



# A Stepwise Algorithmic Approach and External Validation Study for Noninvasive Prediction of Advanced Fibrosis in Nonalcoholic Fatty Liver Disease

Heather Mary-Kathleen Kosick<sup>1</sup> · Aline Keyrouz<sup>2</sup> · Oyedele Adeyi<sup>3</sup> · Giada Sebastiani<sup>2</sup> · Keyur Patel<sup>1</sup>

Received: 18 June 2020 / Accepted: 24 November 2020 / Published online: 3 January 2021  
© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

**Background and Aims** Advanced F3–4 fibrosis predicts liver-related mortality in nonalcoholic fatty liver disease (NAFLD). Noninvasive tests, designed to rule in/out advanced fibrosis, are limited by indeterminates, necessitating biopsy. We aimed to determine whether stepwise combinations of noninvasive serum-based tests and elastography (VCTE) could predict F3–4, reduce indeterminates, and decrease liver biopsies.

**Methods and Results** Five hundred forty-one biopsy-proven NAFLD cases were identified between 2010 and 2018 from two Canadian centers. Characteristics of training ( $n=407$ )/validation ( $n=134$ ) cohorts included: males 54%/59%; mean age 48.5/52.5 years; mean body mass index 32.3/33.6 kg/m<sup>2</sup>; diabetes mellitus 30%/34%; and F3–4 48%/43%. For training/validation cohorts, area under the receiver operating curve (AUROC) for FIB-4, AST-platelet ratio index (APRI), NAFLD fibrosis score (NFS), BARD score, and AST/ALT ratio ranged from 0.70 to 0.83/0.68 to 0.81, with indeterminates 25–39%/34–45%, for F3–4. In the training cohort, parallel FIB-4 + NFS had good accuracy (AUROC = 0.81) but was limited by 38% indeterminates and 16% misclassified. Sequential FIB-4 → NFS reduced indeterminates to 10%, and FIB-4 → VCTE to 0%, misclassified 20–22%, while maintaining high specificity (0.88–0.92) and accuracy (AUROC 0.75–0.78) for combined cohorts. Liver biopsy could have been avoided in 27–29% of patients using sequential algorithms.

**Conclusions** Sequential FIB-4 → NFS/VCTE predicts F3–4 with high specificity and good accuracy, while reducing indeterminates and need for biopsy. Parallel algorithms are limited by high indeterminates. Sequential FIB-4 → NFS had similar accuracy to VCTE-containing algorithms. Validation in low-prevalence cohorts may allow for potential use in community or resource-limited areas for risk stratification.

**Keywords** Nonalcoholic fatty liver disease · Liver fibrosis · Liver biopsy · Noninvasive

## Introduction

Nonalcoholic fatty liver disease (NAFLD) is a silent epidemic with over 12 million cases diagnosed annually in the USA alone [1]. Globally, NAFLD is thought to affect

**Supplementary Information** The online version of this article contains supplementary material available at <https://doi.org/10.1007/s10620-020-06748-8>.

✉ Heather Mary-Kathleen Kosick  
heather.evans@medportal.ca

Aline Keyrouz  
keyrouzaline@gmail.com

Oyedele Adeyi  
adeyio@umn.edu

Giada Sebastiani  
giada.sebastiani@mcgill.ca

Keyur Patel  
keyur.patel@duke.edu

<sup>1</sup> Division of Gastroenterology, University Health Network Toronto, Toronto General Hospital, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada

<sup>2</sup> Division of Gastroenterology and Hepatology, McGill University Health Center, 1001 Decarie Blvd, Montreal, QC H4A 3J1, Canada

<sup>3</sup> Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA

one in four individuals and parallels the rise in obesity and metabolic syndrome [2, 3]. NAFLD is a leading cause of end-stage liver disease, resulting in liver-related complications including hepatocellular carcinoma (HCC) and death [4, 5]. NAFLD is now one of the leading causes of liver transplant [6].

NAFLD encompasses both nonalcoholic fatty liver (NAFL) and nonalcoholic steatohepatitis (NASH) [4, 7]. NAFL is defined by the presence of steatosis ( $\geq 5\%$ ) in the absence of inflammation or hepatocellular injury (hepatocyte ballooning on histology), while NASH refers to steatosis ( $\geq 5\%$ ) occurring in the presence of both [8]. NAFL and NASH exist on a spectrum, with steatohepatitis signifying an increased risk of progressive hepatic injury [9]. The natural history of NAFLD is variable with  $\sim 20\%$  of patients progressing to advanced fibrosis, which is the main predictor of clinical outcomes [10–12]. However, most patients remain asymptomatic, and even advanced NAFLD may be easily overlooked in clinical practice [4].

Liver biopsy is the established standard for diagnosis of NASH and fibrosis, but is an invasive procedure with associated risk, along with diagnostic limitations due to sampling error and interpretation. Several blood-based and imaging elastography noninvasive tests (NIT) have been developed to predict advanced NAFLD fibrosis and reduce the need for biopsy [13, 14]. These include simple, indirect serum marker algorithms such as NAFLD fibrosis score (NFS), FIB-4, aspartate aminotransferase (AST)-platelet ratio index (APRI), and BARD (body mass index (BMI), AST/alanine aminotransferase (ALT) ratio, diabetes), that are routinely available and have comparable diagnostic accuracy for advanced fibrosis [15]. A significant limitation of simple NITs is that at least 30% of patients are classified as “indeterminate,” with index scores that are between cutoffs to “rule in” or “rule out” advanced fibrosis [16–18]. Diagnostic accuracy is further reduced by misclassified cases, and as a result, these simple tests are not yet able to replace biopsy for diagnosis of advanced fibrosis in NAFLD. Other validated proprietary serum marker tests such as FibroTest (BioPredictive, Paris, France; FibroSURE, Labcorp, Burlington, NC), Enhanced Liver Fibrosis (ELF) score (Siemens Healthcare, Erlangen, Germany), or FibroMeter<sup>VCTE</sup> (Echosens, Paris, France) may increase diagnostic accuracy as compared to simple markers, but are not as readily available [13]. Imaging-based NITs such as vibration-controlled transient elastography (VCTE, Echosens, Paris, France), shear-wave elastography, or magnetic resonance (MR)-based techniques are being increasingly adopted in diagnostic algorithms for advanced fibrosis in NAFLD, but limited by variable optimal liver stiffness thresholds across different cohorts and are not as readily available or as cost-effective as simple markers for fibrosis assessment [19].

Sequential combinations of blood-based markers and VCTE were initially developed and validated in chronic hepatitis C, to reduce the need for liver biopsy to diagnose significant fibrosis (F2–4) prior to interferon-based therapy [20, 21]. Recently, the sequential use of NITs such as FIB-4 and ELF or VCTE has been evaluated for risk stratification for NAFLD advanced fibrosis and to improve referral pathways from primary care [22, 23]. Sequential elastography with blood-based markers can improve diagnostic accuracy for advanced fibrosis and reduce indeterminates [24, 25]. However, sequential NIT approaches that still require validation, optimal combinations, or thresholds have not been established, and proprietary tests such as ELF are not yet available for routine use, outside of a research setting, in North America. The European Association for the Study of the Liver (EASL) Clinical Practice guidelines for NAFLD indicate that although combined NIT approaches may improve accuracy, there is no consensus on thresholds or strategies for use in avoiding biopsies for advanced fibrosis [26].

Our aims were to identify if stepwise combination of simple NITs (1) reduces indeterminate rates for predicting advanced NAFLD fibrosis (F3–4) compared to individual NIT, (2) identifies the optimal sequential combination of simple NITs to reduce the need for biopsy for the diagnosis of advanced fibrosis (F3–4), and (3) validates the performance of sequential NITs in an independent external cohort.

## Methods

### Study Design and Population

This was a retrospective observational cohort study of patients with NAFLD from two Canadian tertiary-care centers: University Health Network, Toronto (Training cohort) and McGill University Health Center, Montreal (Validation cohort).

Patients included were  $\geq 18$  years of age and had undergone a liver biopsy to assess for NASH between January 1, 2010 and July 1, 2018. Patients were excluded if they had an alternate cause of chronic liver disease or steatosis at biopsy, including viral hepatitis, significant alcohol use, Wilson’s disease, autoimmune hepatitis,  $\alpha 1$ -antitrypsin deficiency, hemochromatosis, and long-term use of steatogenic medications (including amiodarone, tamoxifen, methotrexate, tetracycline, and glucocorticoids). Further exclusion criteria were presence of malignancy (except HCC) within the past 5 years, immunosuppressive medications within the past 3 years, human immunodeficiency virus (HIV) co-infection, and inadequate liver biopsy ( $< 10$  mm or based on pathology assessment). Significant alcohol use was defined as  $> 14$  units weekly (or  $> 2$  units daily) for women and  $> 21$

units weekly (or > 3 units daily) for men. Anthropometric data and bloodwork within  $\pm 6$  months of liver biopsy were included, and VCTE if a valid liver stiffness measure (LSM) was available within  $\pm 12$  months of liver biopsy.

### Histologic Analysis and Data Acquisition

Liver biopsies were assessed by tertiary center expert histopathologists at each site. In addition, biopsy report summaries outlining the pathologists' clinical impression were verified by an independent clinical researcher, and those identifying an alternate cause of fibrosis or steatosis as per study exclusion criteria were omitted. NASH and fibrosis were scored using the NASH Clinical Research Network (NASH-CRN) Scoring System [8]. The presence or absence of "advanced fibrosis," defined as CRN score 3–4 (bridging fibrosis or cirrhosis) was recorded for each biopsy.

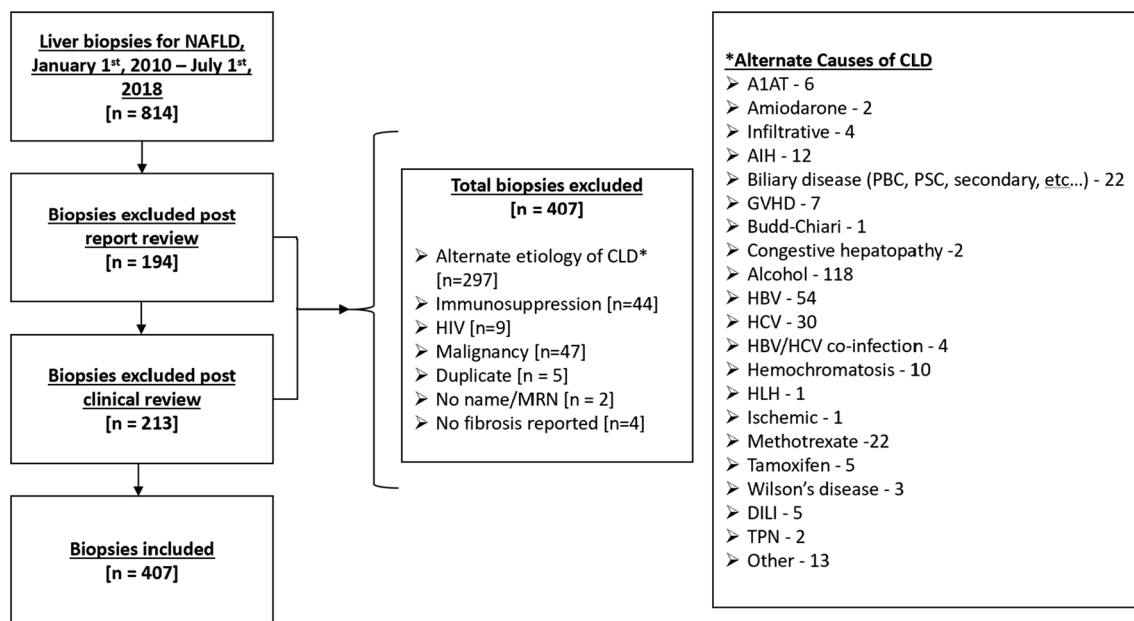
Clinical data were acquired via the local electronic medical record. Baseline data included demographics (age at time of biopsy, gender, comorbidities, medications), anthropometric measurements (height, weight, BMI), biochemical tests (liver enzymes: AST, ALT, alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT); liver function: bilirubin, International Normalized Ratio (INR), albumin, glucose; complete blood count (CBC), electrolytes, creatinine, lipid profile, glycated hemoglobin (HBA1c), ferritin), and VCTE reports. Comorbidities i.e., hypertension, diabetes, dyslipidemia, etc., were recorded based on physician

reporting of these conditions within the medical record. A total of 814 liver biopsies were obtained to assess for NAFLD in the Toronto cohort between January 1, 2010 and July 1, 2018. Following initial review of biopsy reports, 196 biopsies were excluded based on stated inclusion and exclusion criteria. The remaining 620 biopsies underwent clinical review, of which 213 were subsequently excluded. The remaining 407 biopsies comprised the training cohort in this study (Fig. 1).

Serum-based NIT of NAFLD fibrosis, including NFS, FIB-4, BARD, APRI, and AST/ALT ratio, were calculated for all included patients using published formulas [27–30]. Validated thresholds predicting advanced F3–4 fibrosis used for the purposes of this study were NFS  $\geq 0.676$ , FIB-4  $> 2.67$  [31], APRI  $> 1.5$  [32], and AST/ALT ratio  $\geq 0.8$ . Cutoffs predicting F0–2 or the absence of advanced fibrosis were NFS  $< 1.455$ , FIB-4  $< 1.3$ , BARD  $< 2$ , APRI  $< 0.5$ , and AST/ALT ratio  $< 0.8$ . Valid LSM  $\leq 8.4$  kPa, using the "M" or "XL" probe, was selected to rule out F3–4 fibrosis, with 8.4 kPa representing the Youden associated LSM cutoff for VCTE in our training cohort.

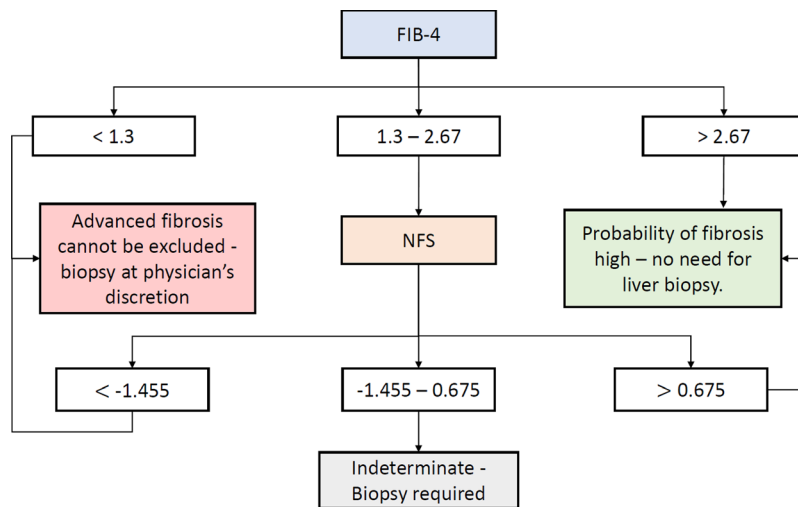
### Sequential Algorithms for Prediction of NAFLD Fibrosis

Three algorithms were constructed in the training cohort using a stepwise combination of NIT: (1) FIB-4  $\rightarrow$  NFS, (2) FIB-4  $\rightarrow$  VCTE, (3) FIB-4  $\rightarrow$  NFS  $\rightarrow$  VCTE (Fig. 2).

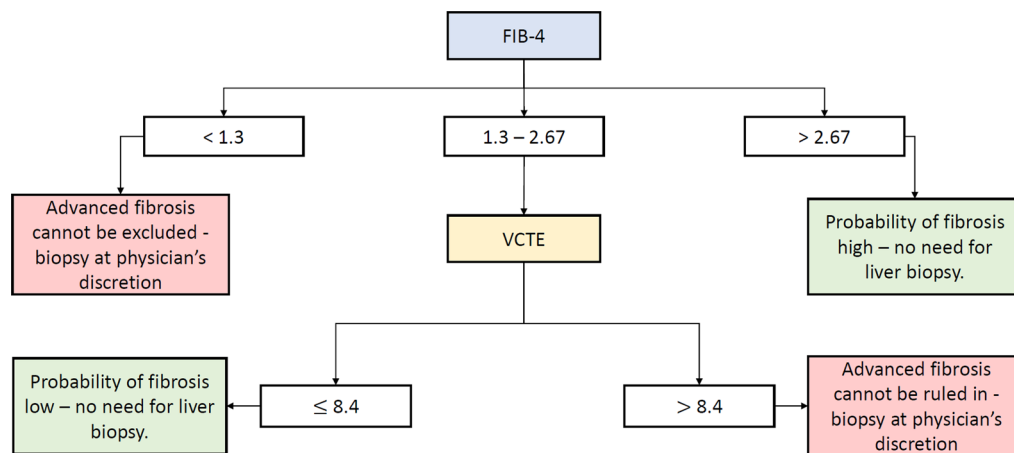


**Fig. 1** Training cohort patient selection and biopsy exclusions. NAFLD nonalcoholic fatty liver disease, CLD chronic liver disease, HIV human immunodeficiency virus, MRN medical record number, A1AT alpha-1 antitrypsin deficiency, AIH autoimmune hepatitis,

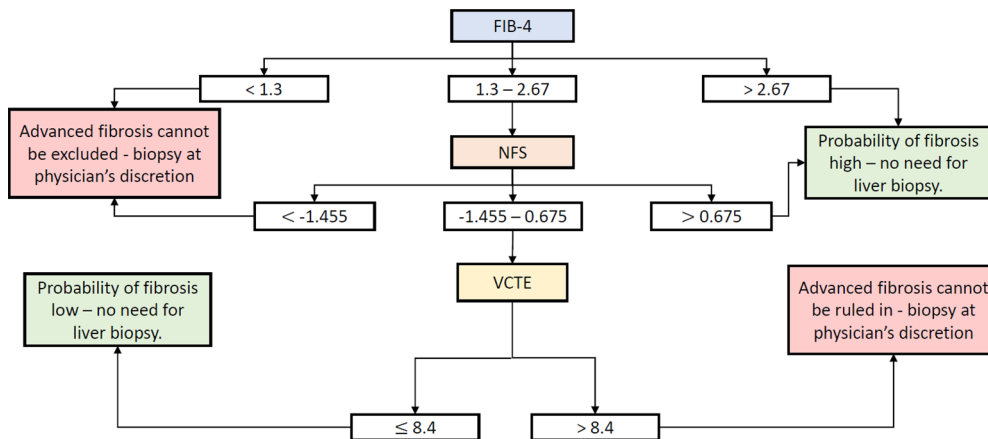
PBC primary biliary cholangitis, PSC primary sclerosing cholangitis, GVHD graft versus host disease, HBV hepatitis B virus, HCV hepatitis C virus, HLH hemophagocytic lymphohistiocytosis, DILI drug-induced liver injury, TPN total parenteral nutrition



**Algorithm 1** – FIB-4 followed by NAFLD fibrosis score (NFS). Biopsy was not required (green boxes) if the following conditions were met: 1) Sensitivity for individual non-invasive test (NIT) > 80%, in order to 'rule out' advanced fibrosis, 2) Specificity >80% for individual NIT, i.e. FIB-4/NFS, in order to 'rule in' advanced fibrosis, and 3) Type II error rate < 20% for any individual NIT. If these criteria were not met, biopsy could be performed at physician's discretion. Biopsies were required for those with indeterminate scores.



**Algorithm 2** – FIB-4 followed by vibration-controlled transient elastography (VCTE). Biopsy was not required (green boxes) if the following conditions were met: 1) Sensitivity for individual non-invasive test (NIT), i.e. VCTE, > 80%, in order to 'rule out' advanced fibrosis, 2) Specificity >80% for individual NIT, i.e. FIB-4, in order to 'rule in' advanced fibrosis, and 3) Type II error rate < 20% for any individual NIT. If these criteria were not met, biopsy could be performed at physician's discretion.



**Algorithm 3** – FIB-4 followed by NAFLD fibrosis score (NFS), followed by vibration-controlled transient elastography (VCTE). Biopsy was not required (green boxes) if the following conditions were met: 1) Sensitivity for individual non-invasive test (NIT), i.e. VCTE, > 80%, in order to 'rule out' advanced fibrosis, 2) Specificity >80% for individual NIT, i.e. FIB-4/NFS, in order to 'rule in' advanced fibrosis, and 3) Type II error rate < 20% for any individual NIT. If these criteria were not met, biopsy could be performed at physician's discretion.

**Fig. 2** Sequential algorithms for advanced fibrosis

Standardized cutoffs were used as described above. The second test was applied to patients with “intermediate” range index scores (FIB-4 = 1.3–2.67 for algorithms 1 and 2, and NFS = – 1.455–0.675 for algorithm 3). Patient scores falling above/below cutoffs were labeled as having “high” or “low” likelihood of F3–4 fibrosis. Misclassification rates were reported at each stage of the decision tree. For an algorithm to avoid the need for biopsy, the following criteria were required: (1) sensitivity  $\geq 80\%$ , (2) specificity  $\geq 80\%$ , and (3) false negative rate  $< 20\%$ .

## Statistical Analysis

Student’s *t* test was used to compare quantitative data. Chi-squared test was used for comparison of frequency data. The Mann–Whitney *U* test was used to compare ordinal data. All tests are two-tailed, with  $p < 0.05$  set as the level of statistical significance. All tests assumed equal variance unless standard deviations between compared groups differed substantially. Qualitative variables are expressed as mean  $\pm$  SD with 95% confidence interval, assuming a normal distribution, based on the central limit theorem. Area under the receiver operating curve (AUROC), as described by DeLong et al. [33], was used to determine accuracy of noninvasive tests and sequential algorithms. To better account for the non-binary nature of biopsy as a reference standard, and spectrum effect across an ordinal scale, the weighted Obuchowski measure was also calculated for each NIT algorithm [34]. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios were determined for all sequential algorithms, following removal of indeterminate range index scores. For individual NIT, performance characteristics are reported at the maximal Youden Index. For sequential algorithms, given the dichotomous nature of these results, predictive values were calculated using standardized statistical formulae via production of 2 by 2 tables.

Statistical analyses were performed using Medcalc (*MedCalc Software Version 19.0.7, Ostend, Belgium*). “R” statistical software (*R Development Core Team, 2008*) was used for “non-binary ROC analysis,” to calculate the Obuchowski measure, using the “nonbinROC” R package created by Paul Nguyen (DOI: 10.1.1.215.7235).

## Results

### Training Cohort

#### Demographics

Baseline clinical characteristics for the 407 patients from the training cohort are shown in Table 1. The prevalence

of advanced fibrosis was 48% ( $n = 196/407$ ), and compared to F0–F2, patients with F3–4 were more likely to be older (53.9 vs. 43.4 years), female (57% vs. 36%), with higher BMI (33.4 vs. 31.0), and receiving treatment for coexisting metabolic disease (hypertension, type 2 diabetes, dyslipidemia). There were no differences between the F0–2 and F3–4 cohorts for reported rates of smoking (30.3% vs. 25.5%,  $p = 0.28$ ).

Index scores for simple markers NFS, FIB-4, APRI, BARD, AST/ALT were all significantly higher in F3–4 patients. However, mean NFS for F3–4 was lower than the threshold of 0.676. As expected, patients with F3–4 had significantly higher LSM compared to stage F0–2 patients ( $6.7 \pm 2.3$  kPa vs.  $20.6 \pm 16.1$  kPa,  $p < 0.0001$ ).

### Single Noninvasive Test Performance for F3–4

The AUROC values for predicting F3–4 using NIT in training cohort ranged from 0.70 to 0.92 (Obuchowski 0.66–0.86) (Table 2). VCTE was only available in a subset of patients ( $n = 80$ ). A LSM threshold of 8.4 kPa, the Youden cutoff for our training cohort, had the highest AUROC of 0.92 with sensitivity 0.86 and specificity 0.87. AUROCs for NFS, FIB-4, and APRI were determined following exclusion of 25–39% of patients with Indeterminate results (Table 2). Of the simple biomarker-based tests, FIB-4 had high specificity of 0.94 but lower sensitivity of 0.66 and AUROC 0.83, as compared to AUROC 0.78 and 0.80 for NFS and APRI, respectively. There were no significant differences between AUROCs for FIB-4, NFS, APRI, or VCTE. Misclassification rates for simple biomarkers ranged from 17 to 35% (FIB-4 17%, NFS 23%, APRI 23%, BARD 35%, AST/ALT 35%). VCTE had a lower misclassification rate of 14% (Table 2).

### Sequential Noninvasive Algorithms for Advanced Fibrosis

We next evaluated three sequential algorithms of NIT for the prediction of advanced F3–4 NAFLD fibrosis, with the aim of reducing the proportion of indeterminate and misclassified patients. FIB-4, NFS, and VCTE were selected for their favorable performance as single NITs in the training cohort and availability in clinical practice. Sequential algorithms were constructed as (1) FIB-4  $\rightarrow$  NFS, (2) FIB-4  $\rightarrow$  VCTE, (3) FIB-4  $\rightarrow$  NFS  $\rightarrow$  VCTE (Fig. 2).

Algorithm 1 (FIB-4  $\rightarrow$  NFS) in patients with FIB-4 score 1.3–2.67 resulted in AUROC of 0.77, with high specificity of 0.95 and lower sensitivity of 0.60. Compared to single NITs, this approach reduced indeterminates to 26/318 (8%) patients, with an overall misclassification rate of 20%.

Algorithm 2 (FIB-4  $\rightarrow$  VCTE) resulted in an AUROC for F3–4 of 0.81, with a sensitivity of 0.68 and a specificity of 0.95. There were no indeterminate scores, and the misclassification rate was 17%.

**Table 1** Baseline clinical characteristics in training cohort

Patient demographics	Total (n=407)	F0–2 (n=211)	F3–4 (n=196)	p
Fibrosis stage		F0 (n=76) F1 (n=71) F2 (n=64)	F3 (n=74) F4 (n=122)	–
% Males (n)	54% (221)	64% (136)	43% (85)	< 0.0001*
Age (mean ± SD) years	48.5 ± 13.1	43.4 ± 12.3	53.9 ± 11.6	< 0.0001*
% Hypertension (n)	33%	22.3% (47)	45.4% (89)	< 0.0001*
% Diabetes (n)	30%	15.6% (33)	46.4% (91)	< 0.0001*
% Dyslipidemia (n)	34%	26.5% (56)	41.8% (82)	0.0011*
% Chronic kidney disease (n)	2% (8)	0% (0)	4.1% (8)	0.0030*
% Smoking (reported) (n)	28% (114)	30.3% (64)	25.5% (50)	0.2817
Antihypertensives (n)	32% (130)	21.8% (46)	42.9% (84)	< 0.0001*
Non-insulin hypoglycemic (n)	23.1% (94)	14.2% (30)	32.7% (64)	< 0.0001*
Insulin (n)	8% (33)	2.4% (5)	14.3% (28)	< 0.0001*
Statins (n)	21% (84)	12.3% (26)	29.6% (58)	< 0.0001*
BMI (mean ± SD) (kg/m <sup>2</sup> )	32.3 ± 6.9	31.0 ± 6.5	33.4 ± 7.1	0.0131*
AST (mean ± SD) [U/L]	80.2 ± 192.6	51.6 ± 55.1	111.9 ± 270.6	0.0045*
ALT (mean ± SD) [U/L]	98.4 ± 113.6	83.6 ± 71.6	115.2 ± 145.6	0.0102*
NAS (median, IQR)	4 (2–5)	3 (1–4)	4 (4–5)	< 0.0001*, <sup>†</sup>
NFS (n=198)	– 1.35 ± 2.48	– 2.51 ± 1.71	– 0.13 ± 2.59	< 0.0001*
FIB-4 (n=363)	2.58 ± 5.96	1.17 ± 1.50	4.20 ± 8.26	< 0.0001*
BARD (median, IQR) (n=218)	2 (1–3)	1 (1–2)	2 (1–3)	< 0.0001* <sup>†</sup>
APRI (n=363)	1.36 ± 4.98	0.60 ± 0.77	2.23 ± 7.12	0.0033*
AST/ALT ratio (n=365)	0.85 ± 0.57	0.70 ± 0.39	1.03 ± 0.68	< 0.0001*
VCTE [kPa] (n=83)	15.4 ± 14.5	6.7 ± 2.3	20.6 ± 16.1	< 0.0001*

BMI body mass index, AST aspartate aminotransferase, ALT alanine aminotransferase, NAS NAFLD activity score, NFS NAFLD fibrosis score, APRI AST-to-platelet ratio index, VCTE vibration-controlled transient elastography, SD standard deviation, IQR interquartile range

n number of patients

\*p < 0.05; <sup>†</sup>p calculated using Mann–Whitney test

**Table 2** Noninvasive tests for the prediction of stage F3–F4 fibrosis in training cohort

Test	PPV	NPV	Sens	Spec	LR <sup>+</sup>	LR <sup>–</sup>	AUROC (95% CI)	Indeterminates	Misclassified	Ob (SE)
NFS (n=198)	0.85	0.74	0.64	0.90	6.38	0.40	0.78 (0.71–0.85)	25% (n=49)	23% (n=34) (FN=30, FP=4)	0.71 (0.03)
BARD (n=218)	0.64	0.67	0.71	0.59	1.74	0.49	0.70 (0.63–0.76)	–	35% (76/218) (FN=32, FP=44)	0.66 (0.02)
FIB-4 (n=363)	0.89	0.81	0.66	0.94	11.67	0.36	0.83 (0.78–0.87)	27% (n=98)	17% (n=46) (FN=38, FP=8)	0.74 (0.02)
APRI (n=363)	0.73	0.79	0.62	0.86	4.30	0.44	0.80 (0.74–0.85)	39% (n=140)	23% (n=52) (FN=39, FP=13)	0.73 (0.02)
AST/ALT ratio (n=365)	0.65	0.73	0.73	0.65	2.06	0.42	0.71 (0.66–0.76)	–	35% (n=127) (FN=76, FP=51)	0.66 (0.02)
VCTE (n=80)	0.92	0.79	0.86	0.87	6.45	0.16	0.92 (0.83–0.97)	–	14% (n=11) (FN=7, FP=4)	0.86 (0.03)

Sensitivity/specificity are reported at the Youden cutoff. Misclassification rates are reported as the percentage of patients with discordant NIT and biopsy results after excluding indeterminate results

PPV positive predictive value, NPV negative predictive value, LR<sup>+</sup> positive likelihood ratio, LR<sup>–</sup> negative likelihood ratio, AUROC area under the receiver operating characteristic, Sens sensitivity, Spec specificity, FN false negative, FP false positive, CI confidence interval, Ob Obuchowski measure, SE standard error, NFS NAFLD fibrosis score, APRI AST-to-platelet ratio index, VCTE vibration-controlled transient elastography

Algorithm 3 (FIB-4 → NFS → VCTE) in patients who were indeterminate by FIB-4 analysis, underwent NFS for further stratification, followed by VCTE for patients still classified as indeterminate by NFS. resulted in the AUROC for F3–4 as 0.78, with a sensitivity of 0.62 and a specificity of 0.94. There were no patients left with indeterminate scores, and misclassification rate was 20% (Table 3).

### Parallel Combination of Noninvasive Tests for Advanced Fibrosis

We also compared performance characteristics of FIB-4 in parallel to NFS (FIB-4 + NFS) for advanced fibrosis. Patients with both FIB-4 and NFS meeting thresholds for presence or absence of advanced F3–4 fibrosis were classified as “likely” or “unlikely” to have advanced fibrosis. If either FIB-4 or NFS was not in agreement, patients were labeled as indeterminate. Overall concordance of FIB-4 and NFS was 73%. Among patients with a FIB-4 score indicating F0–2, 85% ( $n = 85/99$ ) also had a concordant NFS score. Among patients with an indeterminate FIB-4 score, 49% (26/53) had a concordant NFS score. For patients with a FIB-4 score indicating advanced fibrosis, 76% (35/46) had a concordant NFS score. Overall, performance characteristics of this parallel FIB-4 + NFS ( $n = 198$ ) indicated AUROC of 0.81, with sensitivity = 0.67, specificity = 0.96, PPV = 0.91, NPV = 0.81. Use of a parallel analysis thus resulted in comparable diagnostic performance to algorithm 1, with a marginally higher sensitivity, but was limited by a higher indeterminate rate (78/198, 38%). Parallel FIB-4 + NFS did lower overall misclassification rates, with a rate of 16% (19/120 patients) versus 20% for the sequential FIB-4 → NFS algorithm. Overall, 54% were either misclassified or indeterminate for the parallel combination, compared to 28% for this sequential algorithm.

## Validation Cohort

### Demographics

Baseline clinical characteristics for the  $n = 134$  patients from the external validation cohort as compared to the original cohort are shown in Table 4. The prevalence of advanced fibrosis in the validation cohort was 43% (57/134) and comparable to 48% for the training cohort ( $p = 0.31$ ). Patients in the validation cohort were older (52.5 vs. 48.5 years;  $p = 0.0016$ ), with lower mean AST (49.6 vs. 80.2 IU/L;  $p = 0.0036$ ) and ALT (68.4 vs. 98.4 IU/L;  $p < 0.0001$ ), but otherwise comparable for biological sex, BMI, and proportion with diabetes mellitus. Despite a similar prevalence of F3–4, mean index scores for FIB-4, NFS, and APRI were lower in the validation cohort compared to the training cohort, but there were no differences in LSM (Table 4).

### Single Noninvasive Test Performance

Results of blood-based NIT were comparable between training and validation cohorts, with AUROC 0.68–0.81 for F3–4 (Supplementary Table 1). VCTE did not perform as well, with a lower AUROC = 0.66. NIT performed with high specificity ( $> 0.8$ ) for all scores except FIB-4, which performed with higher sensitivity (0.88). Rates of indeterminate scores for NFS, FIB-4, and APRI, were 40%, 34%, and 45%, respectively. These were higher as compared to the training cohort. Misclassification rates ranged from 23 (AST/ALT) to 44% (VCTE). Obuchowski measure ranged from 0.63 (VCTE) to 0.74 (FIB-4).

**Table 3** Sequential algorithms for the prediction of F3–4 fibrosis in training cohort

Algorithm	PPV	NPV	Sens.	Spec.	LR <sup>+</sup>	LR <sup>-</sup>	AUROC (95% CI)	Indeterminates	Misclassified	Ob (SE)
1 FIB-4 → NFS ( $n = 318$ )	0.89	0.76	0.60	0.95	11.13	0.42	0.77 (0.72–0.82)	8% ( $n = 26$ )	20% ( $n = 59$ ) (FN = 50, FP = 9)	0.71 (0.01)
2 FIB-4 → VCTE ( $n = 292$ )	0.91	0.79	0.68	0.95	12.41	0.34	0.81 (0.76–0.86)	0%	17% ( $n = 50$ ) (FN = 41, FP = 9)	0.73 (0.01)
3 FIB-4 → NFS → VCTE ( $n = 302$ )	0.89	0.76	0.62	0.94	10.56	0.40	0.78 (0.73–0.83)	0%	20% ( $n = 60$ ) FN = 50, FP = 10	0.71 (0.01)

Sensitivity/specificity was calculated using a two-by-two table. Misclassification rates are reported as the percentage of patients with discordant NIT and biopsy results after excluding indeterminate results

PPV positive predictive value, NPV negative predictive value, LR<sup>+</sup> positive likelihood ratio, LR<sup>-</sup> negative likelihood ratio, AUROC area under the receiver operating characteristic, Sens sensitivity, Spec specificity, FN false negative, FP false positive, CI confidence interval, Ob Obuchowski measure, SE standard error

**Table 4** Baseline clinical characteristics for training and validation cohorts

Patient demographics	Toronto	McGill	<i>p</i>
Number of patients	407	134	–
F0–2 ( <i>n</i> )	52% (211)	57% (77)	0.3148
F3–4 ( <i>n</i> )	48% (196)	43% (57)	0.3148
% Males ( <i>n</i> )	54% (221)	59% (79)	0.3131
Age* (mean ± SD)	48.5 ± 13.1	52.5 ± 11.5	0.0016
% Diabetes ( <i>n</i> )	30% (124)	34% (46)	0.3856
BMI (mean ± SD) [kg/m <sup>2</sup> ]	32.3 ± 6.9	33.6 ± 5.8	0.0800
AST (mean ± SD) [U/L]	80.2 ± 192.6	49.6 ± 29.2	0.0036*
ALT (mean ± SD) [U/L]	98.4 ± 113.6	68.4 ± 43.3	< 0.0001*
Platelets	220 ± 94	203 ± 73	0.0716
NFS ( <i>n</i> )	– 1.32 ± 2.50 (198)	– 0.77 ± 1.83 (80)	0.0156*
FIB-4 ( <i>n</i> )	2.60 ± 5.96 (363)	1.94 ± 1.60 (107)	< 0.0001*
BARD (median, IQR) ( <i>n</i> )	2 (1–3) (218)	2 (1–3) (90)	0.2282 <sup>†</sup>
APRI ( <i>n</i> )	1.37 ± 4.98 (363)	0.73 ± 0.52 (108)	< 0.0001*
AST/ALT ratio ( <i>n</i> )	0.86 ± 0.57 (365)	0.78 ± 0.29 (110)	0.2156
VCTE [kPa] ( <i>n</i> )	15.4 ± 14.5 (83)	15.0 ± 12.2 (89)	0.8447

BMI body mass index, AST aspartate aminotransferase, ALT alanine aminotransferase, NAS NAFLD activity score, NFS NAFLD fibrosis score, APRI AST-to-Platelet Ratio Index, VCTE vibration-controlled transient elastography, SD standard deviation, IQR interquartile range

*n* number of patients

\**p* < 0.05; <sup>†</sup>*p* calculated using Mann–Whitney test

**Sequential Noninvasive Tests for Advanced Fibrosis in Validation Cohort**

Sequential algorithm AUROCs were 0.67–0.70 (Obuchowski 0.65) in the validation cohort. Overall performance characteristics were lower compared to the training cohort (Table 5). Algorithm 1 (FIB-4 → NFS) performed with highest specificity for identifying F3–4 fibrosis.

Algorithm 1 (FIB-4 → NFS) resulted in modest AUROC of 0.67, with specificity of 0.82 but low sensitivity of 0.56. Compared to single NITs, this approach reduced indeterminates to 15% with an overall misclassification rate of 30%.

Algorithm 2 (FIB-4 → VCTE) resulted in AUROC for F3–4 of 0.69. There were no indeterminate scores, but as expected from the sensitivity and specificity of 0.67–0.71, the misclassification rate was 31%.

Algorithm 3 (FIB-4 → NFS → VCTE) resulted in an AUROC for F3–4 was 0.70, with a sensitivity of 0.62 and a specificity of 0.78. There were no indeterminate scores, and the misclassification rate was 29% (Table 5).

**Table 5** Sequential algorithms for the prediction of F3–4 fibrosis, external validation cohort

Algorithm	PPV	NPV	Sens	Spec	LR <sup>+</sup>	LR <sup>–</sup>	AUROC (95% CI)	Indeterminates	Misclassified	Ob (SE)
1 FIB-4 → NFS ( <i>n</i> = 98)	0.68	0.73	0.56	0.82	3.04	0.54	0.67 (0.58–0.79)	15% (15)	30% ( <i>n</i> = 24) (FN = 15, FP = 10)	0.65 (0.03)
2 FIB-4 → VCTE ( <i>n</i> = 95)	0.59	0.78	0.67	0.71	2.31	0.47	0.69 (0.59–0.78)	0% (0)	31% ( <i>n</i> = 29) (FN = 12, FP = 17)	0.65 (0.03)
3 FIB-4 → NFS → VCTE ( <i>n</i> = 93)	0.67	0.74	0.62	0.78	2.77	0.49	0.70 (0.59–0.79)	0% (0)	29% ( <i>n</i> = 27) (FN = 15, FP = 12)	0.65 (0.03)

Sensitivity/specificity was calculated using a two-by-two table. Misclassification rates are reported as the percentage of patients with discordant NIT and biopsy results after excluding indeterminate results

PPV positive predictive value, NPV negative predictive value, LR<sup>+</sup> positive likelihood ratio, LR<sup>–</sup> negative likelihood ratio, AUROC area under the receiver operating characteristic, Sens sensitivity, Spec specificity, FN false negative, FP false positive, CI confidence interval, Ob Obuchowski measure, SE standard error



## Modeling for Biopsy Avoidance in a Combined Cohort

Based on our data for the combined cohort (Supplementary Tables 2 and 3), we next determined whether a single test or algorithm could reduce the need for liver biopsy diagnosis of advanced fibrosis in the following scenario: (1) negative score (below cutoff) for a test with sensitivity > 80% and (2) positive score (above cutoff) for a test with specificity > 80%, with a type II error rate of < 20%. At a prevalence of 47% (253/541) for advanced fibrosis, liver biopsy could have been avoided in 27% (112/416), 29% (112/387), and 29% (116/395) of patients using sequential algorithms 1, 2, and 3, respectively. This compares to 22% (61/278), 21% ([98/469), and 35% (59/168) for NFS, FIB-4, and VCTE as single tests, respectively. Differences in biopsies saved were significant for Algorithm 1 versus FIB-4 ( $p=0.0363$ ), Algorithm 2 versus NFS ( $p=0.0427$ ), Algorithm 2 versus FIB-4 ( $p=0.0069$ ), Algorithm 3 versus FIB-4 ( $p=0.0066$ ), and Algorithm 3 versus NFS ( $p=0.0419$ ).

## Discussion

The identification of advanced fibrosis is important for risk stratification and clinical management in NAFLD. Simple marker algorithms, such as FIB-4 and NFS, are readily available for the noninvasive prediction of advanced fibrosis but are limited by indeterminate scores in a significant proportion of patients. Our study demonstrates the utility of combining simple NITs for the prediction of advanced fibrosis in NAFLD patients and further validates performance characteristics of reducing indeterminates using sequential algorithms in an independent cohort. Our data indicate that the use of sequential NITs can significantly reduce the number of “indeterminate” results compared to single tests, without compromising diagnostic performance, or increasing misclassification rates. Furthermore, by using sequential tests, liver biopsy for the diagnosis of advanced fibrosis could have been avoided in up to 29% of NAFLD patients in our cohort and 27% using sequential combinations of serum-based tests alone.

Individual NIT identified advanced fibrosis in our cohorts with good performance, and overall diagnostic accuracy and indeterminate rates were comparable to prior studies [17, 24, 25, 27, 31]. Both parallel and sequential simple blood marker NITs were evaluated in the training cohort. Compared to their sequential application (in algorithm 1), FIB-4 + NFS in parallel marginally reduced misclassification rates but led to indeterminate scores of 38%, that were higher than observed with individual tests. A prior multicenter study in 761 NAFLD patients indicated that paired VCTE with NFS or FIB-4 also reduced misclassification rates, but resulted in increased indeterminate rates in over

one-half of patients and thus reduced the overall accuracy for paired tests to diagnose F3–F4 to ~40% [24]. Data from a phase III clinical trial in NASH stage 3–4 patients indicated the simultaneous combination of FIB-4 or NFS with ELF or VCTE resulting in low misclassification rates of 4–8%, but unacceptable indeterminate classification in 64–77% of patients [36]. Based on findings from these larger cohort studies, we did not further evaluate paired VCTE + FIB-4 or NFS in our cohort. We also observed discordance rates of ~30% between FIB-4 and NFS which further limited this diagnostic approach in our cohort. Overall, these findings suggest that parallel combination of current NITs have limited utility for diagnosis of advanced fibrosis in NAFLD.

The strengths of our study include the combined cohort of > 500 patients with biopsy-proven NAFLD confirmed by expert histopathologists, and first to use sequential NIT diagnostic algorithms with an independent validation cohort in a North American NAFLD population. We selected NFS and FIB-4 as first-line tests for ease of use, cost-effectiveness, and access. Importantly, in our overall cohort, sequential FIB-4 → NFS allowed for reduction of indeterminates to 10% compared to 29% for individual tests, while maintaining specificity > 0.9. Overall misclassification rates for FIB-4 → NFS were 22% and comprised of false negatives in > 75% of these misclassified patients, highlighting the lower sensitivity of these simple serum marker tests in our cohorts. Our second sequential algorithm (FIB-4 → VCTE) had an overall specificity of 0.88 for advanced fibrosis, reduced indeterminates to 0%, with a misclassification rate of 20%, of which two-thirds of cases were false negatives (Supplementary Table 3). This compares favorably with recent studies that indicated indeterminate rates of 6–20% and misclassification rates of 16–20% using this sequential algorithm [24, 36]. These differences in indeterminate rates between our data and these recent studies relate to the use of a single lower LSM threshold of 8.4 kPa to “rule out” advanced fibrosis in our cohort, compared to either < 7.9 kPa and ≥ 9.6 kPa [24] or 9.9 kPa and 11.4 kPa [36] as lower and upper thresholds for “ruling out” or “ruling in” advanced fibrosis, respectively. Higher prevalence of F3–4 may also result in higher misclassification and indeterminate results with a sequential approach [37]. Algorithm 3 did not provide any incremental diagnostic accuracy and change in misclassifications compared to FIB-4 → VCTE. In the multicenter study of 761 NAFLD patients, the prevalence of advanced fibrosis was much lower at 30%, and in keeping with our findings, the best-performing single NITs for advanced fibrosis were FIB-4, NFS, and VCTE. Sequential algorithms incorporating VCTE and FIB or NFS in various combinations improved diagnostic accuracy of identifying advanced fibrosis to ~70% [24]. However, performance of a sequential FIB-4 → NFS algorithm was not reported,

which we feel is a unique strength of our study, owing to the simplicity and ease of access in performing these tests, and potentially overcoming some of the variability associated with obtaining reliable LSM assessments in non-tertiary clinical settings [23].

We did not explore performance of sequential tests based on new optimal thresholds derived from our combined dataset, as larger cohort studies have been unable to define single thresholds to optimize sensitivity and specificity for simple markers such as FIB-4 and NFS and thus remove these indeterminate scores [36]. Options to further reduce indeterminates and maintain accuracy with serum markers include the use of proprietary scores such as FibroTest or FibroMeter. However, these are associated with additional costs and were not included in this study. Other proprietary tests with validation for sequential use, including ELF<sup>TM</sup> and FibroMeter [25], are not yet routinely available or covered by provincial health plans in Canada, and VCTE is available for a non-reimbursed fee outside of tertiary centers.

Limitations of our study include that this was a retrospective study at tertiary centers with for-cause biopsy and a higher prevalence of F3–4 fibrosis compared to other studies [27, 31, 38, 39]. As such, further validation of these sequential algorithms is required in lower-prevalence populations. To account for the non-binary nature of our outcome measure and bias toward greater prevalence of F3–4, we evaluated the Obuchowski measure to account for differences in distribution of fibrosis scores in developing our sequential algorithms. However, only a few studies in NAFLD have accounted for the spectrum effect of variable fibrosis distribution in their cohorts during assessment of NIT diagnostic accuracy [25, 40, 41].

Due to increasing comorbidities, advanced age has been shown to reduce the specificity of FIB-4 [42]. Indeed, NIT were principally validated in adult NAFLD patients age < 65 years [16, 17, 31, 42]. We did not further control for this factor, as ~10% of patients were aged over 65 years in our combined cohorts, which reflects the general NAFLD population with respect to age. Ethnicity has also been shown to limit the performance of NIT. A recent study demonstrated the poor sensitivity of several NIT in South Asian patients [43]. This was also not controlled for in our study as ethnicity was inconsistently reported.

Highlighting the need for external validation in biomarker studies, differences in algorithm performance were noted between our training and validation cohorts despite having comparable rates of advanced fibrosis. At the selected thresholds, FIB-4 had a higher sensitivity and lower specificity in the external validation cohort. The higher sensitivity for FIB-4 in the validation cohort reflects the lower transaminase levels in this group. Despite lower AUCs for sequential tests in the external cohort, there was an important reduction in indeterminate rates for the FIB-4 → NFS algorithm by

17–19% and 21–25% for both training and validation cohorts, respectively.

VCTE was only available in a subset of our patients in this retrospective study. A cutoff of  $\leq 8.4$  kPa was selected based on the corresponding Youden index determined from our training cohort. This single low LSM cutoff served to optimize sensitivity and exclude patients without advanced fibrosis. Both M and XL probes were used for VCTE in this study. Prior studies have shown variable LSM diagnostic thresholds for F3–4 of 6.95–11.4 kPa using the M probe and 5.7–9.3 kPa for NAFLD F3–4 [44]. Our AUROC of 0.76 for VCTE in the combined cohort was lower than expected. VCTE performed with high sensitivity and specificity in our training cohort. Our validation cohort was limited by a higher false positive rate and lower specificity than expected, at the selected 8.4 kPa cutoff. This can be explained by the higher median LSM among F0–2 patients in the validation vs. training cohorts (10.8 kPa vs. 6.3 kPa, respectively). Further, differences may have occurred due to variability in operator technique or patient body habitus. Although both sites had very experienced operators and used a similar protocol of attempting measurement with the M probe, prior to use of the XL, if no reliable measurement could be obtained, further analysis of VCTE-specific parameters and reliability criteria would have been helpful to further elucidate observed differences between cohorts.

In this study, VCTE alone reduced need for liver biopsy at rates greater than sequential combinations of NIT using VCTE. This may relate to higher sensitivity for F3–4 for VCTE, but despite the differences in VCTE performance between our two cohorts, the use of VCTE as a sequential test improved specificity. In our region, VCTE is principally available at specialist centers and usually associated with non-reimbursed costs. Use of sequential FIB4 + NFS is a simple means of reducing need for liver biopsy, without compromising accuracy. Although our study did not assess cost-effectiveness, previous work has demonstrated that a sequential combination of serum-based NIT with VCTE have been shown to be more cost-effective than VCTE alone [45].

In future, given the low prevalence of advanced NAFLD in community practices and heterogeneity in disease-related outcomes [46, 47], a “triage” strategy based on combination NIT determined advanced fibrosis could reduce the number of referrals for secondary or invasive testing and reduce associated patient anxiety, discomfort, and healthcare costs [22, 48]. Although validation in a primary care setting could allow for implementation of sequential algorithms of NIT for diagnostic purposes, this may be difficult due to the need for liver biopsy as a comparative reference standard. Future work may instead focus on modeling strategies for screening using validated noninvasive tools such as MR elastography, allowing for assessment of NIT algorithms

performance at a lower population prevalence. Additional assessment of the utility of such sequential algorithms with age-specific thresholds, validation in multi-ethnic cohorts, incorporation of VCTE-controlled attenuation parameter in diagnostic algorithms to reduce false positives [49], and risk stratification for disease progression or clinical outcomes is still required.

In order for a noninvasive testing strategy to be implemented in primary care, multiple variables must be considered and optimized, which are outside the scope of this study. These include optimization of test accuracy, sensitivity, and specificity, and minimizing misclassified (false positive and negative) and indeterminate results. Optimal thresholds for combined tests have yet to be determined, and minimizing false negative results is important for screening populations at risk of advanced fibrosis. However, validation of these diagnostic approaches in lower prevalence cohorts is still required. Cost-effectiveness of a sequential algorithm strategy for diagnosis and management of advanced fibrosis must also be taken into account.

In summary, the use of sequential NITs was able to accurately predict F3–4 fibrosis and reduce indeterminates, thereby reducing need for liver biopsy, compared to serum NITs used alone or in parallel. A combination of FIB-4 followed by NFS performed well and was comparable to sequential tests incorporating VCTE. The ability of simple NIT sequential algorithms to accurately predict advanced fibrosis with high specificity improves the possibility of identifying patients without biopsy who may benefit from closer surveillance as well as potential candidates for future therapies.

**Acknowledgment** Study sponsors did not play any role in the study design, data collection or analysis, interpretation of data, in the writing of the report, or in the decision to submit the manuscript for publication.

**Funding** This study was supported by a research Grant from Gilead Sciences Canada. GS is supported by a Junior 1 and 2 Salary Award from FRQS (Nos. 27127 and 267806) and research salary from the Department of Medicine of McGill University.

## Compliance with Ethical Standards

**Conflict of interest** GS has acted as speaker for Merck, Gilead, Abbvie, Novonordisk, Novartis, served as an advisory board member for Merck, Novartis, Gilead and Intercept, and has received unrestricted research funding from Merck and Theratechnologies Inc. KP has the following declarations: (1) Gilead-Advisory Board, consulting, research grants; (2) Novartis-Advisory Board; (3) Intercept-Advisory Board.

**Ethics and consent** This study was performed according to the protocol and guidelines of Good Clinical Practice/ICH, based upon the principles outlined in the Declaration of Helsinki and local and national guidelines governing the conduct of clinical research studies. All patient data used for the purpose of this retrospective study were

de-identified and anonymized, and patient consent was not required. This study was approved by the Institutional Research Ethics Board at both the University Health Network and the McGill University Health Centre.

## References

1. Younossi ZM, Blissett D, Blissett R, et al. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology*. 2016;64:1577–1586.
2. Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64:73–84.
3. Araujo AR, Rosso N, Bedogni G, et al. Global epidemiology of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: what we need in the future. *Liver Int.* 2018;38:47–51.
4. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67:328–357.
5. Baffy G, Brunt EM, Caldwell SH. Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace. *J Hepatol.* 2012;56:1384–1391.
6. Wong RJ, Aguilar M, Cheung R, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology*. 2015;148:547–555.
7. Byrne CD, Targher G. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease: is universal screening appropriate? *Diabetologia*. 2016;59:1141–1144.
8. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41:1313–1321.
9. Musso G, Cassader M, Gambino R. Non-alcoholic steatohepatitis: emerging molecular targets and therapeutic strategies. *Nat Rev Drug Discov.* 2016;15:249–274.
10. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2015;149:389–397.
11. Hagstrom H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J Hepatol.* 2017;67:1265–1273.
12. Dulai PS, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: systematic review and meta-analysis. *Hepatology*. 2017;65:1557–1565.
13. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2019;156:1264–1281.
14. Vuppalanchi R, Siddiqui MS, Van Natta ML, et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. *Hepatology*. 2018;67:134–144.
15. Bedossa P, Patel K. Biopsy and noninvasive methods to assess progression of nonalcoholic fatty liver disease. *Gastroenterology*. 2016;150:1811–1822.
16. Musso G, Gambino R, Cassader M, et al. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med.* 2011;43:617–649.

17. Sumida Y, Yoneda M, Hyogo H, et al. Validation of the FIB4 index in a Japanese nonalcoholic fatty liver disease population. *BMC Gastroenterol.* 2012;12:2.
18. Loomba R, Adams LA. Advances in non-invasive assessment of hepatic fibrosis. *Gut.* 2020.
19. Patel K, Sebastiani G. Limitations of non-invasive tests for assessment of liver fibrosis. *JHEP Rep.* 2020;2:100067.
20. Sebastiani G, Vario A, Guido M, et al. Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. *J Hepatol.* 2006;44:686–693.
21. Castera L, Sebastiani G, Le Bail B, et al. Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. *J Hepatol.* 2010;52:191–198.
22. Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. *J Hepatol.* 2019;71:371–378.
23. Davyduke T, Tandon P, Al-Karaghoul M, et al. Impact of implementing a “FIB-4 First” strategy on a pathway for patients with NAFLD referred from primary care. *Hepatol Commun.* 2019;3:1322–1333.
24. Petta S, Wong VW, Camma C, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. *Aliment Pharmacol Ther.* 2017;46:617–627.
25. Boursier J, Guillaume M, Leroy V, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. *J Hepatol.* 2019;71:389–396.
26. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol.* 2016;64:1388–402.
27. Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology.* 2007;45:846–854.
28. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology.* 2006;43:1317–1325.
29. Harrison SA, Oliver D, Arnold HL, et al. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut.* 2008;57:1441–1447.
30. Wai CT, Greenon JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology.* 2003;38:518–526.
31. Shah AG, Lydecker A, Murray K, et al. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol.* 2009;7:1104–1112.
32. Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology.* 2013;145:782–789.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–845.
34. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Stat Med.* 2006;25:481–493.
35. Sun W, Cui H, Li N, 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.* 2016;46:862–870.
36. Anstee QM, Lawitz EJ, Alkhouri N, et al. Noninvasive tests accurately identify advanced fibrosis due to NASH: Baseline data from the STELLAR trials. *Hepatology.* 2019.
37. Chan WK, Treeprasertsuk S, Goh GB, et al. Optimizing use of nonalcoholic fatty liver disease fibrosis score, fibrosis-4 score, and liver stiffness measurement to identify patients with advanced fibrosis. *Clin Gastroenterol Hepatol.* 2019;17:2570–2580.
38. Stauffer K, Halilbasic E, Spindelboeck W, et al. Evaluation and comparison of six noninvasive tests for prediction of significant or advanced fibrosis in nonalcoholic fatty liver disease. *United European Gastroenterol J.* 2019;7:1113–1123.
39. Onnerhag K, Hartman H, Nilsson PM, et al. Non-invasive fibrosis scoring systems can predict future metabolic complications and overall mortality in non-alcoholic fatty liver disease (NAFLD). *Scand J Gastroenterol.* 2019;54:328–334.
40. Lambert J, Halfon P, Penaranda G, et al. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem.* 2008;54:1372–1378.
41. Boursier J, Vergniol J, Guillet A, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *J Hepatol.* 2016;65:570–578.
42. McPherson S, Hardy T, Dufour JF, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *Am J Gastroenterol.* 2017;112:740–751.
43. De Silva S, Li W, Kemos P, et al. Non-invasive markers of liver fibrosis in fatty liver disease are unreliable in people of South Asian descent. *Frontline Gastroenterol.* 2018;9:115–121.
44. Xiao G, Zhu S, Xiao X, et al. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. *Hepatology.* 2017;66:1486–1501.
45. Tapper EB, Sengupta N, Huniuk MGM, Afdhal NH, Lai M. Cost-effective evaluation of nonalcoholic fatty liver disease with NAFLD fibrosis score and vibration controlled transient elastography. *Am J Gastroenterol.* 2015;110:1298–1304.
46. Armstrong MJ, Houlihan DD, Benthall L, et al. Presence and severity of non-alcoholic fatty liver disease in a large prospective primary care cohort. *J Hepatol.* 2012;56:234–240.
47. Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology.* 2015;61:1547–1554.
48. Crossan C, Majumdar A, Srivastava A, et al. Referral pathways for patients with NAFLD based on non-invasive fibrosis tests: diagnostic accuracy and cost analysis. *Liver Int.* 2019;39:2052–2060.
49. Petta S, Wong VW, Camma C, et al. Improved noninvasive prediction of liver fibrosis by liver stiffness measurement in patients with nonalcoholic fatty liver disease accounting for controlled attenuation parameter values. *Hepatology.* 2017;65:1145–1155.