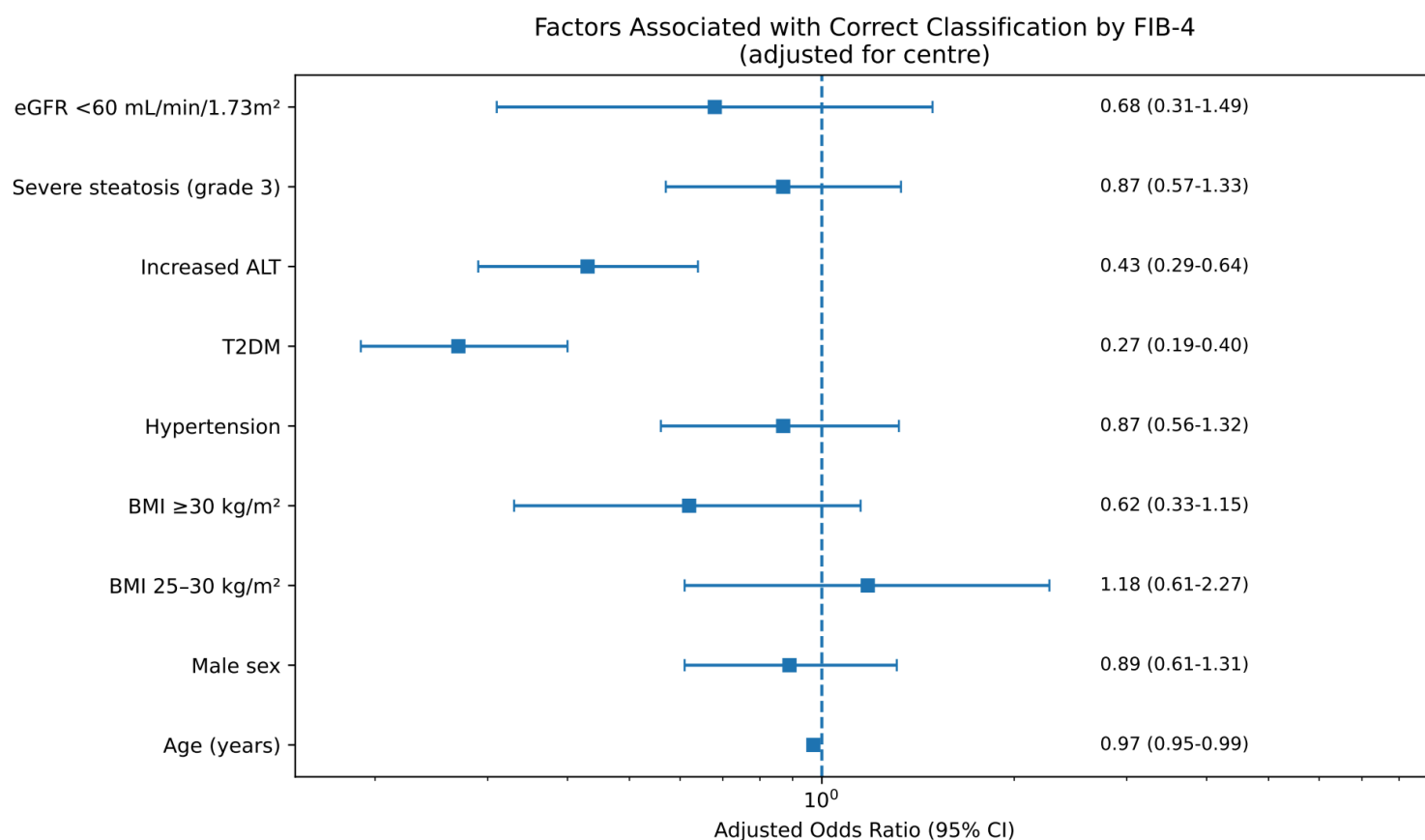


Figure 3

Multivariate analysis of factors affecting misclassification (LSM<8kPa) in patients with positive FIB-4 ($\geq 1.3/2$).

Abbreviations: ALT, alanine aminotransferase; BMI, body mass index; T2DM: type 2 diabetes mellitus CI, confidence interval; LSM, liver stiffness measurement; OR, odds ratio.

*adjusted for centre

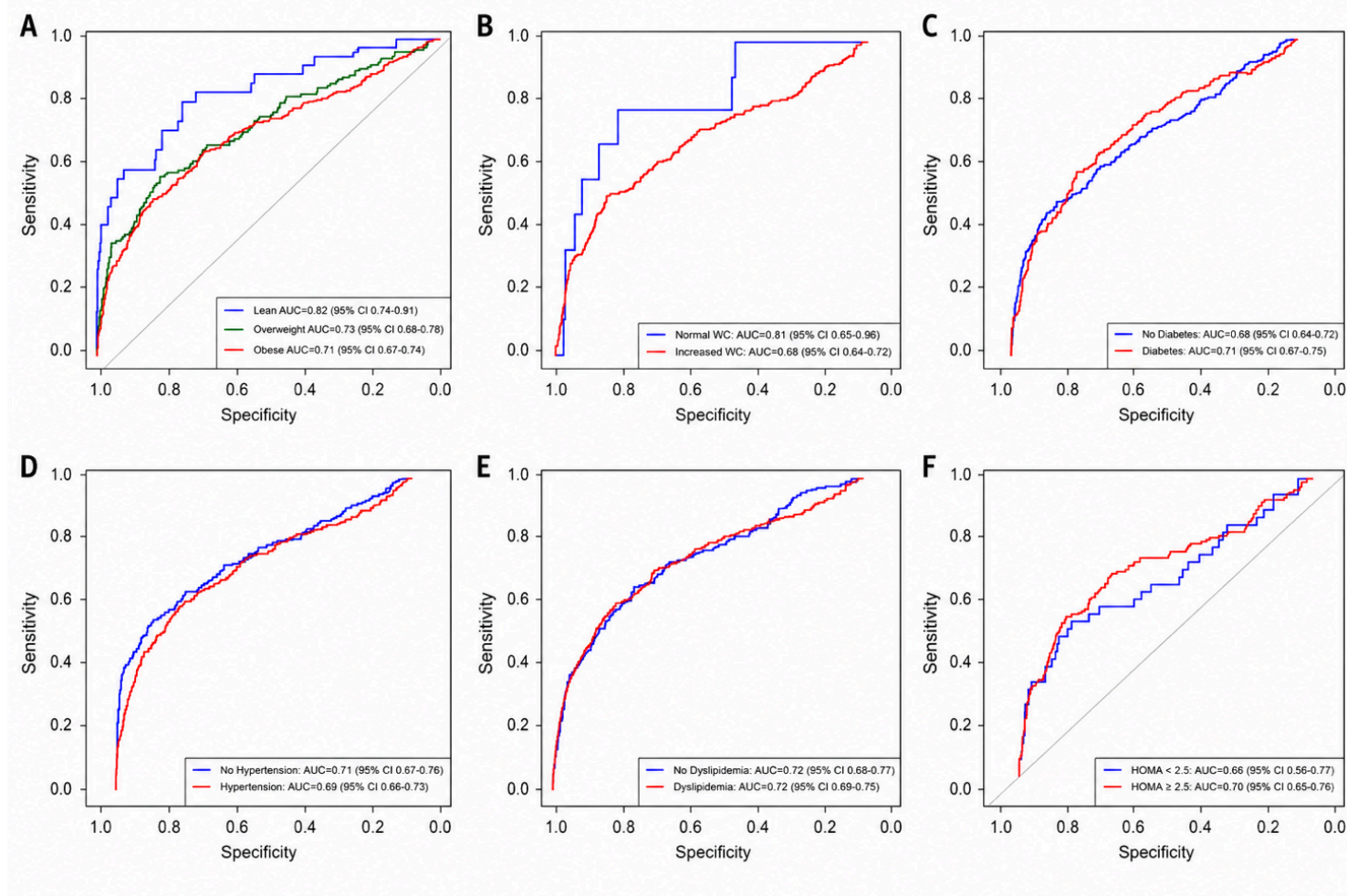


Figure 4

Diagnostic performance of FIB-4 for predicting LSM ≥ 8 kPa across metabolic comorbidity subgroups. ROC curves are shown according to BMI category (A), waist circumference (B), type 2 diabetes (C), hypertension (D), dyslipidemia (E), and insulin resistance (F).

Abbreviations: AUC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; FIB-4, Fibrosis-4 index; HOMA-IR, homeostatic model assessment of insulin resistance; LSM, liver stiffness measurement; ROC, receiver operating characteristic; T2DM, type 2 diabetes mellitus; WC, waist circumference.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicalabstract.pdf](#)
- [Supplementaryfiles.docx](#)