

Is FIB-4 the right tool for screening for liver fibrosis?

Leen JM. Heyens , Demi P.A.van Malde , Gediz Dogay Us ,
Francesco Innocenti , Mathieu Struyve ,
Christophe Van Steenkiste , Sven Francque , Geert Robaey ,
Ger H. Koek , the MASLD research group

PII: S1665-2681(25)00401-6
DOI: <https://doi.org/10.1016/j.aohep.2025.102176>
Reference: AOHEP 102176



To appear in: *Annals of Hepatology*

Received date: 10 November 2025
Accepted date: 10 December 2025

Please cite this article as: Leen JM. Heyens , Demi P.A.van Malde , Gediz Dogay Us , Francesco Innocenti , Mathieu Struyve , Christophe Van Steenkiste , Sven Francque , Geert Robaey , Ger H. Koek , the MASLD research group, Is FIB-4 the right tool for screening for liver fibrosis?, *Annals of Hepatology* (2025), doi: <https://doi.org/10.1016/j.aohep.2025.102176>

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier España, S.L.U. on behalf of Fundación Clínica Médica Sur, A.C.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Is FIB-4 the right tool for screening for liver fibrosis?

Leen JM. Heyens^{a,b,c,*}, Demi P.A. van Malde^{d#}, Gediz Dogay Us^{b,e}, Francesco Innocenti^f, Mathieu Struyve^c, Christophe Van Steenkiste^{g,h,i}, Sven Francque^{g,i}, Geert Robaey^a, Ger H. Koek^{b,d} and the MASLD research group[†]

^aHasselt University, Faculty of Health and Life Sciences/LCRC, Diepenbeek, Belgium

^bMaastricht University, School of Nutrition and Translational Research in Metabolism, NUTRIM, Maastricht, the Netherlands

^cZiekenhuis Oost-Limburg, Department of Gastro-enterology and Hepatology and Department Future Health, Genk, Belgium

^dMUMC+, Department of Gastroenterology and Hepatology, Maastricht University Medical Centre, Netherlands,

^ePax Clinic, Department of Gastroenterology, Istanbul, Türkiye

^fDepartment of Methodology and Statistics, Care and Public Health Research Institute (CAPHRI), Maastricht University

^gAntwerp University Hospital, Department of Gastroenterology and Hepatology, Antwerp, Belgium

^hAZ Maria Middelaers Ghent, Department of Gastroenterology and Hepatology, Ghent, Belgium

ⁱLaboratory of Experimental Medicine and Paediatrics, Faculty of Medicine and Health Sciences, InflaMed, University of Antwerp, Antwerp, Belgium

[#]Authors contributed equally.

[†]MASLD research group: Yves Kockaerts¹, Ine Lowyck¹, Kirsten Stinkens¹, Heddy. Van Leeuwen-Wintjens¹, Liesbeth Vernijns², Jacques Germaux³, Toon Meyers⁴, Jan Gyselaers⁵, Sjors Pietermans⁶, Ruth Bollen⁷, Anneleen Robaey⁸, Canio Gilio⁹, Roy Remmen¹⁰, Jean Muris¹¹, Marjolijn Uitterhoeve-Prins¹², Michelle Bremers¹³, Loes van Bokhoven¹⁴, Judith Wellens¹⁵, Marlies Devos¹⁶, Kitty de Munck¹⁷, Eva Rubens¹⁷, Katrien Joris¹⁷, M.H. Us¹⁸, M. Ugur¹⁸, I.N. Akpinar¹⁸

¹Ziekenhuis Oost-Limburg, Department of Endocrinology, Genk, Belgium

²Huisartsenbox, Genk, Belgium

³Huisartsenpraktijk Medi-Mine, Genk, Belgium

⁴Groepspraktijk Luce, Genk, Belgium

⁵Gezondheidscentrum Sirona vzw, Genk, Belgium

⁶Groepspraktijk De Dam, Maasmechelen, Belgium

⁷W-Care Hoeselt, Hoeselt, Belgium

⁸Huisartsen Termolen, Zonhoven, Belgium

⁹Huisarts Gilio, Maasmechelen, Belgium

¹⁰University of Antwerp, Maatschappelijke Gezondheidszorg en Eerstelijnszorg, Wilrijk, Belgium

¹¹CAPHRI Care and Public Health Research Institute, Department of General Practice, UM, The Netherlands

¹²Dokters van Hier, Maastricht, The Netherlands

¹³Huisartsenpraktijk Bandkeramiek, Elsloo, The Netherlands

¹⁴CAPHRI Care and Public Health Research Institute, UM, The Netherlands

¹⁵Department of Gastroenterology and Hepatology, Leuven University Hospitals.

¹⁶Groepspraktijk Het Spoor, Baasrode Belgium

¹⁷Antwerp University Hospital, Department of Gastroenterology and Hepatology, Antwerp, Belgium

¹⁸Pax Clinic, Department of Gastroenterology, Istanbul, Türkiye

***Corresponding author:**

Leen Heyens

leen.heyens@uhasselt.be or l.heyens@maastrichtuniversity.nl

Word count:

Figures:

Tables: 5

ABSTRACT

Introduction and Objectives: The screening accuracy of non-invasive fibrosis tests like FIB-4 in metabolic dysfunction-associated steatotic liver disease (MASLD) remains unclear. Using standard cut-offs, this study evaluated FIB-4's agreement with vibration-controlled transient elastography (VCTE) and identified and validated new thresholds.

Patients and Methods: A prospective cohort study (2019–2024) in Belgian and Dutch primary care used VCTE by FibroScan® (Echosens, France) as a proxy for the fibrosis stage. The FIB-4 index was derived from electronic patient data and study blood samples. Agreement between VCTE and FIB-4 was analysed using Weighted Cohen's kappa. New fibrosis cut-offs ($\geq F2$; ≥ 8 kPa) for ≤ 65 and > 65 years were determined via Youden's Index and validated in a Turkish primary care cohort and a Belgian secondary care T2DM cohort.

Results: Among 563 participants (median age 62 years, 47.1% male, 14.2% with T2DM, median BMI 28.2 kg/m²), FIB-4 showed poor agreement with VCTE ($\kappa = 0.138$, 95% CI: 0.069–0.207). Suggested new cut-offs of 1.29 (≤ 65 years) and 1.72 (> 65 years) were proposed. The 1.29 cut-off performed similarly to the existing 1.3 in validation cohorts. In the Türkiye and T2DM cohorts, the 1.72 cut-off improved sensitivity over 2.0 but had lower specificity.

Conclusions: The FIB-4 index showed poor agreement with VCTE and low sensitivity, making it an unreliable standalone diagnostic tool for liver fibrosis in people with MASLD in both primary and secondary care. Alternative non-invasive tests or improved cut-off values are needed for accurate fibrosis detection in clinical practice.

Keywords: FIB-4 score, fibrosis, screening, metabolic dysfunction-associated steatotic liver disease, vibration-controlled transient elastography

Abbreviations: ALT, Alanine transaminase; AST, Aspartate transaminase; AUROC, Area under the receiver operating characteristic curve; BMI, Body mass index; CAP, Controlled attenuation parameter; CI, Confidence interval; CVD, Cardiovascular disease; EASD, European Association for the Study of Diabetes; EASL, European Association for the Study of the Liver; EASO, European Association for the Study of Obesity; ELF, Enhanced liver fibrosis (test); FIB-4, Fibrosis-4 index; GGT, Gamma-glutamyl transferase (also called gamma-glutamyl transpeptidase); GP, General practitioner; HCC, Hepatocellular carcinoma; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; LSM, Liver stiffness measurement; MAF-5, Metabolic dysfunction-associated fibrosis 5 (score); MASLD, Metabolic dysfunction-associated steatotic liver disease; MASH, Metabolic dysfunction-associated steatohepatitis; MetS, Metabolic syndrome; NAFLD, Non-alcoholic fatty liver disease; NASH, Non-alcoholic steatohepatitis; NPV, Negative predictive value; PCP, Primary care practice; PPV, Positive predictive value; T2DM, Type 2 diabetes mellitus; TR, Türkiye;

VCTE, Vibration-controlled transient elastography.

Journal Pre-proof

1. Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD, formerly known as non-alcoholic fatty liver disease; NAFLD¹), is defined by macrovesicular lipid accumulation in more than five percent of hepatocytes in combination with at least one cardiometabolic risk factor like dyslipidaemia, hypertension, and diabetes. MASLD, in the absence of classical causes of steatogenesis (e.g., alcohol and steatogenic drugs), largely overlaps with NAFLD, but in contrast to the latter, it can co-exist with other chronic liver diseases². MASLD encompasses a wide spectrum of clinical conditions, ranging from isolated steatosis to inflammation (metabolic dysfunction-associated steatohepatitis; MASH), liver fibrosis, cirrhosis, and hepatocellular carcinoma (HCC)³. The global prevalence of MASLD, currently estimated at 32.4%, parallels the global increase in obesity and type 2 diabetes mellitus (T2DM)^{4,5}. In addition, MASLD is also causally linked to the development of metabolic disturbances, such as T2DM (hence implying a complex multidirectional pathophysiological relation), and is associated with an increased risk of cardiovascular disease (CVD), non-liver malignancies, and various extra-hepatic complications⁶⁻⁹.

The stage of liver fibrosis is the strongest predictor for liver-related mortality and the development of other hepatic and extra-hepatic comorbidities¹⁰⁻¹⁴. The severity of fibrosis is scaled from F0 to F4, as proposed by the MASH Clinical Research Network Scoring System. An increase in fibrosis from stage $F \geq 2$ or higher (called “significant fibrosis”) is correlated with a more severe liver- and non-liver-related outcome¹⁵. Therefore, it is vital to correctly rule out significant liver fibrosis ($F \geq 2$), preferably with non-invasive methods like vibration-controlled transient elastography (VCTE) or the FIB-4 Index. VCTE determines liver stiffness based on the propagation of a shear wave through the liver. It is a reliable technique used as a surrogate marker for liver fibrosis and is incorporated in the FibroScan® device¹⁶. The FIB-4 is a freely available score that consists of the parameters aspartate transaminase (AST), alanine transaminase (ALT), age, and thrombocytes¹⁷. Because these parameters are usually readily available, the FIB-4 is easy to use for screening. Consequently, the European Association for the Study of the Liver (EASL) recommends using the FIB-4 and VCTE as the first and second steps to rule out liver fibrosis¹⁶. The reliability of the FIB-4 varies, however, across different populations, including those who are living with overweight, obesity, or T2DM and having CVD. Furthermore, it appears less accurate in populations with T2DM and CVD¹⁸⁻²¹. The screening performance of the FIB-4 in people with MASLD remains to be adequately investigated further.

Therefore, we aimed to investigate the agreement between the FIB-4 and VCTE with the existing age-dependent cut-off values in a primary care cohort ¹⁶. We then searched for new age-dependent cut-off values that might perform better than the current ones. Finally, these new cut-off values were validated in two external cohorts.

Journal Pre-proof

2. Patients and Methods

2.1 Study population

The following prospective cohort studies, hereinafter referred to as the discovery cohorts, were conducted from October 2020 to February 2024 in nine primary care practices (PCPs) throughout Belgium (NCT04647409) and five PCPs throughout the Netherlands (NCT04918732). The study protocols were conducted according to the Helsinki Declaration after approval by the Ethics Committee of Hasselt University (CME2020 019), University Hospital Antwerp (19/44/495), Maria Middelaers (MMS.2019.019), and the Medical Ethical Committee of Maastricht University/Maastricht University Medical Centre (NL73265.068.20). The prospective cohort studies used as validation cohorts were conducted from September 2019 to October 2024 in Türkiye (NCT05194553) and at the endocrinology department of Ziekenhuis Oost-Limburg (NCT04999124).

2.2 Participants

Participants were recruited by the general practitioner (GP), or e-mails were sent through the electronic patient record system used in the GP practice. Additionally, flyers and posters were available in the waiting room. People with T2DM were recruited by their endocrinologists. Participants were eligible for inclusion if they were older than 18 years old, able to understand Dutch, had no excessive alcohol use (more than 2 or 3 glasses of alcohol per day for women or men, respectively), and had no other known liver diseases or secondary causes of steatosis. Other liver diseases were defined as having Hepatitis B virus, Hepatitis C virus, autoimmune hepatitis, primary biliary cholangitis, haemochromatosis, Wilson's disease, or Alpha 1 antitrypsin deficiency. Secondary causes of steatosis are mainly drug-induced liver injury, *e.g.*, by amiodarone, tamoxifen, and methotrexate (Supplemental Table S1). Inclusion and exclusion criteria for the validation cohorts can be found in the Supplemental materials (Supplemental Table S1).

2.3 Non-invasive tests

As a surrogate for liver fibrosis, liver stiffness was measured by Vibration Controlled Transient Elastography (VCTETM) and liver steatosis by Controlled Attenuation Parameter (CAPTM). VCTE and CAP were measured utilising a FibroScan® 430+ mini (Echosens, Paris, France) in the right liver lobe by intercostal approach. Subjects were asked to be in fasting conditions for at least 3 hours before the examination and were placed in a supine position with arms in maximal abduction. For screening, the M probe (3.5MHz) or the XL probe (2.5 MHz) was selected as indicated by the device. VCTE and CAP values were considered reliable when the

interquartile range (IQR) was equal to or less than 30% of the median liver stiffness measurement (LSM) value (IQR/med) and at least ten measurements were performed. A CAP value of >248 dB/m was considered as having steatosis ²². Fibrosis stages based on VCTE values are categorised as follows: a VCTE value of <8.0 kPa, $\geq 8-12$ kPa, or ≥ 12 kPa was considered as having no to little fibrosis (F0-F1), significant fibrosis (F2) or advanced fibrosis and cirrhosis (F3-F4), respectively ²².

The FIB-4 was calculated for all subjects using their age (in years), platelet count (in $10^9/L$), AST (in U/L), and ALT (in U/L) using the following calculation: $(\text{age} \times \text{AST}) / (\text{platelet count} \times (\sqrt{\text{ALT}}))$ ¹⁷. A value below 1.3 (if aged ≤ 65 years) or 2.0 (if aged >65 years) was considered indicative of a low risk for significant fibrosis (F2), and a score higher than 2.67 indicated a high risk ¹⁶.

2.4 Data collection

Anthropometric data (waist circumference, weight, height, and blood pressure), smoking status, alcohol usage, medical background, medication usage, and demographic data were collected on the day of evaluation. Blood samples were taken within one month of the VCTE measurement. Missing anthropometric data, laboratory values (within 6 months of the VCTE measurement), medical history, and medication usage were collected retrospectively from medical records. A body mass index (BMI) of <25 kg/m², 25-30 kg/m², and >30 kg/m² were considered as normal weight, overweight, and obesity, respectively ^{23,24}. Furthermore, metabolic syndrome (MetS) diagnosis was based on the International Diabetes Federation consensus ²⁵. High waist circumference was defined as >94 cm for men and >80 cm for women ²⁵.

2.5 Sample size calculation

The sample size required for this study was calculated to be 475, using the sample size formula for agreement estimation with binary outcomes as outlined by Shoukri *et al.* ²⁶. The calculation assumed a kappa value of 0.65, based on findings from a prior study on liver fibrosis ²⁷. A proportion of 0.7 was used, reflecting the prevalence of MASLD in high-risk groups ²⁸. The precision, defined as half the width of the confidence interval (CI), was set at 0.075, with a confidence level of 95%.

2.6 Missing data

Considering the variables required to calculate the FIB-4 score, data were missing for 16.8% of the 563 participants. There were no missing data on VCTE values. Missing data on covariates and outcome variables were imputed using the multiple imputation procedure in SPSS version

29.0.2.0 (for details, see Supplement Missing Data). The imputation model included the following variables to enhance the plausibility of the missing at random (MAR) assumption (see Supplement Missing Data for the rationale behind each inclusion): AST, ALT, thrombocytes, FIB-4, VCTE, cohort indicator, waist circumference, BMI, gender, triglycerides, CAP, and gamma-glutamyl transferase (GGT). The number of imputations ($n = 17$) was based on the fraction of missing information. The results obtained with multiple imputation, valid under the MAR assumption, were compared with those from complete case analyses, valid under the more restrictive and less realistic assumption of data missing completely at random (MCAR), as recommended by Sterne ²⁹.

2.7 Statistical analysis

Statistical analysis was performed using Statistical Package for Social Sciences (SPSS) (Version 29.0, Armonk, New York, IBM Corp). Categorical variables were presented as absolute values with percentages. The Shapiro-Wilk or Kolmogorov-Smirnov test was used to test for the normality of continuous variables. All skewed numeric variables are presented as median with IQR. The nonparametric Mann-Whitney U and Kruskal-Wallis tests were used to compare two or more groups for continuous variables with a non-normal distribution. As a post-hoc analysis, the Dunn's test (adjusted with Bonferroni correction for multiple testing) was performed for the Kruskal-Wallis test. The Chi-Square or Fisher's Exact test was used to compare discrete and categorical variables. The Fisher's Exact test was only used if the cells' expected count was less than five. To investigate the agreement between the FIB-4 and VCTE, the weighted Cohen's kappa was used ³⁰. Cohen's kappa values are interpreted based on Landis *et al.*'s cut-off ranges ³⁰. The Area Under the Receiver Operating Characteristic curve (AUROC) was determined for the FIB-4 with VCTE as the gold standard. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the cut-off values were determined. The Youden's Index was used to find the optimal sensitivity-specificity trade-off for new cut-off values. The number of subjects included in every analysis is designated by 'n'. A p-value ≤ 0.05 is considered statistically significant.

2.8 Ethical considerations

The study protocols were approved by the Ethics Committee of the Istanbul Health Sciences University Clinical Studies Ethics Board (2021/224) and the Committee on Medical Ethics of Ziekenhuis Oost-Limburg (CTU2020015), respectively. Good clinical practice (GCP) guidelines were followed throughout the study. All authors had access to the study data and reviewed and approved the final manuscript.

3. Results

3.1 Cohort characteristics

In total, 1,730 people were screened, of whom 563 (71.4%) for the discovery cohort were eligible according to the exclusion criteria (Figure 1). For the validation cohorts, 493 (90.6%) from Türkiye and 229 (57.7%) T2DM participants were eligible, respectively. Table 1 shows the participant characteristics of all analysed cohorts. In the discovery cohort, the median age was 62 [54-70] years, 265 (47.1%) were men, 80 (14.2%) had T2DM, and the median BMI was 28.2 [25.0-31.6] kg/m². For the Türkiye validation cohort, the median age was 52 [44-61] years, 307 (62.3%) were men, 148 (30%) had T2DM, and the median BMI was 28.0 [25.2-31.3] kg/m². In the T2DM cohort, the median age was 61 [56-68] years, 136 (59.4%) were men, and the median BMI was 31.6 [28.5-35.6] kg/m².

3.2 Assessment of agreement

The Weighted Cohen's kappa was applied to assess agreement between F0-F1, F2, and F3-F4 classifications for VCTE, and <1.3, between 1.3 (or 2.0 depending on age) and 2.67, and >2.67 for the FIB-4. Analysis of the total cohort showed a weighted kappa of 0.117 (95% CI, 0.027-0.208), indicating a slight agreement between VCTE and the FIB-4 (Table 2). The sub-analyses done to determine the agreement for different demographic categories, *i.e.*, male or female, age ≤ 65 years or >65 years, or risk groups like living with overweight, obesity, T2DM, and metabolic syndrome, showed a slight to fair agreement between VCTE and FIB-4. The imputed case analysis revealed a weighted kappa value for the total cohort of 0.138 (95% CI, 0.069-0.207), which was slightly higher than that of the non-imputed dataset but still indicated a slight agreement. A similar trend was observed in the sub-analyses except for the people with normal weight, overweightness, or obesity (Table 2). While the complete case analysis showed that the weighted Cohen's kappa of people with normal weight was better than that of obese people, the imputed data showed the opposite. Since the imputation model included key patient characteristics, accounting for notable differences between complete and incomplete cases, the outcome measures, and variables associated with steatosis and fibrosis (for details, see Supplement Missing Data), agreement estimates from multiple imputation are likely less biased than those from complete case analysis, which relies on the restrictive MCAR assumption. Nevertheless, both approaches suggest only a slight agreement between VCTE and FIB-4.

3.3 Cut-off determination

The AUROC for the diagnosis of significant fibrosis of the FIB-4 was found to be 0.654 (95% CI, 0.555-0.753) for people aged ≤ 65 years and 0.625 (95% CI, 0.524 – 0.726) for people older

than 65 years (Figure S1). Similar AUROC values were found in the imputed case analysis for the age-divide groups, respectively, 0.651 (95% CI, 0.557-0.745) and 0.632 (95% CI, 0.534-0.730) (data not shown). The sensitivity, specificity, PPV, and NPV of the current FIB-4 cut-off value of 1.3 were 40.6 % (95% CI, 25.5-57.7), 80.1% (95% CI, 74.9-84.4), 19.7% (95% CI, 11.9-30.8), and 91.8% (95% CI, 87.6-94.7), respectively. For the current cut-off value of 2.0, the sensitivity, specificity, PPV, and NPV were 27.0% (95% CI, 15.4-43.0), 81.0% (95% CI, 73.8-86.5), 26.3% (95% CI, 15.0-42.0), and 81.5 % (95% CI, 74.4-87.0) (Table 3). Similar results were again observed for the imputed case analysis (Table 3).

In an attempt to overcome the slight agreement between VCTE and FIB-4, as well as the poor AUROC values, new cut-offs were established. The Youden's Index was determined with the highest value of 0.234, both for the lower (aged ≤ 65 years) and the higher age (aged >65 years) group. This resulted in a FIB-4 cut-off value of 1.29 for people aged 65 years or younger and 1.72 for those aged 65 years or older. The sensitivity, specificity, PPV, and NPV for both cut-off values were 40.6% (95% CI, 25.5-57.7) and 56.8% (95% CI, 40.9-71.3), 79.7% (95% CI, 74.5-84.1) and 66.7% (95% CI, 58.7-73.8), 19.4% (95% CI, 11.7-30.4) and 30.0% (95% CI, 20.5-41.5), 91.8% (95% CI, 87.5-94.7) and 86.9% (95% CI, 78.4-91.2), respectively (Table 3). Comparable results were also observed in the imputed case analysis (Table 3).

3.4 Validation

The newly proposed cut-off values were validated in two external cohorts: a primary care (Türkiye cohort) and a secondary care (T2DM cohort) population. The AUROC for the diagnosis of advanced fibrosis of the FIB-4 was 0.635 (95% CI, 0.501-0.768) and 0.760 (95% CI, 0.564-0.957) for people aged ≤ 65 years and > 65 years in the Türkiye cohort. In the T2DM cohort, the AUROC values were 0.643 (95% CI, 0.551-0.735) and 0.605 (95% CI, 0.482-0.727) for both age groups (data not shown). These AUROC values were similar to those in the discovery cohort.

The weighted Cohen's kappa of the current cut-off values (1.3/2.0 and 2.67) in these cohorts was 0.126 (95% CI, 0.004-0.248) and 0.266 (95% CI, 0.009-0.524), 0.155 (95% CI, -0.007-0.317) and 0.187 (95% CI, -0.012-0.386) for the Türkiye and T2DM cohorts respectively (Table 4). For the suggested cut-off values of 1.29 and 1.72 in the Türkiye and T2DM cohorts, the weighted Cohen's kappa was 0.115 (95% CI, -0.001-0.231) and 0.187 (95% CI, 0.029-0.346), 0.182 (95% CI, -0.015-0.379) and 0.101 (95% CI, -0.013-0.314) (Table 4). When compared to

the current cut-off value, only for the 1.72 cut-off value in the Türkiye cohort was the weighted Cohen's kappa better than the current cut-off value of 2.0.

The proposed cut-off value of 1.29 had the following sensitivity, specificity, PPV, and NPV in the Türkiye and T2DM cohorts: 33.3 (95% CI, 17.2-54.6)% and 44.4 (95% CI, 32.8-56.7)%, 87.5% (95% CI, 83.9-90.4) and 73.8% (95% CI, 63.5-82.0), 12.3% (95% CI, 6.1-23.2) and 56.0% (95% CI, 42.3-68.8), 96.1% (95% CI, 93.6-97.7) and 63.9% (95% CI, 54.0-72.8) (Table 5). The performance of the 1.3 and 1.29 cut-offs are very similar in both validation cohorts. For the cut-off value of 1.72, the sensitivity, specificity, PPV, and NPV were, respectively, 66.7% (95% CI, 35.4-87.9) and 52.3% (95% CI, 37.9-66.2), 67.2% (95% CI, 55.0-77.4) and 57.9% (95% CI, 42.2-72.1), 22.2% (95% CI, 10.6-40.8) and 59.0% (95% CI, 43.4-72.9), 93.5% (95% CI, 82.5-97.8) and 51.2% (95% CI, 36.8-65.4) in the Türkiye and T2DM cohorts (Table 5). In both the Türkiye and T2DM cohorts, the suggested cut-off value of 1.72 outperforms the 2.0 cut-off value when it comes to sensitivity, but specificity is lower, creating both fewer false negatives and more false positives.

4. Discussion

We are the first to assess the agreement between VCTE and the FIB-4 in both a primary and secondary care MASLD population using the existing age-dependent cut-off values ¹⁶. The study results indicate a slight agreement between VCTE and FIB-4, independent of demographics or risk group. This results in a high false positive or negative rate when people are referred to secondary or tertiary care based on the FIB-4 cut-offs. In addition, it leads to over-referrals and extra costs for the healthcare system. Kjaergaard *et al.* found that using the ≥ 1.3 cut-off value for the FIB-4 in a general population cohort leads to a false positive rate of 35% for the detection of significant fibrosis ³¹. Similarly, Graupera *et al.* found a high false positive rate of 29% using data gathered from five independent cohorts across different countries ²⁰. In our study cohort, an over-referral rate of at least 81.6% was found for people aged ≤ 65 years with a cut-off ≥ 1.3 , and 73.7% for people aged > 65 years with a cut-off ≥ 2.0 . The higher result could be attributed to differences in population characteristics, such as a higher prevalence of metabolic risk factors, *e.g.*, obesity, T2DM, or metabolic syndrome. The high false positive rate could also be attributed to the fact that, like many other non-invasive scores and tools, the FIB-4 has been developed in a selected population of patients with HIV/HCV co-infection ¹⁷. It has subsequently been promoted to screen larger populations as it is easy to use, although it has been poorly validated for that purpose or validated only in highly selected MASLD cohorts.

Next, the AUROC of FIB-4 was assessed, and it was found that FIB-4 had poor discriminative ability in our study cohort. Previous studies found better results concerning the AUROC value when using VCTE or liver biopsy as a reference. For example, Ding *et al.* found an AUROC of 0.728 (0.651-0.806) for diagnosing fibrosis $\geq F2$ and 0.821 (0.741-0.901) for diagnosing fibrosis $\geq F3$ ³². The study cohort in their research differs from ours, exhibiting higher AST (32.2 ± 23.9 U/L) and ALT (50.3 ± 41.7 U/L) levels but lower thrombocyte values ($222.6 \pm 82.4 \times 10^9/L$). Additionally, their mean VCTE is higher (8.9 kPa ± 10.6), leading to a greater proportion of participants with advanced fibrosis stages than our study cohort. Furthermore, they included individuals seen at the gastroenterology department, which led to a selection bias that could explain the better AUROC that was found, as performance is dependent on the prevalence. Screening is, however, to be performed in low-prevalence settings, as in our cohort, making our observation very relevant when it comes to the reliability of FIB-4 as a first-line screening test. A meta-analysis by Sun *et al.* compared the FIB-4 to the MASLD fibrosis score and BARD score to predict advanced fibrosis in a biopsy-proven MASLD cohort ³³. They found

that the 1.30 cut-off value had a sensitivity of 0.844 (0.772-0.901) and a specificity of 0.685 (0.654-0.716) with an AUROC of 0.8496 ± 0.0680 . These results were, hence, also considerably better than ours, though this could be attributed to differences in the selected study cohorts. Only one of the four study cohorts in the meta-analysis collected data cross-sectionally; the other cohorts were retrospective and possibly suffered from selection bias.

As the current age-dependent cut-off values underperformed in the study cohort with low sensitivity but high NPV, we determined two new age-dependent cut-off values, 1.29 (≤ 65 years) and 1.72 (> 65 years). The newly suggested cut-off values were based on a higher sensitivity, as the prevalence of MASLD influences NPV. Validation in the two external cohorts showed that the newly proposed cut-offs only performed slightly better, with only a marginally higher sensitivity and NPV. Moreover, it was also noticed that the AUROC for the FIB-4 was acceptable in the Türkiye cohort, especially in people older than 65 years. These results should be interpreted with caution, as the likely small sample size in this subgroup is reflected in the excessively wide 95% CI. In the T2DM cohort, however, the performance of the FIB-4 was mediocre at best, independent of the current or suggested cut-off values used. Additionally, the agreement was especially low in the population without T2DM, while it was slightly higher in the population with T2DM. This underscores the fact that FIB-4 does not perform well in the well-established risk group of people with T2DM³⁴, as seen in previous research^{18,35,36}. For instance, Kim *et al.* compared the FIB-4 with liver biopsy and found an overall AUROC value of 0.79. When separating the study population into a group with T2DM and a group without, the AUROC of the FIB-4 for detecting advanced fibrosis was 0.68 for individuals with T2DM versus 0.85 for individuals without T2DM ($p = 0.003$)¹⁸. Another study by Bril *et al.* comparing the FIB-4 with VCTE results found an AUROC of 0.59 for significant fibrosis and 0.63 for advanced fibrosis in predicting fibrosis in individuals with T2DM³⁵. Furthermore, the FIB-4 and VCTE agreement was also poor in obese individuals in our study cohort. Foregoing research has shown that the FIB-4 has significantly lower AUROCs for advanced fibrosis in obese MASLD patients versus non-obese MASLD patients³⁷.

Lastly, in the study cohort, a false negative rate of 8.8% was found for those aged ≤ 65 years with a cut-off ≥ 1.3 , but when compared to those aged > 65 years, a false negative rate of 18.5% was found with a cut-off ≥ 2.0 . Usage of the FIB-4 consequently would lead to missing 10-20% of the fibrotic cases, or between 266 and 531 million people worldwide. These results, combined with the knowledge that the FIB-4 cannot be used in people younger than 35 years (the latter being a problem as the MASLD population grows younger³⁸), lead us to conclude

that the FIB-4 is not an ideal screening tool to use both in the general population and in at-risk groups³⁹. This also emphasises that it is vital to keep the context of use in mind when designing new scores or determining different cut-off values. Moreover, one must bear in mind that the parameters used in the FIB-4 formula are indirect markers for estimating fibrosis severity.

We therefore propose using other non-invasive scores specifically designed to screen for fibrotic MASLD, like the recently proposed MAF-5 score⁴⁰. The MAF-5 outperformed the FIB-4 in our dataset with an AUROC of 0.766 (0.701-0.830). However, the FIB-4 could possibly be used as a predictor for fibrosis progression, liver and non-related liver events, and mortality, as several studies have shown⁴¹⁻⁴⁵. The FIB-4 had a prognostic accuracy for fibrosis progression ranging from an AUROC of 0.65 (0.64-0.76) to 0.81 (0.73-0.89). Nonetheless, the details of the corresponding threshold values are poorly reported. One study used a threshold of 0.2 with an AUROC of 0.68 (0.60-0.76), an NPV of 91%, and a PPV of 38.3%⁴⁵. As for liver-related events and mortality, the AUROC ranged from 0.71 to 0.89 and 0.67 to 0.82, respectively⁴⁴. While the data from these studies seem promising, Hagström *et al.* concluded that the FIB-4 has limited predictive value for liver-related events in a primary care cohort, especially on an individual patient level⁴⁶. Although it may be useful on a population scale, its reliability for individual risk assessment is poor. Furthermore, the FIB-4 cannot be used as a treatment response marker, as a lot of other factors influence liver transaminase levels.

This study possesses several notable strengths that enhance its scientific rigour and the reliability of its findings. A key strength of this study is that the main cohort was recruited in primary care, making it highly representative of the real-world context in which screening approaches are intended to be used. Unlike most studies that develop biomarkers for screening in secondary care or highly selected populations, this study directly assesses their performance in a primary care setting. Additionally, the relatively low prevalence of advanced disease in this cohort enables a more reliable evaluation of diagnostic accuracy within the intended screening population of those in primary care. The next major strength is the large sample size, which met the calculated requirements, ensuring sufficient precision in estimating agreement. The results were further strengthened by using validated measurement methods, contributing to their credibility. A unique aspect of the study was its single-meeting design, which simultaneously minimises temporal variability by collecting all measurements, such as anthropometric data and FibroScan® assessments. This approach eliminates potential discrepancies arising from changes in participants' conditions over time and ensures that the data represent their health status at a specific moment. Additionally, this design reduced the risk of participant dropout,

further supporting the achievement of a robust sample size. The study's broad inclusion criteria are another significant strength, as they allowed for the inclusion of individuals from diverse risk groups, such as those with T2DM and obesity. This diversity ensures that the results are not limited to a single demographic or clinical population, enhancing the generalizability of the findings to various groups. Finally, the well-characterised population and detailed data collection ensured robust data quality, providing a strong foundation for the study's conclusions and adding value to its contribution to the field.

This study has several limitations that should be considered when interpreting the findings. First, the lack of liver biopsy, the gold standard for diagnosing liver fibrosis, is a potential limitation⁴⁷. While FibroScan® is widely used, it does not entirely replace the precision of biopsy-based diagnosis. This was addressed by using FibroScan® as the reference standard, given its widespread validation and use for diagnosing liver fibrosis. Secondly, the number of participants with fibrosis grades F2 or F3-F4 was limited, making it impossible to assess whether the FIB-4 performs differently in individuals with higher grades of fibrosis. Moreover, FibroScan results can vary slightly depending on the operator's skill, although the investigators ensured adequate training before initiating the study. Also, only measurements meeting the quality standard criteria were included. Third, selection bias could not be completely avoided; flyers were distributed in the waiting rooms of endocrinology departments and primary care clinics, but these were mainly taken by older people, as children were excluded from the study. This led to a higher median age in the discovery cohort. Lastly, some participants were referred by their GP or endocrinologist based on abnormal laboratory results, which could further influence the study population.

5. Conclusions

The agreement between VCTE and the FIB-4 in this study was too low to support the use of FIB-4 as a reliable diagnostic test for liver fibrosis in individuals with and without risk of fibrosis. Moreover, the FIB-4 demonstrated poor discriminative ability compared to VCTE, with low sensitivity across the different cut-off values. Additionally, the high false-negative rate of the FIB-4 further limits its clinical applicability, making its use in routine practice challenging. These findings and previous research indicate that the FIB-4 is unsuitable for liver fibrosis screening. Further research is needed to refine cut-off values or explore alternative non-invasive diagnostic tests.

Declaration of interests

None.

Funding

Author L.JM.H is funded by a fellowship of Fonds Wetenschappelijk Onderzoek (FWO) (1S73421N).

Author Contributions

L.JM.H. and D.v.M. contributed to the conception and design, acquired data, analysed and interpreted data, drafted the manuscript, revised the manuscript critically for important intellectual content, and provided final approval of the version to be published. L.JM.H. and D.v.M. are also the guarantors of this work and, as such, had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. The co-authors G.D.U, F.I., M.S., C.v.S., S.F., G.R. and G.H.K. contributed to the conception and design, revised the manuscript critically for important intellectual content, and provided final approval of the version to be published.

Acknowledgments

The PhD author of this review is part of the 'Limburg Clinical Research Centre (LCRC), supported by the foundation Limburg Sterk Merk, province of Limburg, Flemish government, Hasselt University, Ziekenhuis Oost-Limburg, and Jessa Hospital. We want to acknowledge the staff of the following general practitioner practices: Groepspraktijk De Dam, Huisartsenbox, Gezondheidscentrum Sirona, Groepspraktijk Luce, W-Care Hoeselt, Dr. Gilio, Huisartsenpraktijk Medi-Mine, Huisartsen Termolen, Gezondheidscentrum Heer, Dokters van Hier, Huisartsenpraktijk Dorine Verschure and Huisartsenpraktijk Bandkeramiek.

References

1. Rinella ME, Lazarus JV, Ratzliff V, Francque SM, Sanyal AJ, Kanwal F, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. *J Hepatol*. Jun 20 2023;doi:10.1016/j.jhep.2023.06.003
2. Pappachan JM, Babu S, Krishnan B, Ravindran NC. Non-alcoholic Fatty Liver Disease: A Clinical Update. *Journal of clinical and translational hepatology*. Dec 28 2017;5(4):384–393. doi:10.14218/jcth.2017.00013
3. Tanaka N, Kimura T, Fujimori N, Nagaya T, Komatsu M, Tanaka E. Current status, problems, and perspectives of non-alcoholic fatty liver disease research. *World J Gastroenterol*. Jan 14 2019;25(2):163–177. doi:10.3748/wjg.v25.i2.163
4. Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. Sep 2022;7(9):851–861. doi:10.1016/s2468-1253(22)00165-0
5. Miao L, Targher G, Byrne CD, Cao YY, Zheng MH. Current status and future trends of the global burden of MASLD. *Trends Endocrinol Metab*. Aug 2024;35(8):697–707. doi:10.1016/j.tem.2024.02.007
6. Byrne CD, Targher G. NAFLD: a multisystem disease. *J Hepatol*. Apr 2015;62(1 Suppl):S47–64. doi:10.1016/j.jhep.2014.12.012
7. Francque SM, van der Graaff D, Kwanten WJ. Non-alcoholic fatty liver disease and cardiovascular risk: Pathophysiological mechanisms and implications. *J Hepatol*. Aug 2016;65(2):425–43. doi:10.1016/j.jhep.2016.04.005
8. Targher G, Byrne CD, Tilg H. NAFLD and increased risk of cardiovascular disease: clinical associations, pathophysiological mechanisms and pharmacological implications. *Gut*. Sep 2020;69(9):1691–1705. doi:10.1136/gutjnl-2020-320622
9. Velarde-Ruiz Velasco JA, Garcia-Jimenez ES, Garcia-Zermeno KR, Morel-Cerda EC, Aldana-Ledesma JM, Castro-Narro GE, et al. Extrahepatic complications of non-alcoholic fatty liver disease: Its impact beyond the liver. *Rev Gastroenterol Mex*. Sep 2 2019;Complicaciones extrahepaticas de la enfermedad por hígado graso no alcoholico: impacto mas alla del hígado. doi:10.1016/j.rgmex.2019.05.004
10. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatology (Baltimore, Md)*. May 2017;65(5):1557–1565. doi:10.1002/hep.29085
11. Younossi ZM, Stepanova M, Rafiq N, Makhlof H, Younoszai Z, Agrawal R, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. *Hepatology (Baltimore, Md)*. Jun 2011;53(6):1874–82. doi:10.1002/hep.24268
12. Ekstedt M, Hagström H, Nasr P, Fredrikson M, Stål P, Kechagias S, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology (Baltimore, Md)*. May 2015;61(5):1547–54. doi:10.1002/hep.27368
13. Sebastiani G, Alshaalan R, Wong P, Rubino M, Salman A, Metrakos P, et al. Prognostic Value of Non-Invasive Fibrosis and Steatosis Tools, Hepatic Venous Pressure Gradient (HVPG) and Histology in Nonalcoholic Steatohepatitis. *PLoS One*. 2015;10(6):e0128774. doi:10.1371/journal.pone.0128774
14. Leung JC, Loong TC, Wei JL, Wong GL, Chan AW, Choi PC, et al. Histological severity and clinical outcomes of nonalcoholic fatty liver disease in nonobese patients. *Hepatology (Baltimore, Md)*. Jan 2017;65(1):54–64. doi:10.1002/hep.28697
15. Kleiner DE, Brunt EM, Wilson LA, Behling C, Guy C, Contos M, et al. Association of Histologic Disease Activity With Progression of Nonalcoholic Fatty Liver Disease. *JAMA Netw Open*. Oct 2 2019;2(10):e1912565. doi:10.1001/jamanetworkopen.2019.12565

16. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J Hepatol.* Sep 2024;81(3):492–542. doi:10.1016/j.jhep.2024.04.031
17. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology (Baltimore, Md).* Jun 2006;43(6):1317–25. doi:10.1002/hep.21178
18. Kim RG, Deng J, Reaso JN, Grenert JP, Khalili M. Noninvasive Fibrosis Screening in Fatty Liver Disease Among Vulnerable Populations: Impact of Diabetes and Obesity on FIB-4 Score Accuracy. *Diabetes Care.* Oct 1 2022;45(10):2449–2451. doi:10.2337/dc22-0556
19. Bril F, Godínez Leiva E, Lomonaco R, Shrestha S, Kalavalapalli S, Gray M, et al. Assessing strategies to target screening for advanced liver fibrosis among overweight and obese patients. *Metab Target Organ Damage.* 2022;2doi:10.20517/mtod.2022.08
20. Graupera I, Thiele M, Serra-Burriel M, Caballeria L, Roulot D, Wong GL, et al. Low Accuracy of FIB-4 and NAFLD Fibrosis Scores for Screening for Liver Fibrosis in the Population. *Clin Gastroenterol Hepatol.* Nov 2022;20(11):2567–2576.e6. doi:10.1016/j.cgh.2021.12.034
21. Thévenot T, Vendeville S, Weil D, Akkouché L, Calame P, Canivet CM, et al. Systematic screening for advanced liver fibrosis in patients with coronary artery disease: The CORONASH study. *PLoS One.* 2022;17(5):e0266965. doi:10.1371/journal.pone.0266965
22. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD): Executive Summary. *Diabetologia.* Nov 2024;67(11):2375–2392. doi:10.1007/s00125-024-06196-3
23. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organ Tech Rep Ser.* 2000;894:i–xii, 1–253.
24. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet.* Jan 10 2004;363(9403):157–63. doi:10.1016/s0140-6736(03)15268-3
25. Alberti KG, Zimmet P, Shaw J. The metabolic syndrome--a new worldwide definition. *Lancet.* Sep 24–30 2005;366(9491):1059–62. doi:10.1016/s0140-6736(05)67402-8
26. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research.* 2004;13:251 – 271.
27. Paranaguá-Vezozzo DC, Andrade A, Mazo DF, Nunes V, Guedes AL, Ragazzo TG, et al. Concordance of non-invasive mechanical and serum tests for liver fibrosis evaluation in chronic hepatitis C. *World J Hepatol.* Mar 18 2017;9(8):436–442. doi:10.4254/wjh.v9.i8.436
28. Quek J, Chan KE, Wong ZY, Tan C, Tan B, Lim WH, et al. Global prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in the overweight and obese population: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol.* Jan 2023;8(1):20–30. doi:10.1016/s2468-1253(22)00317-x
29. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. doi:10.1136/bmj.b2393
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* Mar 1977;33(1):159–74.
31. Kjaergaard M, Lindvig KP, Thorhauge KH, Andersen P, Hansen JK, Kastrup N, et al. Using the ELF test, FIB-4 and NAFLD fibrosis score to screen the population for liver disease. *J Hepatol.* Aug 2023;79(2):277–286. doi:10.1016/j.jhep.2023.04.002

32. Ding Y, Wang Z, Niu H, Deng Q, Wang Y, Xia S. FIB-4 is closer to FibroScan screen results to detecting advanced liver fibrosis and maybe facilitates NAFLD warning. *Medicine (Baltimore)*. Aug 25 2023;102(34):e34957. doi:10.1097/md.00000000000034957
33. Sun W, Cui H, Li N, Wei Y, Lai S, Yang Y, et al. Comparison of FIB-4 index, NAFLD fibrosis score and BARD score for prediction of advanced fibrosis in adult patients with non-alcoholic fatty liver disease: A meta-analysis study. *Hepatol Res*. Aug 2016;46(9):862–70. doi:10.1111/hepr.12647
34. Nogueira JP, Cusi K. Role of insulin resistance in the development of nonalcoholic fatty liver disease in people with type 2 diabetes: from bench to patient care. *Diabetes Spectrum*. 2024;37(1):20–28.
35. BRIL F, GRAY M. 1321-P: Screening for Liver Fibrosis in Diabetes—Assessing the Performance of Current Recommendations in a Large Unselected Population. *Diabetes*. 2023;72(Supplement_1)doi:10.2337/db23-1321-P
36. Ito T, Nguyen VH, Tanaka T, Park H, Yeh ML, Kawanaka M, et al. Poor Diagnostic Efficacy of Noninvasive Tests for Advanced Fibrosis in Obese or Younger Than 60 Diabetic NAFLD patients. *Clin Gastroenterol Hepatol*. Apr 2023;21(4):1013–1022.e6. doi:10.1016/j.cgh.2022.05.015
37. Joo SK, Kim W, Kim D, Kim JH, Oh S, Lee KL, et al. Steatosis severity affects the diagnostic performances of noninvasive fibrosis tests in nonalcoholic fatty liver disease. *Liver Int*. Feb 2018;38(2):331–341. doi:10.1111/liv.13549
38. Chacón C, Arteaga I, Martínez-Escudé A, Ruiz Rojano I, Lamonja-Vicente N, Caballeria L, et al. Clinical epidemiology of non-alcoholic fatty liver disease in children and adolescents. The LiverKids: Study protocol. *PLoS One*. 2023;18(10):e0286586. doi:10.1371/journal.pone.0286586
39. McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, et al. Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. *Am J Gastroenterol*. May 2017;112(5):740–751. doi:10.1038/ajg.2016.453
40. van Kleef LA, Francque SM, Prieto-Ortiz JE, Sonneveld MJ, Sanchez-Luque CB, Prieto-Ortiz RG, et al. Metabolic Dysfunction-Associated Fibrosis 5 (MAF-5) Score Predicts Liver Fibrosis Risk and Outcome in the General Population With Metabolic Dysfunction. *Gastroenterology*. Jul 2024;167(2):357–367.e9. doi:10.1053/j.gastro.2024.03.017
41. Anstee QM, Berentzen TL, Nitze LM, Jara M, Jensen AB, Kjær MS, et al. Prognostic utility of Fibrosis-4 Index for risk of subsequent liver and cardiovascular events, and all-cause mortality in individuals with obesity and/or type 2 diabetes: a longitudinal cohort study. *Lancet Reg Health Eur*. Jan 2024;36:100780. doi:10.1016/j.lanepe.2023.100780
42. Mózes FE, Lee JA, Vali Y, Alzoubi O, Staufer K, Trauner M, et al. Performance of non-invasive tests and histology for the prediction of clinical outcomes in patients with non-alcoholic fatty liver disease: an individual participant data meta-analysis. *Lancet Gastroenterol Hepatol*. Aug 2023;8(8):704–713. doi:10.1016/s2468-1253(23)00141-3
43. Boursier J, Hagström H, Ekstedt M, Moreau C, Bonacci M, Cure S, et al. Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events. *J Hepatol*. May 2022;76(5):1013–1020. doi:10.1016/j.jhep.2021.12.031
44. Lee J, Vali Y, Boursier J, Spijker R, Anstee QM, Bossuyt PM, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review. *Liver Int*. Feb 2021;41(2):261–270. doi:10.1111/liv.14669
45. Chalasani N, Abdelmalek MF, Loomba R, Kowdley KV, McCullough AJ, Dasarathy S, et al. Relationship between three commonly used non-invasive fibrosis biomarkers and improvement in fibrosis stage in patients with non-alcoholic steatohepatitis. *Liver Int*. May 2019;39(5):924–932. doi:10.1111/liv.13974

46. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. *J Hepatol*. Nov 2020;73(5):1023–1029. doi:10.1016/j.jhep.2020.06.007
47. Chowdhury AB, Mehta KJ. Liver biopsy for assessment of chronic liver diseases: a synopsis. *Clin Exp Med*. Jun 2023;23(2):273–285. doi:10.1007/s10238-022-00799-z

Journal Pre-proof

Table 1: Characteristics of the different cohorts.

Characteristic	Discovery cohort (n=563)	Validation cohort	
		Türkiye (n=493)	T2DM (n=229)
Setting	Primary care	Primary care	Hospital
Demographic			
Age (years)	62 [54-70]	52 [44-61]	61 [56-68]
Gender (male)	265 (47.1)	307 (62.3)	136 (59.4)
Smoker			
Never smoked	299 (53.1)	300 (60.9)	94 (41.0)
Ex-smoker	209 (37.1)	41 (8.3)	94 (41.0)
Smoker	55 (9.8)	152 (30.8)	41 (17.9)
Alcohol use ^a	413 (73.5)	263 (53.3)	65 (28.4)
Metabolic health			
BMI (kg/m ²)	28.2 [25.0-31.6]	28.0 [25.2-31.3]	31.6 [28.5-35.6]
BMI \geq 25 kg/m ²	426 (75.7)	418 (84.8)	214 (93.4)
Waist circumference (cm)	99.5 \pm 14.5	93.0 [87.0-103.0]	108.5 [98.0-117.0]
Diabetes type 2	80 (14.2)	148 (30.0)	229 (100.0)
Hypertension or treatment with anti-hypertensives	220 (39.1)	260 (52.47)	188 (82.1)
Metabolic syndrome	298 (55.6)	240 (48.7)	181 (95.8) ^b
Biochemistry			
HDL (mg/dL)	53.0 [43.0-65.0]	48.0 [39.0-59.0]	42.0 [36.0-49.0]
LDL (mg/dL)	106.0 [82.0-133.0]	115.0 [86.0-150.0]	60.0 [46.0-82.0]
Total cholesterol	187.0 [161.0-214.0]	195.0 [159.0-228.0]	138.0 [121.0-169.0]
Triglycerides (mg/dL)	116.0 [84.0-163.0]	123.0 [88.0-169.8]	163.0 [110.0-232.0]
AST (U/L)	23.0 [19.0-29.0]	19.8 [16.0-24.0]	26.0 [19.0-36.0]
ALT (U/L)	24.0 [17.0-36.0]	21.0 [15.9-31.0]	30.0 [20.5-50.5]
GGT (U/L)	25.0 [17.0-39.3]	21.0 [14.0-33.4]	42.0 [25.0-84.0]
Thrombocytes	250.0 [210.0-286.0]	248.0 [212.0-288.5]	215.0 [178.0-275.0]
Liver parameters			
VCTE (kPa)	5.2 [4.1-6.5]	4.9 [4.0-6.1]	7.6 [5.6-11.4]
Fibrosis \geq F2	78 (13.9)	30 (6.1)	107 (46.7)
Fibrosis \geq F3-F4	20 (3.6)	8 (1.6)	51 (22.3)
CAP (dB/m)	270.0 [225.0-316.0]	241.5 \pm 48.4	312.0 [267.0-352.5]
Steatosis	362 (64.3)	219 (44.4)	190 (83.0)

Data are presented as mean \pm standard deviation, median [interquartile range], or number (%).

^aAlcohol use was defined as the consumption of alcohol but without excessive alcohol usage (<2/3E/day women/men) as those participants were excluded. ^bn=189

Abbreviations: ALT: alanine aminotransferase, AST: aspartate aminotransferase, BMI: body mass index, CAP: controlled attenuation parameter, GGT: gamma-glutamyl transpeptidase, FAST: FibroScan-AST, HDL: high-density lipoprotein, LDL: low-density lipoprotein, MetS: metabolic syndrome, T2DM: type 2 diabetes mellitus, VCTE: vibration-controlled transient elastography.

Journal Pre-proof

Table 2: Weighted Cohen's kappa for the complete and imputed case analysis stratified for different risk groups.

Population	Complete case analysis		Imputed case analysis	
	n	Weighted kappa (95% CI)	n	Weighted kappa (95% CI)
Overall	482	0.117 (0.027-0.208)	563	0.138 (0.069-0.207)
Demographic				
Male	227	0.089 (-0.039-0.218)	265	0.090 (-0.012-0.193)
Female	255	0.130 (0.008-0.253)	298	0.178 (0.086-0.269)
Age \leq 65	298	0.139 (0.024-0.253)	347	0.150 (0.037-0.262)
Age > 65	184	0.093 (-0.050-0.236)	216	0.075 (-0.003-0.154)
Clinical characteristic				
Normal weight	108	0.147 (-0.008 -0.303)	133	0.071 (-0.020-0.162)
Overweight	187	0.122 (-0.027-0.271)	221	0.141 (0.028-0.253)
Obese	183	0.116 (-0.024-0.255)	205	0.192 (0.069-0.315)
High waist circumference	388	0.131 (0.033-0.229)	453	0.168 (0.088-0.245)
Metabolic health				
T2DM	104	0.236 (0.056-0.415)	112	0.272 (0.119-0.426)
No T2DM	378	0.056 (-0.032-0.144)	451	0.074 (0.007-0.141)
Metabolic syndrome	264	0.149 (0.030-0.268)	298	0.173 (0.079-0.268)

Abbreviations: CI: confidence interval, T2DM: type 2 diabetes mellitus.

Table 3: Sensitivity, specificity, PPV, and NPV of the current and proposed cut-off values in the discovery cohort.

Cut-off value	Complete case analysis (n=482)				Imputed case analysis (n=563)			
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
1.3 (current for age \leq 65)	40.6 (25.5 – 57.7)	80.1 (74.9 – 84.4)	19.7 (11.9 – 30.8)	91.8 (87.6 – 94.7)	40.6 (23.8-57.4)	80.5 (75.9-85.2)	20.9 (11.2-30.6)	91.5 (88.0-94.9)
1.29 (suggested for age \leq 65)	40.6 (25.5 – 57.7)	79.7 (74.5 – 84.1)	19.4 (11.7 – 30.4)	91.8 (87.5 – 94.7)	40.6 (23.8-57.4)	79.8 (75.1-84.5)	20.3 (10.8-29.7)	91.4 (87.9-94.9)
2.0 (current for age $>$ 65)	27.0 (15.4 – 43.0)	81.0 (73.8 – 86.5)	26.3 (15.0 – 42.0)	81.5 (74.4 – 87.0)	26.1 (12.2-40.0)	82.3 (76.5-88.2)	24.6 (11.2-37.9)	83.5 (78.0-89.0)
1.72 (suggested for age $>$ 65)	56.8 (40.9 – 71.3)	66.7 (58.7 – 73.8)	30.0 (20.5 – 41.5)	86.0 (78.4 – 91.2)	55.8 (39.9-71.8)	68.1 (60.9-75.3)	27.8 (17.8-37.9)	87.5 (81.9-93.1)

CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value.

Table 4: Weighted Cohen's kappa with new cut-off values in the validation cohorts.

Population	1.3 (current for age \leq 65)	1.29 (suggested for age \leq 65)	2.0 (current for age $>$ 65)	1.72 (suggested for age $>$ 65)
	Weighted kappa (95% CI)	Weighted kappa (95% CI)	Weighted kappa (95% CI)	Weighted kappa (95% CI)
Türkiye (n=493)	0.126 (0.004-0.248)	0.115 (-0.001-0.231)	0.266 (0.009-0.524)	0.182 (-0.015-0.379)
T2DM (n=229)	0.155 (-0.007-0.317)	0.187 (0.029-0.346)	0.187 (-0.012-0.386)	0.101 (-0.113-0.314)

CI: confidence interval.

Table 5: Sensitivity, specificity, PPV, and NPV of the current and proposed cut-off values in the validation cohorts.

	Türkiye	T2DM
--	---------	------

Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
1.3 (current for age \leq 65)	33.3 (17.2-54.6)	88.4 (84.8-91.2)	13.2 (6.5-24.8)	96.1 (93.6-97.7)	42.6 (31.0-55.1)	72.5 (61.9-81.1)	54.2 (40.3-67.4)	62.4 (52.2-71.5)
1.29 (suggested for age \leq 65)	33.3 (17.2-54.6)	87.5 (83.9-90.4)	12.3 (6.1-23.2)	96.1 (93.6-97.7)	44.4 (32.8-56.7)	73.8 (63.5-82.0)	56.0 (42.3-68.8)	63.9 (54.0-72.8)
2.0 (current for age $>$ 65)	55.6 (26.7-81.1)	81.3 (70.0-88.9)	29.4 (13.3-53.1)	92.9 (83.0-97.2)	45.5 (31.7-59.9)	73.7 (58.0-85.0)	66.7 (48.8-80.8)	53.8 (40.5-66.7)
1.72 (suggested for age $>$ 65)	66.7 (35.4-87.9)	67.2 (55.0-77.4)	22.2 (10.6-40.8)	93.5 (82.5-97.8)	52.3 (37.9-66.2)	57.9 (42.2-72.1)	59.0 (43.4-72.9)	51.2 (36.8-65.4)

CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value.

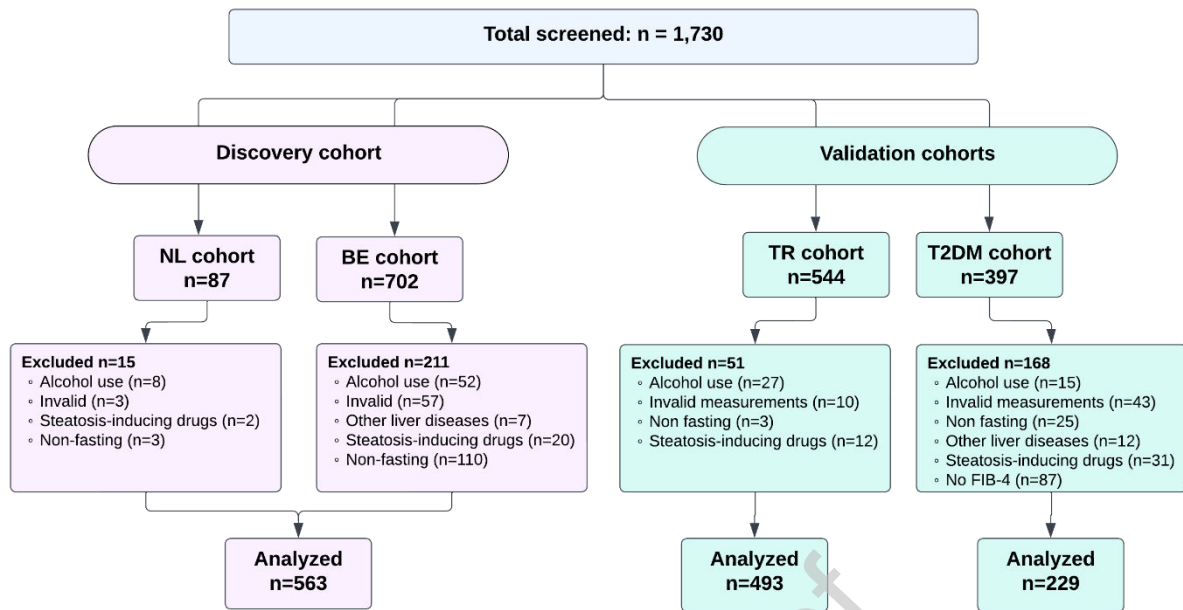


Figure 1: Flowchart of the inclusion process. NL: The Netherlands, BE: Belgium, TR: Türkiye, T2DM: type 2 diabetes mellitus.

