ADDRESSING VERIFICATION BIAS IN MASLD SCREENING: A BAYESIAN LATENT CLASS EVALUATION OF FIB-4 AND APRI IN A LOW-PREVALENCE OCCUPATIONAL COHORT

Nguyen Huu Ngoc Tuan✉, Nguyen Minh Ha

Pham Ngoc Thach University of Medicine

Non-invasive tests (NITs), such as the Fibrosis-4 (FIB-4) index and the AST to Platelet Ratio Index (APRI), are recommended for stratifying risk of advanced fibrosis in metabolic dysfunction-associated steatotic liver disease (MASLD), yet their true diagnostic accuracy in primary care is often obscured by verification bias. This study evaluated the operational yield and model-based diagnostic performance of these NITs in a low-prevalence occupational health-check cohort of 238 adults. Operational analysis revealed high overall rule-out yields using pre-specified low-risk thresholds (FIB-4 < 1.30: 93.3%; APRI < 0.50: 92.4%), although the FIB-4 yield declined precipitously in patients aged >60 years old (28.6%). To compensate for the absence of a histological reference standard, a Bayesian Latent Class Analysis (LCA) adjusting for conditional dependence was employed, estimating a true advanced fibrosis prevalence of 2.6% (95% CrI: 0.5%-7.3%). Adjusted for correlated errors, FIB-4 demonstrated superior sensitivity compared to APRI (73.1% vs. 57.8%) while maintaining comparable specificity (92.7% vs. 91.8%). Furthermore, Decision Curve Analysis confirmed that an FIB-4 guided triage strategy maximizes net clinical benefit across relevant threshold probabilities. These findings indicate that FIB-4 is a highly efficient primary care gatekeeping tool, superior to APRI, provided age-adjusted thresholds are applied to older adults.

Keywords: APRI, Bayesian latent class analysis, Decision curve analysis, FIB-4, MASLD.

I. INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD) has emerged as the most prevalent chronic liver disease globally, affecting approximately 30% of the adult population and imposing a profound clinical and economic burden.[1-3] The 2023 international Delphi consensus, which transitioned the nomenclature from non-alcoholic fatty liver disease (NAFLD) to MASLD, reflects a modernized, affirmative understanding of the cardiometabolic drivers inherent to hepatic steatosis and fibrogenesis.[4] Within primary care and routine occupational health-check settings, the paramount clinical challenge is identifying the small subset of patients harboring advanced fibrosis-who are at the highest risk for progression to cirrhosis and hepatocellular carcinoma-among a vast, largely asymptomatic population.

Major international hepatology societies, including recent 2024 updates from the European Association for the Study of the Liver (EASL), strongly advocate for the use of simple, blood-based non-invasive tests (NITs) as the mandatory first-line triage step.[5,6] Algorithms such as the Fibrosis-4 (FIB-4) index and the AST to Platelet Ratio Index (APRI) rely on routine laboratory parameters to estimate the risk of fibrosis.[7,8]

Among available blood-based NITs, several composite scores have been developed, including the NAFLD fibrosis score (NFS), the

Corresponding author: Nguyen Huu Ngoc Tuan
Pham Ngoc Thach University of Medicine
Email: nhntuan@pnt.edu.vn
Received: 11/03/2026
Accepted: 01/04/2026

BARD score, and the Enhanced Liver Fibrosis (ELF) test. However, FIB-4 and APRI remain the most widely endorsed by major clinical practice guidelines from EASL, AASLD, and the American Gastroenterological Association, principally because they rely exclusively on routinely available laboratory parameters (age, AST, ALT, and platelet count) without requiring proprietary biomarker assays. This renders them uniquely suitable for deployment in primary care and resource-constrained occupational health-check settings, where specialized testing infrastructure is often unavailable.

Despite the widespread integration of NITs into clinical pathways, validating their diagnostic performance in primary care cohorts presents methodological challenges. Most retrospective screening studies are affected by verification bias. While standard Latent Class Analysis (LCA) is occasionally employed to estimate test accuracy without a gold standard, it strictly assumes local statistical independence between the tests being evaluated. This assumption is critically violated when evaluating FIB-4 and APRI simultaneously, as both algorithms are mathematically coupled via shared aspartate aminotransferase (AST) and platelet parameters.[9-11]

To address these critical methodological gaps, this study evaluates the operational performance and true diagnostic utility of FIB-4 and APRI within a Vietnamese MASLD health-check cohort -a population of working adults subjected to routine annual screening, representative of the primary care tier where early detection efforts are most needed and where disease prevalence is expected to be low- using a Bayesian LCA that accounts for the conditional dependence between FIB-4 and APRI. Moreover, we translate these corrected parameters into a Decision Curve Analysis (DCA) framework to quantitatively evaluate the net clinical benefit of NIT-guided referral strategies.[12] This integration of operational metrics and disease modeling may provide evidence-based guidance for optimizing MASLD screening pathways in primary care ecosystems.

II. SUBJECTS AND METHODS

1. Study Design and Setting

This study is a secondary, cross-sectional analysis utilizing existing data derived from a broader prospective health-check cohort. The primary objective of this specific analysis was to evaluate the operational rule-out yield and model the diagnostic accuracy of non-invasive liver fibrosis tests (FIB-4 and APRI). The source population comprised medical and administrative staff from Nguyen Tri Phuong (NTP) Hospital and Pham Ngoc Thach (PNT) University of Medicine, who underwent their annual occupational health examinations. It is important to note that this occupational health-check population may be subject to a 'healthy worker effect,' potentially underrepresenting individuals with advanced metabolic disease; the implications of this selection bias for test performance estimates are addressed in the Discussion.

2. Study Population and Sampling

The parent cohort data were collected between October 31, 2022 (7 days after the earliest ethical committee approval) and December 21, 2023 (30 days after the latest ethical committee approval). During this period, a total of 700 participants were recruited for the primary studies.

For this secondary manuscript, a refined analytical subset of 238 participants was extracted from the existing dataset. Inclusion in this specific sub-analysis required participants to meet the 2023 consensus criteria for MASLD,

defined by ultrasonographic evidence of hepatic steatosis in the presence of at least one cardiometabolic risk factor (e.g., Asian-specific BMI ≥ 23 kg/m², dyslipidemia, or impaired fasting glucose).[4] Exclusion criteria included positive serology for viral hepatitis (HBsAg or anti-HCV), significant alcohol consumption, or missing concurrent data for the critical laboratory variables required to compute the index tests (age, AST, ALT, and platelet count).

Data Collection and Index Tests

All clinical, anthropometric, and laboratory data utilized in this analysis were retrieved from the same clinical visit to prevent temporal bias. The non-invasive indices were calculated a priori using the following established formulas:

- FIB-4 Index = (Age [years] × AST [U/L]) / (Platelets [10^9/L] × √ALT [U/L])

- APRI Score = ((AST [U/L] / ULN AST [U/L]) / Platelets [10^9/L]) × 100

The upper limit of normal (ULN) for AST was defined according to local institutional laboratory reference standards. While the use of a standardized ULN (e.g., 40 U/L) is often preferred for external comparability, the local ULN was retained in this analysis to reflect real-world clinical practice. A sensitivity analysis employing a standardized ULN was beyond the scope of this secondary analysis but is acknowledged as a limitation. The primary outcome was the rule-out yield, defined as the proportion of participants scoring strictly below established low-risk cutoffs (FIB-4 < 1.30 and APRI < 0.50).

Statistical Analysis

Statistical reporting was adapted from the STARD (Standards for Reporting of Diagnostic Accuracy) guidelines to account for the absence of an elastographic or histological reference standard.[13] Categorical variables were summarized as frequencies and percentages, while continuous variables exhibiting right-skewed distributions were presented as medians with interquartile ranges (IQR). Rule-out yields were reported with 95% Wilson binomial confidence intervals. Diagnostic concordance between FIB-4 and APRI classifications was quantified using Cohen's kappa (κ) and McNemar's test. Multivariable logistic regression was applied to identify independent determinants of achieving a rule-out classification.

To formally estimate the true diagnostic accuracy and the prevalence of unobserved advanced fibrosis, a Bayesian Latent Class Analysis (LCA) was employed. The Bayesian hierarchical model utilized Markov Chain Monte Carlo (MCMC) sampling (50,000 iterations). Weakly informative beta priors (Beta(1,1)) were assigned for sensitivity and specificity parameters to allow the data to predominantly drive the posterior estimates, while a uniform prior on prevalence reflected the absence of strong external prevalence data in this specific occupational setting and explicitly parameterized the conditional dependence between FIB-4 and APRI (which share AST and platelet parameters) to prevent biased estimates. Subsequently, Decision Curve Analysis (DCA) was utilized to calculate the net clinical benefit of the index tests across a continuum of theoretical risk thresholds. All analyses were executed using R statistical software, with two-sided p-values < 0.05 considered statistically significant.

3. Ethical Considerations

The parent studies from which this data was derived, as well as the data collection protocols, were conducted in rigorous accordance with the Declaration of Helsinki. Informed consent was obtained from all participating staff members

prior to their enrollment in the primary cohort. Institutional review board and ethical committee approvals covering the participant recruitment and data collection were independently granted by three separate regulatory bodies: The Biomedical Research Ethics Committee of the University of Medicine and Pharmacy at Ho Chi Minh City (Approval No. 788/HĐĐĐ-ĐHYD, dated October 24, 2022); The Biomedical Research Ethics Committee of Pham Ngoc Thach University of Medicine (Approval No. 859/TĐHYKPNT-HĐĐĐ, dated April 20, 2023); and The Biomedical Research Ethics Committee of Nguyen Tri Phuong Hospital (Approval No. 2572/NTP-HĐĐĐ, dated November 21, 2023). Official administrative clearance for data collection at the clinical site was granted by the Pham Ngoc Thach University of Medicine Polyclinic (Decision No. 319/QĐ-PKĐK, dated November 30, 2022).

III. RESULTS

1. Baseline Characteristics and Score Distributions

A total of 238 adult participants who met the diagnostic criteria for MASLD and had concurrent laboratory data were included in the primary analysis. The cohort was predominantly composed of young to middle-aged adults (median age 37.0 years old, IQR 32.0-46.0), with an approximately equal gender distribution (48.7% male, 51.3% female). The median BMI was 25.0 kg/m² (IQR 23.2-27.1), classifying the majority of the cohort as overweight or obese according to Asian-specific criteria. Type 2 diabetes was present in 4.2% of the cohort. Median aminotransferase levels were within standard normal limits (AST 22.7 U/L; ALT 26.6 U/L), and the median platelet count was 268.5 × 10^9/L.

Continuous distributions for both non-invasive indices were heavily right-skewed, reflecting a predominantly healthy, early-disease primary care population (Figure 1). The median FIB-4 score was 0.6 (IQR 0.5-0.9), and the median APRI score was 0.2 (IQR 0.2-0.3).

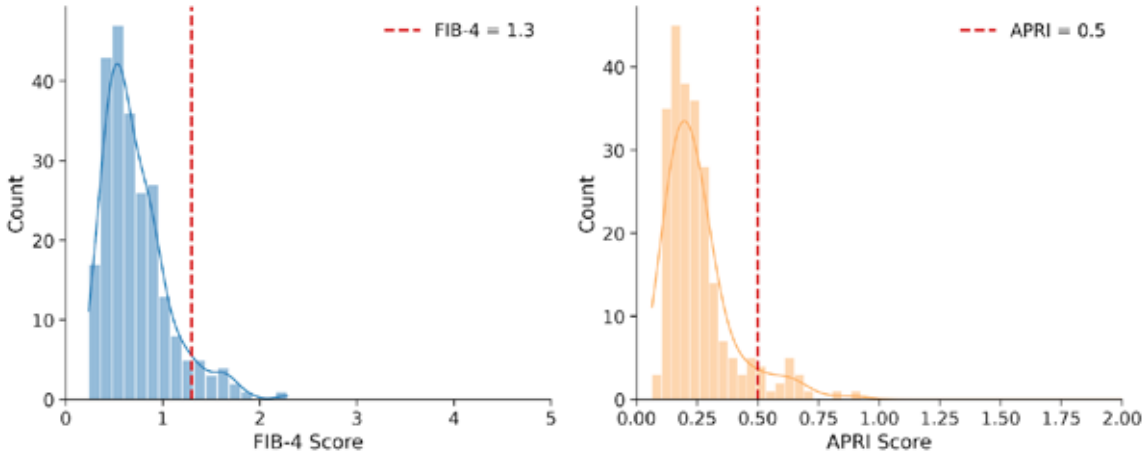

Figure 1. Density distributions of continuous FIB-4 and APRI scores in the MASLD cohort

The histograms, overlaid with kernel density estimates, display the right-skewed distributions of the Fibrosis-4 (FIB-4) index (left) and the AST to Platelet Ratio Index (APRI) (right) within the primary analytical sample (N = 238). The vertical red dashed lines denote the pre-specified low-risk "rule-out" clinical thresholds evaluated in this study (FIB-4 < 1.30 and APRI < 0.50).

Abbreviations: APRI, AST to Platelet Ratio Index; AST, aspartate aminotransferase; FIB-4, Fibrosis-4; MASLD, metabolic dysfunction-associated steatotic liver disease.

2. Primary Operational Metric: Rule-out Yield

Application of the pre-specified low-risk thresholds demonstrated a remarkably high operational yield for both tests (Table 1). In the overall cohort, the FIB-4 < 1.30 threshold successfully classified 93.3% (95% CI: 89.4-95.8) of patients as low risk, thereby theoretically excluding them from immediate requirement for specialized hepatology referral or transient elastography. The APRI < 0.50 threshold yielded a comparable rule-out proportion of 92.4% (95% CI: 88.4-95.2).

Stratification by demographic and metabolic parameters revealed significant variability in test performance, particularly concerning age. While the FIB-4 rule-out yield was 98.8% in patients under 45 years old, it fell to 84.5% in those aged 45-60 years old, and dropped precipitously to 28.6% (95% CI: 8.2-64.1) in patients over 60 years old. By contrast, APRI maintained a more stable yield across age strata, remaining at 85.7% in the oldest cohort. BMI and sex did not demonstrate clinically substantial impacts on the rule-out yield for either index.

Table 1. Primary Operational Metric - Rule-out Yield (%) by Pre-specified Cutoffs

Stratum	N	FIB-4 < 1.30 Yield % (95% CI)	APRI < 0.50 Yield % (95% CI)
Overall Cohort	238	93.3% (89.4 - 95.8)	92.4% (88.4 - 95.2)
Age Group			
< 45 years old	173	98.8% (95.9 - 99.7)	92.5% (87.6 - 95.6)
45 - 60 years old	58	84.5% (73.1 - 91.6)	93.1% (83.6 - 97.3)
> 60 years old	7	28.6% (8.2 - 64.1)	85.7% (48.7 - 97.4)
BMI Category (Asian criteria)			
Normal (< 23)	56	92.9% (83.0 - 97.2)	96.4% (87.9 - 99.0)
Overweight (23 - 27.5)	130	93.1% (87.4 - 96.3)	92.3% (86.4 - 95.8)
Obese (≥ 27.5)	52	94.2% (84.4 - 98.0)	88.5% (77.0 - 94.6)
Type 2 Diabetes			
No	228	93.9% (90.0 - 96.3)	92.5% (88.4 - 95.3)
Yes	10	80.0% (49.0 - 94.3)	90.0% (59.6 - 98.2)
Gender			
Male	116	92.2% (85.9 - 95.9)	87.9% (80.8 - 92.7)
Female	122	94.3% (88.6 - 97.2)	96.7% (91.9 - 98.7)

Note. 'Rule-out yield' is defined as the proportion of participants scoring strictly below the pre-specified low-risk diagnostic thresholds (FIB-4 < 1.30 and APRI < 0.50), theoretically avoiding the need for further specialized referral or transient elastography. 95% Confidence Intervals (CI) were computed using the Wilson score interval for binomial proportions. BMI was categorized utilizing established Asian-specific cutoffs (Normal < 23.0 kg/m², Overweight 23.0-27.4 kg/m², Obese ≥ 27.5 kg/m²).

Abbreviations: APRI, AST to Platelet Ratio Index; BMI, body mass index; CI, confidence interval; FIB-4, Fibrosis-4.

3. Diagnostic Concordance

Cross-tabulation of the binary risk classifications (low risk versus indeterminate/ high risk) showed a high absolute raw agreement between the two indices, driven primarily by the large number of dual "rule-out" classifications (n=209). However, statistical agreement beyond chance was only fair (Cohen's κ = 0.240). Discordant classifications occurred in 24 patients (10.1% of the cohort). McNemar's test indicated no significant directional bias in the discordance (p = 0.838), suggesting neither test was systematically categorizing more patients into the high-risk stratum than the other.

4. Clinical Determinants of Rule-Out Classification

Multivariable logistic regression was utilized to identify independent predictors of achieving a low-risk classification (Table 2). Consistent with the algebraic structure of the index, advancing age was a strong negative predictor for ruling out advanced fibrosis using FIB-4 (adjusted Odds Ratio [aOR] 0.84 per year; 95% CI: 0.78-0.90; p < 0.001). Higher baseline ALT levels also independently reduced the likelihood of meeting the rule-out criteria for both FIB-4 (aOR 0.98 per U/L; p = 0.002) and APRI (aOR 0.94 per U/L; p < 0.001). Neither sex, BMI, nor the presence of type 2 diabetes independently predicted the operational classification when controlling for aminotransferases and age.

Table 2. Multivariable Predictors of Achieving a "Rule-out" Classification

Predictor	FIB-4 < 1.30 aOR (95% CI)	p-value	APRI < 0.50 aOR (95%CI)	p-value
Age (per year)	0.84 (0.78 - 0.90)	< 0.001	0.93 (0.87 - 1.01)	0.070
Gender (Male vs. Female)	1.12 (0.32 - 3.83)	0.862	0.28 (0.04 - 2.00)	0.204
BMI (per kg/m²)	1.11 (0.88 - 1.40)	0.391	0.95 (0.73 - 1.23)	0.701
Diabetes (Yes vs. No)	0.60 (0.06 - 6.22)	0.669	0.31 (0.02 - 5.02)	0.408
ALT (per U/L)	0.98 (0.96 - 0.99)	0.002	0.94 (0.91 - 0.96)	< 0.001

Note. Results are derived from multivariable logistic regression models. Adjusted Odds Ratios (aOR) > 1.00 indicate a higher probability of being classified as low risk (i.e., achieving the rule-out threshold), whereas an aOR < 1.00 indicates a lower probability. Models were mutually adjusted for age, sex, BMI, type 2 diabetes status, and ALT levels. AST and platelet counts were explicitly excluded from the covariate matrix to avoid multicollinearity, as they are mathematically coupled components of the dependent variables (the index scores).

Abbreviations: ALT, alanine aminotransferase; aOR, adjusted odds ratio; APRI, AST to Platelet Ratio Index; BMI, body mass index; CI, confidence interval; FIB-4, Fibrosis-4.

5. Model-Based Diagnostic Accuracy and Clinical Utility

To account for verification bias in the absence of a histological reference standard, a Bayesian LCA incorporating conditional dependence adjustments was executed (Table 3). The model estimated a very low underlying latent prevalence of advanced fibrosis within this screening cohort at 2.6% (95% CrI: 0.5%-7.3%).

Adjusted for the correlated errors between the two tests, FIB-4 demonstrated superior sensitivity compared to APRI (73.1% [95% CrI: 52.2-88.4] vs. 57.8% [95% CrI: 35.4-77.6]). Both tests exhibited excellent specificity in this setting: 92.7% (95% CrI: 88.7-96.1) for FIB-4 and 91.8% (95% CrI: 87.2-95.1) for APRI.

Table 3. Bayesian LCA Posterior Estimates of Diagnostic Accuracy

Parameter	Posterior Median	95% Credible Interval (CrI)
Latent Prevalence of Advanced Fibrosis	2.6%	0.5% - 7.3%
FIB-4 (< 1.30 cutoff)		
Sensitivity	73.1%	52.2% - 88.4%
Specificity	92.7%	88.7% - 96.1%
APRI (< 0.50 cutoff)		
Sensitivity	57.8%	35.4% - 77.6%
Specificity	91.8%	87.2% - 95.1%

Note. In the absence of a histological reference standard, parameters were estimated utilizing a Bayesian Latent Class Analysis (LCA). Estimates reflect the median of the posterior probability distributions derived via Markov Chain Monte Carlo (MCMC) sampling (50,000 iterations; 10,000 burn-in). The 95% Credible Interval (CrI) is the Bayesian analogue to a confidence interval, indicating a 95% probability that the true parameter value falls within this range given the observed data and prior distributions. To prevent biased estimates, the hierarchical model explicitly parameterized the conditional dependence (covariance) between FIB-4 and APRI within the latent classes.

Abbreviations: APRI, AST to Platelet Ratio Index; CrI, credible interval; FIB-4, Fibrosis-4; LCA, Latent Class Analysis.

Translation of these accuracy parameters into a DCA framework confirmed the clinical utility of the non-invasive tests (Figure 2). Across a wide range of low threshold probabilities (pt) relevant to primary care (0.5% to 20%), initiating clinical pathways based on either the FIB-4 ≥ 1.30 or APRI ≥ 0.50 threshold yielded a higher net clinical benefit than either the default strategy of referring all patients or referring no patients. Furthermore, due to its superior sensitivity profile, the FIB-4-guided strategy consistently dominated the APRI strategy, yielding the highest net benefit across all evaluated risk thresholds.

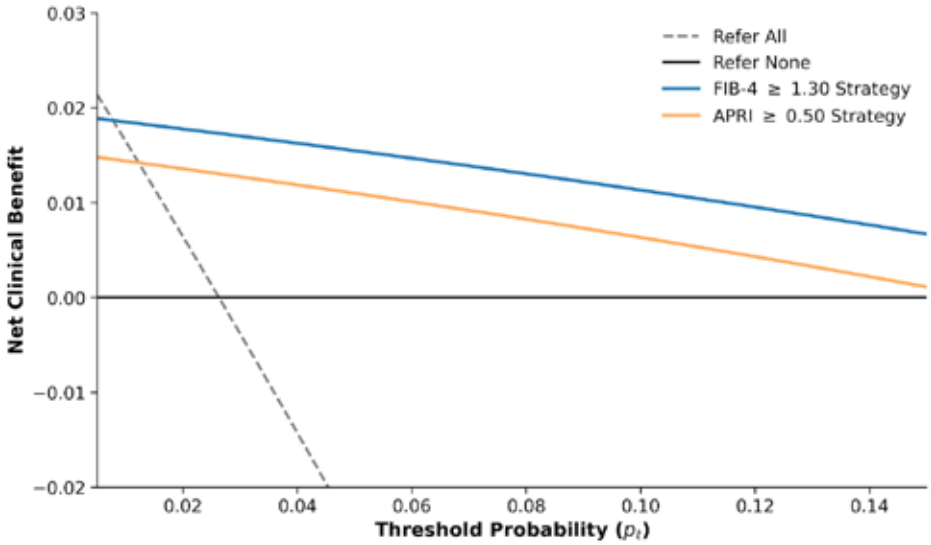

Figure 2. Model-based Decision Curve Analysis (DCA) of non-invasive test strategies for advanced fibrosis

The net clinical benefit of utilizing the FIB-4 ≥ 1.30 (blue line) and APRI ≥ 0.50 (green line) thresholds to guide clinical referral is plotted against a range of threshold probabilities (pt). Due to the absence of a biopsy reference standard, the net benefit calculations utilize the true prevalence and diagnostic accuracy parameters derived from the Bayesian Latent Class Analysis (LCA). Both index test strategies demonstrate superior net benefit compared to the default clinical strategies of referring all patients (dotted black line) or referring no patients (solid black line) across the clinically relevant low-probability thresholds typical of a primary care setting.

Abbreviations: APRI, AST to Platelet Ratio Index; DCA, Decision Curve Analysis; FIB-4, Fibrosis-4.

IV. DISCUSSION

In this cross-sectional analysis of a primary care MASLD cohort, we demonstrated that both the FIB-4 and APRI indices are highly effective gatekeeping tools, successfully classifying over 92% of the cohort as low risk for advanced fibrosis. By employing a Bayesian LCA to overcome the absence of a histological reference standard, we estimated the true underlying prevalence of advanced fibrosis in this setting to be exceptionally low (2.6%). Adjusted for verification bias and conditional dependence, FIB-4 exhibited superior sensitivity compared to APRI (73.1% vs. 57.8%) while maintaining comparable specificity (>91%). Crucially, DCA confirmed that a FIB-4-guided referral strategy provides the highest net clinical benefit across all relevant threshold probabilities, thereby supporting its role as a preferred primary care screening tool.

The global transition from NAFLD to the more inclusive MASLD nomenclature reflects a growing recognition of the synergistic hepatotoxic effects of metabolic dysfunction.[4] With MASLD now affecting up to 30% of the global population, specialist referral for all suspected cases is economically and logistically unfeasible. Our findings align seamlessly with recent clinical practice guidelines, which strongly advocate for the use of NITs as the first line of risk stratification.[6,14,15]

Our data underscore that the primary clinical value of FIB-4 and APRI lies in their negative predictive value. Recent health-economic models emphasize that, while NITs may miss a fraction of cases when used as confirmatory diagnostic tools, they are highly reliable as rule-out instruments, preventing unnecessary transient elastography or liver biopsy in primary care.[16] The 93.3% rule-out yield observed in our cohort indicates that implementing a FIB-4 < 1.30 threshold would safely spare the vast majority of patients from costly, specialized hepatology evaluations. The relatively modest Cohen's κ (0.240) despite high raw agreement (87.8%) is consistent with the well-documented 'kappa paradox' described by Feinstein and Cicchetti, wherein highly imbalanced marginal distributions-as expected when both tests classify over 90% of participants identically-mathematically suppress the κ statistic regardless of true concordance. This phenomenon does not invalidate the clinical utility of either test but rather reflects the statistical artifact inherent to agreement metrics in low-prevalence settings.

A critical observation in our cohort was the precipitous decline in the FIB-4 rule-out yield among older adults. While 98.8% of patients under 45 years old were successfully ruled out, this proportion collapsed to just 28.6% in patients over 60 years old. Because age operates as a direct multiplier in the numerator of the FIB-4 equation, physiological aging disproportionately inflates the score, independent of true fibrogenesis. Recent longitudinal cohort studies have shown that using a static 1.30 cutoff in geriatric populations yields an unacceptably high false-positive rate, overburdening hepatology clinics. Our demographic stratification robustly supports the current consensus to utilize age-adjusted cutoffs (e.g., FIB-4 < 2.00) for individuals aged 65 and older to maintain the test's specificity and operational efficiency.[17]

A primary methodological strength of this study is the rigorous application of Bayesian LCA. The vast majority of retrospective screening studies are fundamentally flawed by verification bias, as only patients with high NIT scores typically proceed to receive a "gold standard" biopsy or magnetic resonance elastography. Furthermore, standard LCA models fail to evaluate FIB-4 and APRI simultaneously because the tests are mathematically coupled (both use AST and platelet counts). By explicitly modeling the conditional dependence covariance matrix within our Bayesian framework, we generated less biased estimates of sensitivity and specificity that reflect true clinical performance.[9-11]

Additionally, the integration of DCA elevates the clinical relevance of our findings.[12] While traditional studies rely solely on the Area Under the Receiver Operating Characteristic (AUROC) curve-a metric divorced from clinical consequences-DCA directly quantifies the epidemiological harm versus benefit of adopting a testing strategy. Our DCA model provides visual and mathematical evidence that utilizing FIB-4 is superior to treating all patients or treating none, offering actionable evidence for local health policymakers formulating MASLD referral guidelines.

Despite these methodological strengths, our study has several notable limitations that warrant a cautious interpretation of the findings. First, the cross-sectional design precludes our ability to assess the longitudinal prognostic value of baseline FIB-4 and APRI scores for predicting downstream clinical events, such as hepatocellular carcinoma or liver-related mortality. Second, the sample size within certain demographic sub-strata was constrained. Specifically, the cohort included only a small

number of participants aged 60 years or older (n = 7). Consequently, the estimated FIB-4 rule-out yield for this older demographic exhibited a very wide 95% confidence interval (8.2%-64.1%); therefore, while the precipitous decline in yield observed in this group aligns with broader epidemiological expectations, the precise point estimate lacks high statistical precision in our specific sample and must be interpreted with caution. Third, while the Bayesian LCA effectively compensated for the lack of a histological reference standard, it remains a mathematical approximation fundamentally dependent on the accuracy, appropriateness, and generalizability of its prior distributions. Finally, reliance on the local laboratory upper limit of normal (ULN) for AST can introduce inherent variability into the APRI calculation, potentially limiting the direct comparability and external validity of our APRI rule-out yields across different healthcare networks. Additionally, the occupational health-check setting introduces a potential 'healthy worker effect,' whereby the study population may have a lower burden of metabolic disease compared to the general primary care population, potentially influencing the estimated prevalence and test performance parameters.

V. CONCLUSION

In a primary care MASLD cohort with a low prevalence of advanced disease, FIB-4 is demonstrated strong performance as a cost-effective rule-out tool. The application of age-adjusted thresholds is critical to maintaining its utility in older populations. Supported by rigorous Bayesian modeling and Decision Curve Analysis, these findings support the potential integration of FIB-4 into routine primary care workflows as a means to optimize specialist resource allocation, pending validation in larger, multi-center cohorts.

ACKNOWLEDGMENTS

The authors declare that no specific funding was received for this study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest regarding the publication of this manuscript.

REFERENCES

1. Teng ML, Ng CH, Huang DQ. The global landscape of lean metabolic dysfunction-associated steatotic liver disease: insight from Asia and the West. *Frontiers in Gastroenterology*. 2025; 4: 1699508.

2. Younossi ZM, Paik JM, Golabi P. Epidemiology of metabolic dysfunction-associated steatotic liver disease. *Clinical and Molecular Hepatology*. 2024; 30(Suppl 1): S1-S15.

3. Le P, Tatar M, Dasarathy S, et al. Estimated Burden of Metabolic Dysfunction-Associated Steatotic Liver Disease in US Adults, 2020 to 2050. *JAMA Netw Open*. Jan 2 2025; 8(1): e2454707. doi:10.1001/jamanetworkopen.2024.54707.

4. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Ann Hepatol*. Jan-Feb 2024; 29(1): 101133. doi:10.1016/j.aohep.2023.101133.

5. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology*. Mar 17 2023;doi:10.1097/HEP.0000000000000323.

6. 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). *Journal of Hepatology*. 2024; 81(3): 492-542.

7. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006; 43(6): 1317-1325.

8. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003; 38(2): 518-526.

9. Albert PS, Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics*. 2004; 60(2): 427-435.

10. Subtil A. Detecting departures from the conditional independence assumption in diagnostic latent class models: a simulation study. *Research methodology modeling*. 2024; 24(1): 45-58.

11. Dendukuri N, Hadgu A, Wang L. Modeling conditional dependence between diagnostic tests: a multiple latent variable model. *Statistics in Medicine*. 2009; 28(3): 441-461.

12. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*. 2006; 26(6): 565-574.

13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015; 351: h5527.

14. American Diabetes A. Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) in People With Diabetes: The Need for Screening and Early Intervention. A Consensus Report of the American Diabetes Association. *Diabetes Care*. 2025; 48(7): 1057-1075.

15. El-Kordi A. Identifying and managing liver disease in primary care. *BMJ*. 2025; 391:r2332.

16. Park H, Yoon EL, Kim M, et al. Cost-effectiveness of advanced hepatic fibrosis screening in individuals with suspected MASLD identified by serologic noninvasive tests. *Sci Rep*. Jul 7 2025; 15(1): 24186. doi:10.1038/s41598-025-08434-z.

17. McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *The American Journal of Gastroenterology*. 2017; 112(5): 740-751.

18. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990; 43(6): 543-549.