

A Single Blood Test Adjusted for Different Liver Fibrosis Targets Improves Fibrosis Staging and Especially Cirrhosis Diagnosis

Paul Calès ^{1,2} Jérôme Boursier,^{1,2} Frédéric Oberti,^{1,2} Valérie Moal,^{2,3} Isabelle Fouchard Hubert,^{1,2} Sandrine Bertrais,² Gilles Hunault,² Marie Christine Rousselet,^{2,4} and multicentric groups (SNIFF, ANRS HC 23)

Fibrosis blood tests are usually developed using significant fibrosis, which is a unique diagnostic target; however, these tests are employed for other diagnostic targets, such as cirrhosis. We aimed to improve fibrosis staging accuracy by simultaneously targeting biomarkers for several diagnostic targets. A total of 3,809 patients were included, comprising 1,012 individuals with chronic hepatitis C (CHC) into a derivation population and 2,797 individuals into validation populations of different etiologies (CHC, chronic hepatitis B, human immunodeficiency virus/CHC, nonalcoholic fatty liver disease, alcohol) using Metavir fibrosis stages as reference. FibroMeter biomarkers were targeted for different fibrosis-stage combinations into classical scores by logistic regression. Independent scores were combined into a single score reflecting Metavir stages by linear regression and called Multi-FibroMeter Version Second Generation (V2G). The primary objective was to combine the advantages of a test targeted for significant fibrosis (FibroMeter^{V2G}) with those of a test targeted for cirrhosis (CirrhoMeter^{V2G}). In the derivation CHC population, we first compared Multi-FibroMeter^{V2G} to FibroMeter^{V2G} and observed significant increases in the cirrhosis area under the receiver operating characteristic curve (AUROC), Obuchowski index (reflecting all fibrosis-stage AUROCs), and classification metric (six classes expressed as a correctly classified percentage) and a nonsignificant increase in significant fibrosis AUROC. Thereafter, we compared it to CirrhoMeter^{V2G} and observed a nonsignificant increase in the cirrhosis AUROC. In all 3,809 patients, respective accuracies for Multi-FibroMeter^{V2G} and FibroMeter^{V2G} were the following: cirrhosis AUROC, 0.906 versus 0.878 ($P < 0.001$; versus CirrhoMeter^{V2G}, 0.897, $P = 0.014$); Obuchowski index, 0.795 versus 0.791 ($P = 0.059$); classification, 86.0% versus 82.1% ($P < 0.001$); significant fibrosis AUROC, 0.833 versus 0.832 ($P = 0.366$). Multi-FibroMeter^{V2G} had the highest correlation with the area of portoseptal fibrosis and the highest reproducibility over time. Correct classification rates of Multi-FibroMeter with hyaluronate (V2G, 86.0%) or without (V3G, 86.1%) did not differ ($P = 0.938$). **Conclusion:** Multitargeting biomarkers significantly improves fibrosis staging and especially cirrhosis diagnosis compared to classical single-targeted blood tests. (*Hepatology Communications* 2018;2:455-466)

Introduction

Liver fibrosis is the main prognostic factor in chronic liver diseases (CLDs). Bridging fibrosis, often called significant fibrosis, indicates decreased life expectancy, and cirrhosis, the most severe step, exposes patients to liver complications. Therefore, guidelines recommend staging for

liver fibrosis and state that “cirrhosis represents the most relevant clinical endpoint.”⁽¹⁾ Pathologic evaluation, usually through liver biopsy, is the measurement reference for liver fibrosis. Recent guidelines state that liver biopsy can be replaced by noninvasive tests in certain circumstances, especially in chronic viral hepatitis and nonalcoholic fatty liver disease (NAFLD).⁽¹⁾

Abbreviations: ALD, alcoholic liver disease; AUROC, area under the receiver operating characteristics; CHB, chronic hepatitis B; CHC, chronic hepatitis C; CLD, chronic liver disease; F, fibrosis stage; HIV, human immunodeficiency virus; NAFLD, nonalcoholic fatty liver disease; PPV, positive predictive value; V2G, version second generation; VCTE, vibration-controlled transient elastometry.

Received October 26, 2017; accepted January 29, 2018.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1161/full.

Following the development of pathologic staging in the early 1990s, the first blood tests to combine biomarkers were targeted for significant liver fibrosis.⁽²⁾ Since then, blood tests have become popular; a systematic review performed of chronic hepatitis C (CHC) evaluated no less than 18 tests.⁽³⁾ Significant fibrosis was initially chosen as the diagnostic target in CHC for two reasons. The first was clinical; significant fibrosis was, at that time, an important cutoff at which antiviral treatment was indicated. The second was statistical; modeling a binary outcome is much easier than modeling a multistage outcome. However, the precise targeting of these tests for significant fibrosis limits their diagnostic pertinence for more distant diagnostic targets, such as cirrhosis. Therefore, in 2009, we proposed a blood test, called CirrhoMeter, specifically designed to diagnose cirrhosis.⁽⁴⁾ CirrhoMeter incorporated the same biomarkers used in our earlier test, called FibroMeter, for the diagnosis of significant fibrosis.⁽⁵⁾ However, in CirrhoMeter, those biomarkers were weighted differently to significantly improve the accuracy of cirrhosis diagnosis. There is nonetheless a downside to having two tests as clinicians may find it difficult to determine which test is appropriate for a given patient, especially when the results of the two tests might differ.

Thus, our primary objective for the present work was to develop and evaluate a new method for constructing diagnostic tests in order to significantly improve the diagnostic accuracy through the use of a

multitarget diagnostic system in CHC. The resulting unique test had to combine good performance for overall fibrosis staging, as does FibroMeter, and good performance for cirrhosis diagnosis, as does CirrhoMeter. The secondary objectives were to validate the various clinically useful characteristics of the new test (e.g., accuracy, robustness in different etiologies, reproducibility).

Patients and Methods

POPULATIONS

A total of 3,809 patients were included in the present study (Table 1). Patient duplication between studies was corrected to ensure that all patients were included only once in the statistical analysis. A flow chart of the four populations used for the present work is provided in Table 2. The studies were approved by an institutional review board, and written informed consent was obtained from all patients.

Derivation Population

Derivation population #1, detailed elsewhere,⁽⁶⁾ included 1,012 patients with CHC. This population provided individual patient data from five centers independent of patient recruitment, blood marker determination, and liver histology interpretation by expert pathologists.

Supported in part by the Agence Nationale de Recherches sur le Sida et les Hépatites Virales through the HC EP 23 Fibrostar study, which furnished a portion of the subjects.

Copyright © 2018 The Authors. Hepatology Communications published by Wiley Periodicals, Inc., on behalf of the American Association for the Study of Liver Diseases. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep4.1161

Potential conflict of interest: Dr. Calès is the inventor of the FibroMeter tests; Dr. Boursier is a consultant for Echosens. The other authors have nothing to report.

ARTICLE INFORMATION:

From the ¹Liver-Gastroenterology Department, University Hospital, Angers, France; ²HIFIH Laboratory, Angers University, Bretagne Loire University, Angers, France; ³Department of Biochemistry and Genetics and ⁴Pathology Department, University Hospital, Angers, France.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Paul Calès, M.D.
Service d'hépatogastroentérologie, CHU, 49933
Angers Cedex 09, France

E-mail: paul.cales@univ-angers.fr
Tel: +33 2 41 35 34 10

TABLE 1. POPULATION CHARACTERISTICS

	Derivation		Validation				
	#1	#2	#3a	#3b	#3c	#3d	#4
Etiology	CHC	CHC	CHB	HIV/CHC	NAFLD	Alcohol	Miscellaneous
Patients (n)	1,012	641	152	444	225	115	1,220
Male (%)	59.6	60.5	81.5	68.7	65.3	64.3	67.3
Age (years)	45.4 ± 12.5	51.4 ± 11.2	40.0 ± 11.3	40.5 ± 5.8	56.5 ± 12.0	50.8 ± 23.9	50.7 ± 13.3
Body mass index (kg/m ²)	NA	24.8 ± 4.0	NA	NA	31.3 ± 5.0	23.9 ± 4.2	29.2 ± 6.3
Metavir (%):							
F0	4.3	3.7	15.1	5.9	25.3	11.3	10.1
F1	43.3	38.7	44.1	24.3	37.3	14.8	32.5
F2	27.0	25.4	25.7	38.5	16.9	14.8	25.0
F3	13.9	18.4	6.6	19.6	15.6	7.0	17.5
F4	11.4	13.7	8.6	13.7	4.9	52.2	14.8
Score	1.85 ± 1.08	2.00 ± 1.13	1.49 ± 1.10	2.11 ± 1.10	1.37 ± 1.16	2.74 ± 1.49	1.94 ± 1.22
Significant fibrosis (%)	52.3	57.6	40.8	69.8	37.3	73.9	57.4
Biopsy length (mm)	21.2 ± 7.9	24.4 ± 8.7	21.6 ± 7.4	20.8 ± 9.9	30.8 ± 12.0	NA	27.6 ± 11.4
Biopsy length ≥15 mm (%)	83.8	92.1	87.1	73.0	92.4	-	88.0

Abbreviation: NA, not available.

Validation Populations

Validation population #2, also detailed elsewhere,^(7,8) included 641 patients with CHC and all tests required for the present work. Validation population #3 included 936 patients with etiologies other than CHC.⁽⁹⁾ Patients are described per etiology hereafter but grouped in the statistical analyses. Subpopulation #3a with chronic hepatitis B (CHB) was extracted from a published database.⁽¹⁰⁾ It included 152 patients with chronic hepatitis (30.4% hepatitis B e antigen positive); inactive carriers of hepatitis B e antigen were excluded. Validation population #3b comprised 444 patients with CHC and human immunodeficiency virus (HIV) infection prospectively included from April 1997 to August 2007 if they had hepatitis C virus RNA and anti-HIV antibodies in serum.⁽¹¹⁾ Subpopulation #3c comprised 225 patients with biopsy-proven NAFLD consecutively included in the study from January 2002 to March 2013 at Angers University Hospital and from September 2005 to July 2011 at Pessac University Hospital. This subpopulation was extracted from a recently published database.⁽¹²⁾ Population #3d was extracted from a database used in a previous study⁽⁵⁾ and included 115 patients with alcoholic liver disease (ALD). Validation population #4 included 1,220 patients with different CLD etiologies as follows: CHC, 41.3%; NAFLD, 31.3%; alcohol, pure (ALD), 8.1% or mixed, 11.7%; CHB, 5.7%; coinfections (HIV/CHC, HIV/CHB, CHB/VHD hepatitis delta virus, others), 1.2%; other combinations of previous etiologies, 0.7%. These patients were consecutively included between 2011 and 2016 in the Angers and

Pessac centers. They thus reflect current clinical practice where liver biopsy is more often indicated when blood tests and vibration-controlled transient elastometry (VCTE) are discordant.⁽¹⁾ For that reason, validation population #4 was considered separately.

DIAGNOSTIC METHODS

Histologic Assessment

Liver biopsies were performed using Menghini's technique with a 1.4-1.6-mm diameter needle. Biopsy specimens were fixed in a formalin-alcohol-acetic solution and embedded in paraffin; 5- μ m-thick sections were then cut and stained with hematoxylin-eosin-saffron. Liver fibrosis was evaluated according to Metavir fibrosis stage (F)⁽¹³⁾ by two senior experts with a consensus reading in case of discordance at Angers and in the Fibrostar study⁽¹⁴⁾ (part of validation population #2) and by a senior expert in other centers. These liver specimen findings served as a reference for the liver fibrosis evaluation by noninvasive tests. The area of portoseptal fibrosis in population #2 was centrally measured by automated morphometry as recently described.⁽¹⁵⁾

TABLE 2. FLOW CHART OF POPULATIONS

Population #	Etiology of Chronic Liver Disease	Patients (n)	Test Aim
1	CHC	1,012	Derivation
2	CHC	641	Validation
3	CHB, CHC/HIV, NAFLD, ALD	936	Validation
4	Miscellaneous	1,220	Validation
Total	Miscellaneous	3,809	-

Fibrosis Biomarkers

Blood markers were those previously used in our blood tests to diagnose different lesions in chronic viral hepatitis.^(5,16) We included platelets, aspartate aminotransferase, hyaluronate, urea, prothrombin index, and alpha2-macroglobulin as used in FibroMeter^{V2G(5,6)} plus gamma-glutamyl transpeptidase used in FibroMeter^{V3G(16)} and alanine aminotransferase used in InflaMeter targeted for liver activity.⁽¹⁷⁾ We also included demographic data (age and sex as used in FibroMeter^{V2G}). Thus, 10 biomarkers were available. The new test was constructed with hyaluronate (second generation as for FibroMeter^{V2G}) or without (third generation as for FibroMeter^{V3G}, a cheaper test). Reference blood tests for comparison with the new test were mainly FibroMeter^{V2G} or FibroMeter^{V3G}, targeted for significant fibrosis ($F \geq 2$); and accessorially CirrhoMeter^{V2G} or CirrhoMeter^{V3G}, targeted for cirrhosis.

VCTE was used as an independent reference for the noninvasive diagnosis of liver fibrosis, especially for cirrhosis diagnosis; details are provided in the [Supporting Materials](#).

METHODOLOGICAL DESIGN

Objectives

The aim of the primary objective was to develop Multi-FibroMeters and to compare their accuracies with FibroMeters or CirrhoMeters in the CHC derivation population. The aims of the secondary objectives were to validate the various clinically useful characteristics of the new test: accuracy in validation populations or as a function of etiology or between Multi-FibroMeter^{V2G} and Multi-FibroMeter^{V3G}, correlation with pathologic measurements, classification metric, and reproducibility.

Study Type

The present work was a retro-prospective diagnostic study (data collection organized upstream of the performance of index tests and the reference standard).⁽¹⁸⁾ It combined types 2b (nonrandomly split data) and 3 (separate data) studies according to the transparent reporting of individual prognosis or diagnosis (TRI-POD) classification.⁽¹⁹⁾

Judgment Criteria

The judgment criteria applied to the accuracy characteristics of the primary and secondary objectives. The

primary objective was to combine the advantages of a test targeted for significant fibrosis (FibroMeter) and those of a test targeted for cirrhosis (CirrhoMeter). Thus, the main judgment criterion was a composite based on three statistical comparisons. First, the area under the receiver operating characteristic curve (AUROC) for cirrhosis of the multitargeted test had to be superior or equal to that of CirrhoMeter. Second, this implied that the multitargeted test had to have an AUROC for cirrhosis significantly superior to that of FibroMeter, as demonstrated.⁽⁴⁾ Third, the AUROC for significant fibrosis of the multitargeted test had to be superior or equal to that of FibroMeter.

The secondary criteria evaluated the overall accuracy of the tests. This included an Obuchowski index and a classification metric (see below) of the multitargeted tests that were expected to be significantly superior to those of FibroMeter.

Finally, as ancillary criteria, we also determined the statistical gains of the multitargeted test. The reference was the test used in judgment criteria comparisons (FibroMeter or CirrhoMeter in one case). For example, FibroMeter was significantly inferior to CirrhoMeter for the cirrhosis AUROC, while the corresponding multitargeted test was statistically superior (or not different) to CirrhoMeter.

Populations

The multitargeted test was developed in derivation population #1 with CHC (1,012 patients) because this etiology provides the most robust fibrosis test among the main CLD etiologies.⁽⁹⁾ Thereafter, it was validated in the validation populations (2,797 patients), first in population #2 (also CHC) and then in populations #3 and #4. The main reported results are those observed in the largest populations without optimism bias for comparison.

Fibrosis Metrics

Liver fibrosis is measured according to two main metrics. The first is a scoring metric. Considering noninvasive measurement by blood tests, liver fibrosis is usually measured according to a score expressing the probability (i.e., from 0 to 1 [or 100%]) of a single diagnostic target (usually significant fibrosis) through a statistical transformation. This scoring metric is the most popular. Considering liver biopsy, the quantitative morphometric measurement provides an area (or surface, expressed in %) of fibrosis.⁽¹⁵⁾

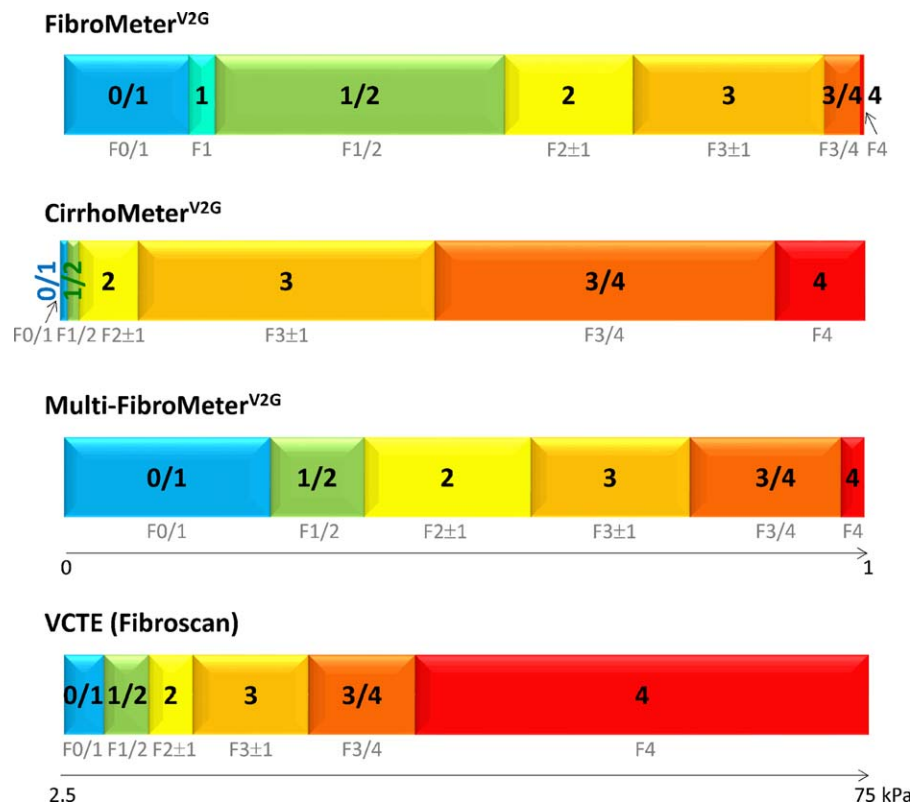


FIG. 1. Fibrosis classifications for FibroMeter^{V2G}, CirrhoMeter^{V2G}, Multi-FibroMeter^{V2G}, and VCTE (Fibroscan). Each fibrosis class, depicted in boxes numbered from 0/1 to 4 (determined by two cutoffs of the test scores), corresponds to around 90% of the Metavir fibrosis stages indicated below each box. The horizontal arrows indicate the scale, range, and unit for score values. This diagram shows that classifications of CirrhoMeter^{V2G} and VCTE are close as the majority of their ranges are devoted to cirrhosis. It also shows that the Multi-FibroMeter^{V2G} classification displays an equilibrated fibrosis spectrum with intermediate values falling between FibroMeter^{V2G} and CirrhoMeter^{V2G} classifications.

The second metric is a classification metric. A blood score is proportional to, and thus can be translated into, a classification of fibrosis stages reflecting pathologic staging.⁽²⁰⁾ Classifications of FibroMeters,⁽²¹⁾ CirrhoMeters,⁽²²⁾ VCTE,⁽²²⁾ and multitargeted tests generally include six fibrosis classes reflecting Metavir staging (Fig. 1). The usual semiquantitative staging performed by pathologists (e.g., stages 0 to 4 in the Metavir system) falls into this metric category.

TEST CONSTRUCTION

The construction of the multitarget staging system was performed in four successive steps.

Step 1 was biomarkers acquisition, and we listed selected biomarkers as single markers or as combined markers, such as a ratio. Step 2 was single-target test construction. These tests were built using a conventional binary logistic regression providing scores from 0 to 1. We used as many diagnostic targets as possible through the five Metavir fibrosis stages. These targets included classical binary targets with a single cutoff: fibrosis ($F \geq 1$), significant fibrosis ($F \geq 2$), severe fibrosis ($F \geq 3$), and cirrhosis ($F = 4$). Another six

targets were obtained by binary targets using two cutoffs, e.g., F1 or F1 + F2 or F1 + F2 + F3 versus other stages. We thus obtained 10 single-targeted tests. It should be noted that FibroMeter and CirrhoMeter might be slightly different from the corresponding single-target tests due to step 1. In step 3, the single-target test selection, the 10 single-targeted tests were included in stepwise multiple linear regression targeted for the five Metavir stages. The resulting score was normalized to 1, i.e., divided by 4, to obtain a score between 0 and 1. This new score was called the multitargeted FibroMeter (Multi-FibroMeter) score. In step 4, the multitarget test classification, we derived the correspondence between the Multi-FibroMeter score and Metavir stages according to our published classification metric.⁽²¹⁾ The six fibrosis classes developed for Multi-FibroMeter in the present optional step were: 0/1 (corresponding to Metavir F0/1), 1/2 (F1/2), 2 (F2 ± 1), 3 (F3 ± 1), 3/4 (F3/4), and 4 (F4).

The construction for FibroMeter and CirrhoMeter was performed twice with (Multi-FibroMeter^{V2G}) or without (Multi-FibroMeter^{V3G}) hyaluronate, which is a costly biomarker. Thus, notably in the Results, the various meters may be referred to in the plural form.

STATISTICS

The statistical calculations of Multi-FibroMeter are complex and thus are not provided; however, calculations can be freely performed for significant studies upon request by contacting the corresponding author of this article.

Accuracy

The diagnostic accuracy of each test score (scoring metric) was expressed with two descriptors. The main descriptor for score accuracy was the AUROC, i.e., the classical index for binary diagnostic targets. Multi-FibroMeter AUROC \geq FibroMeter or CirrhoMeter AUROC was tested with a unilateral classical test for significant difference and a noninferiority test with a margin close to 0. Multi-FibroMeter \approx VCTE for cirrhosis AUROC was tested with a bilateral classical test for significant difference and an equivalence test.

The other descriptor was the Obuchowski index⁽²³⁾ to reflect overall staging and to better take into account differences in fibrosis stage prevalence between populations and thus limit spectrum bias. This index is a multinomial version of the AUROC adapted to ordinal references, such as pathologic fibrosis staging. With n ($n = 5$, F0-F4) categories of the gold standard outcome and AUROC_{*st*}, the index estimates the AUROC of diagnostic tests differentiating between categories s and t . The Obuchowski index is a weighted average of the $n(n-1)/2 = 10$ different AUROC_{*st*} corresponding to all the pair-wise comparisons between two of the n categories. Additionally, the Obuchowski index was assessed using a penalty function proportional to the difference in fibrosis stages, i.e., a penalty of 1 when the difference between stages was 1, 2 when the difference was 2, and so on. The reference prevalence was standardized according to the largest series of CHC with liver biopsies⁽²⁴⁾ to standardize comparisons between etiologies, as reported.⁽⁹⁾ Thus, the result can be interpreted as the probability that the noninvasive test will correctly rank 2 randomly chosen patients with different fibrosis stages. The overall accuracy of the classification metric was assessed by the percentage rate of patients who were well-classified according to Metavir fibrosis stage.

Optimism Bias

By definition, optimism bias maximizes performance in the population where a test is constructed; this affected FibroMeters, CirrhoMeters, and Multi-FibroMeters in derivation population #1. Thus, this

bias was always noted in concerned results, and external validation was necessarily performed outside these populations.

Sample Size Calculation

The size of the main populations (derivation #1 and validation #2) was the size necessary to detect a significant difference between two tests for the diagnosis of cirrhosis (main judgment criterion of the primary objective). With an α risk of 0.05, a β risk of 0.05, a cirrhosis prevalence of 0.12, an AUROC correlation of 0.82, and bilateral testing, the required sample size was 659 patients for the following expected AUROC values for cirrhosis: Multi-FibroMeter^{V2G}, 0.92; FibroMeter^{V2G}, 0.90.⁽⁴⁾

Miscellaneous

Quantitative variables were expressed as mean \pm SD. All P values were bilateral unless otherwise specified. Data were reported according to the 2015 Standards for Reporting of Diagnostic Accuracy Studies (STARD)⁽¹⁸⁾ and Liver FibroSTARD⁽²⁵⁾ statements. Data were raw without correction or exclusion. Thus, analyses were based on the intention-to-diagnose principle; missing data were not replaced. The main statistical analyses were performed under the control of professional statisticians (S.B., G.H.) using SPSS version 18.0 (IBM, Armonk, NY) and SAS 9.2 (SAS Institute Inc., Cary, NC).

Results

MAIN CHARACTERISTICS

The main characteristics of the studied populations are depicted in Table 1. Liver biopsy length was ≥ 15 mm in 85.5% of patients. The main results on accuracy by scoring and classification metrics in single or combined populations are listed in Tables 2 to 6. Details of the statistical comparisons are available in the [Supporting Materials](#).

PRIMARY OBJECTIVE

Development

We obtained a multitargeted test made of several independent single-targeted tests (e.g., seven for Multi-FibroMeter^{V2G}; Fig. 2). The Multi-FibroMeter score distribution as a function of Metavir fibrosis

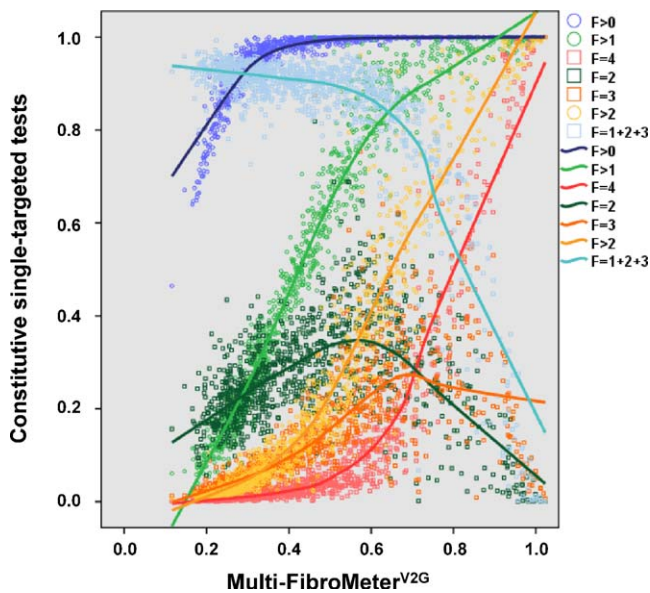


FIG. 2. Scatter plot of Multi-FibroMeter^{V2G} with its seven constitutive single-targeted tests in population #1 (1,012 patients). Curves depict nonlinear locally weighted scatterplot smoothing regression.

stages fell between those of FibroMeter and CirrhoMeter (Fig. 3).

Accuracy

In derivation population #1 (Table 3; Supporting Table S9), the AUROCs for cirrhosis of the Multi-FibroMeters were higher than those of FibroMeters (significant difference) and CirrhoMeters (nonsignificant difference). The AUROCs for significant fibrosis of Multi-FibroMeters were higher than those of FibroMeters (nonsignificant difference).

The Obuchowski indexes of Multi-FibroMeters were significantly higher than those of the corresponding FibroMeters (Table 3). The rates of patients correctly classified in classification metrics were significantly higher ($P < 0.001$) in Multi-FibroMeters versus the corresponding FibroMeters. Thus, the primary objective criteria were fulfilled in derivation population #1.

SECONDARY OBJECTIVES

Accuracy Validation as a Function of Etiology

The differences observed in the derivation population were also observed in CHC validation population #2

(Table 4; Supporting Table S10) with the following exceptions: a better result with the cirrhosis AUROCs of Multi-FibroMeters, which were significantly improved versus CirrhoMeters, and a poorer result with the Obuchowski indexes of Multi-FibroMeters, which remained higher than those of FibroMeters but with a nonsignificant difference. Thus, the main criterion was fulfilled in validation population #2 and was even surpassed in one out of its three composite comparisons.

The AUROCs for cirrhosis and Obuchowski indices of Multi-FibroMeters were superior to FibroMeters in the non-CHC etiologies of population #3 (Supporting Table S1), especially in CHB, HIV/CHC, NAFLD (except Multi-FibroMeter^{V3G} for the Obuchowski index), ALD (except Multi-FibroMeter^{V2G} for the cirrhosis AUROC) (Supporting Table S2), and in the miscellaneous etiologies of validation population #4 (Supporting Table S3). Multi-FibroMeters were significantly superior to FibroMeters in the classification metric in populations #3 and #4 (Supporting Table S4).

Accuracy in Combined Populations

Considering the previous results, the four populations were combined and the diagnostic performance evaluated in the whole population because there was

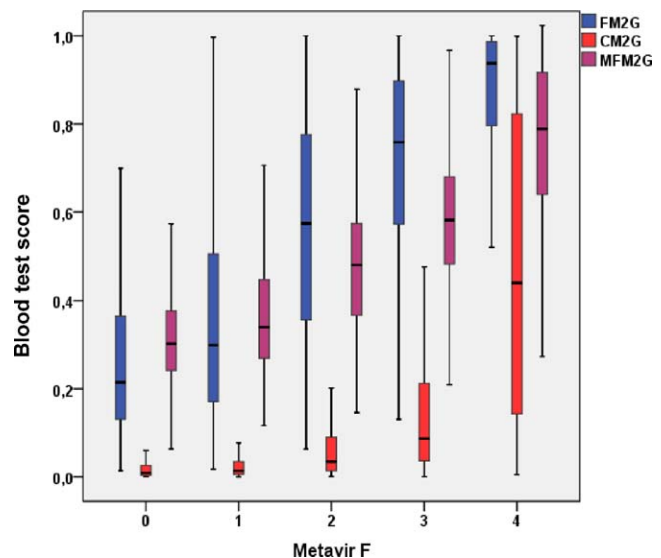


FIG. 3. Distribution of Multi-FibroMeter^{V2G}, FibroMeter^{V2G}, and CirrhoMeter^{V2G} scores as a function of Metavir fibrosis stages in population #1 to #4 (3,809 patients). Box plots depict medians, 25%, and 75% quartiles and extremes. Abbreviations: CM2G, CirrhoMeter^{V2G}; FM2G, FibroMeter^{V2G}; MFM2G, Multi-FibroMeter^{V2G}.

TABLE 3. DIAGNOSTIC PERFORMANCE WITHIN THE FIBROMETER FAMILY IN CHC DERIVATION POPULATION #1 (1,012 PATIENTS)

	AUROC				Obuchowski Index		Classification Metric	
	F ≥ 1	F ≥ 2	F ≥ 3	F = 4	Value	Rank	Rate	
FibroMeter ^{V2G}	0.854	0.853	0.884	0.907	0.843	3	87.6	
CirrhoMeter ^{V2G}	0.825	0.811	0.874	0.919	0.819	5	-	
Multi-FibroMeter ^{V2G}	0.862	0.855*	0.901	0.932*	0.854*	1	91.7*	
FibroMeter ^{V3G}	0.852	0.851	0.880	0.893	0.838	4	86.9	
CirrhoMeter ^{V3G}	0.821	0.814	0.874	0.911	0.818	6	-	
Multi-FibroMeter ^{V3G}	0.859	0.852*	0.896	0.923*	0.850*	2	92.5*	

The best result per diagnostic target is indicated in bold. * depicts criterion of primary objective reached. Details on *P* values of pair comparisons are reported in [Supporting Table S4](#) for classification and in Additional Material for scoring ([Table A1](#)).

TABLE 4. DIAGNOSTIC PERFORMANCE WITHIN THE FIBROMETER FAMILY IN CHC VALIDATION POPULATION #2 (641 PATIENTS)

	AUROC				Obuchowski Index		Classification Metric	
	F ≥ 1	F ≥ 2	F ≥ 3	F = 4	Value	Rank	Rate	
FibroMeter ^{V2G}	0.827	0.812	0.830	0.863	0.797	2	84.2	
CirrhoMeter ^{V2G}	0.783	0.785	0.816	0.858	0.771	5	-	
Multi-FibroMeter ^{V2G}	0.824	0.814*	0.844	0.888*	0.803*	1	88.3*	
FibroMeter ^{V3G}	0.819	0.798	0.816	0.844	0.785	4	81.7	
CirrhoMeter ^{V3G}	0.769	0.771	0.796	0.840	0.756	6	-	
Multi-FibroMeter ^{V3G}	0.815	0.805*	0.827	0.870*	0.791*	3	88.9*	

The best result per diagnostic target is indicated in bold. * depicts criterion of primary objective reached. *P* values of pair comparisons are reported in [Supporting Table S4](#) for classification and in Additional Material for scoring ([Table A2](#)).

no optimism bias in statistical comparisons within the FibroMeter family ([Table 5](#); [Supporting Table S11](#)).

Four items within the main and secondary judgment criteria were fulfilled and one item was surpassed, with the cirrhosis AUROCs of Multi-FibroMeters significantly improved versus the corresponding CirrhoMeters. These criteria were also validated in the three combined validation populations #2 to #4 except for one of the two comparisons for the secondary criteria, the Obuchowski index of Multi-FibroMeter^{V2G}, which was higher than that of FibroMeter^{V2G} but with a nonsignificant difference ([Supporting Table S5](#)).

Finally, all the accuracy criteria (superiority to FibroMeter except for noninferiority of the significant fibrosis AUROC of FibroMeter and the cirrhosis AUROC of CirrhoMeter) were reached with Multi-FibroMeter^{V2G} and Multi-FibroMeter^{V3G} ([Supporting Table S6](#)).

All accuracies in the scoring metrics (Obuchowski index and AUROCs) of tests^{V2G} (FibroMeter^{V2G}, CirrhoMeter^{V2G}, Multi-FibroMeter^{V2G}) were significantly higher than corresponding tests^{V3G} ([Table 5](#); [Supporting Table S11](#) and [Supporting Materials](#)). Similarly, the correct classification rate in the

TABLE 5. DIAGNOSTIC PERFORMANCE WITHIN THE FIBROMETER FAMILY IN COMBINED POPULATIONS #1 TO #4 (3,809 PATIENTS)

	AUROC				Obuchowski Index		Classification	
	F ≥ 1	F ≥ 2	F ≥ 3	F = 4	Value	Rank	Rate	Rank
FibroMeter ^{V2G}	0.788	0.832	0.849	0.878	0.791	2	82.1	3
CirrhoMeter ^{V2G}	0.747	0.800	0.846	0.897	0.769	5	81.8	4
Multi-FibroMeter ^{V2G}	0.778	0.833*	0.863	0.906*	0.795*	1	86.0*	2
FibroMeter ^{V3G}	0.767	0.823	0.837	0.855	0.776	4	79.5	6
CirrhoMeter ^{V3G}	0.722	0.790	0.835	0.879	0.754	6	80.8	5
Multi-FibroMeter ^{V3G}	0.764	0.823*	0.849	0.886*	0.782*	3	86.1*	1

Best result per diagnostic target is indicated in bold. * depicts a criterion of primary objective reached (details in [Supporting Table S6](#)). Details on *P* values of pair comparisons are reported in [Table 6](#) for classification and in Additional Material for scoring ([Table A3](#)).

TABLE 6. CORRECTLY CLASSIFIED PATIENT RATES (%) IN THE CLASSIFICATION METRIC WITHIN THE FIBROMETER FAMILY IN COMBINED POPULATIONS #1 TO #4 (3,809 PATIENTS)

Combined Populations	#1 to #4	#2 to #4
Patients (n)	3,809	2,797
FibroMeter ^{V2G}	82.1	80.2
Multi-FibroMeter ^{V2G}	86.0*	84.0*
<i>P</i> [†]	<0.001	<0.001
FibroMeter ^{V3G}	79.5	76.8
Multi-FibroMeter ^{V3G}	86.1*	83.7*
<i>P</i> [†]	<0.001	<0.001
Test ^{V2G} vs Test ^{V3G} (<i>P</i> [‡]):	-	-
FibroMeters	<0.001	<0.001
Multi-FibroMeters	<u>0.938</u>	<u>0.592</u>

Significant differences (*P*) are shown in bold. Underlined numbers indicate a significant gain for Multi-FibroMeter^{V3G} versus Multi-FibroMeter^{V2G} in the comparison with corresponding FibroMeters. *depicts a criterion of primary objective reached; [†]Comparison of Multi-FibroMeter and corresponding FibroMeter by paired McNemar test; [‡]Comparison of FibroMeter^{V2G} vs FibroMeter^{V3G} or Multi-FibroMeter^{V2G} vs Multi-FibroMeter^{V3G} by paired McNemar test.

classification metric was significantly higher in FibroMeter^{V2G} than in FibroMeter^{V3G} (Table 6; Supporting Table S12). In contrast, classification rates were not significantly different between Multi-FibroMeter^{V2G} and Multi-FibroMeter^{V3G}, which indicates that the Multi-FibroMeter^{V3G} classification compensated for the deficit of FibroMeter^{V3G} against FibroMeter^{V2G} observed in any metric. This result was also observed in the three combined validation populations #2 to #4 (Table 6).

A statistical gain was observed for Multi-FibroMeters (Supporting Table S6) in four out of five accuracy comparisons of judgment criteria (except for the comparison of AUROC for significant fibrosis with FibroMeter, as expected).

Other Validations

Pathologic validation was performed in population #2, without optimism bias, where liver morphometry was available in 510 patients. Among the available tests, the scores of Multi-FibroMeters had the highest correlations with Metavir fibrosis stages and area of portoseptal fibrosis, which was a pathologic reference independent of its construction (Supporting Table S7). The relationship between Multi-FibroMeters and Metavir fibrosis stages or area of portoseptal fibrosis showed a larger value range and greater linearity of Multi-FibroMeters to reflect fibrosis level than FibroMeters or CirrhoMeters (Supporting Fig. S1); this indicates a better reflection of the fibrosis spectrum.

Multi-FibroMeter classification was also validated. First, Multi-FibroMeter^{V2G} classes and scores were well correlated with Metavir fibrosis stages ($r_s = 0.63, 0.65$, respectively), offering higher coefficients than other tests (Supporting Table S7). Second, Multi-FibroMeter^{V2G} classification accuracy ranked first.

Multi-FibroMeter reproducibility over time (Supporting Table S8) was better than other single-targeted blood tests or pathologic measurements.

Discussion

In the present study, we used a statistical approach to improve noninvasive fibrosis staging. By multitargeting biomarkers, different tests single targeted toward various diagnostic targets can be combined into one final test.

The main advantage of combining single targeted tests is the significant increase in diagnostic performance of Multi-FibroMeters compared to their corresponding FibroMeters. Thus, the AUROCs for cirrhosis of Multi-FibroMeters were significantly increased compared to their FibroMeter counterparts. It should be noted that for cirrhosis diagnosis, the most relevant comparator for the evaluation of Multi-FibroMeter is FibroMeter (the reference test for Multi-FibroMeter) and not CirrhoMeter. Our objective was to add the cirrhosis diagnosis performance of CirrhoMeter to FibroMeter. Unexpectedly, Multi-FibroMeters had significantly higher AUROCs for cirrhosis than the CirrhoMeter counterparts. This result was due to a significantly higher accuracy in patients without cirrhosis with Multi-FibroMeter versus CirrhoMeter (details in Supporting Materials). Considering the discrimination of Metavir fibrosis stages, the performance of Multi-FibroMeters, evaluated by the Obuchowski index, was significantly increased compared to the FibroMeter counterparts. Regarding fibrosis classifications reflecting Metavir stages, Multi-FibroMeters had significantly higher accuracy than FibroMeters.

In the noninvasive diagnosis of cirrhosis, the usual reference is liver elastometry. Our results (details in Supporting Material) showed that the AUROCs for cirrhosis of VCTE and Multi-FibroMeter^{V2G} were equivalent. Furthermore, the Multi-FibroMeter advantage will likely be superior in real conditions where VCTE may be not applicable due to failure or unreliability. However, the quality criteria for VCTE have been improved in recent years (operator, probe). For fibrosis staging, the AUROC for significant

fibrosis and the Obuchowski index were significantly increased in Multi-FibroMeters compared to the corresponding FibroMeters. This last result was confirmed by the classification metric. Finally, all judgement criteria were reached with Multi-FibroMeters, i.e., a statistical gain for four criteria compared to FibroMeters.

Multi-FibroMeter^{V3G} classification, which provided accuracy similar to that of Multi-FibroMeter^{V2G} classification, compensated for the deficit of FibroMeter^{V3G} observed against FibroMeter^{V2G} or VCTE in any metric and for the deficit of Multi-FibroMeter^{V3G} observed against VCTE in the scoring metric (details in the [Supporting Material](#)). Thus, Multi-FibroMeter^{V3G} can replace the costlier Multi-FibroMeter^{V2G} provided that the classification metric is used. The advantage of the classification metric over the scoring metric has recently been shown; it dramatically reduces the gray zone due to discordance between FibroMeter and VCTE.⁽²⁶⁾ Finally, the originality of the new method is that it provides a unique polyvalent test offering good performance for significant fibrosis (like FibroMeter) and good performance for cirrhosis (like VCTE). Concerning the latter diagnostic endpoint, the new method outperforms cirrhosis-dedicated blood tests, such as CirrhoMeter. Thus, Multi-FibroMeter is a synergistic test that significantly improves CirrhoMeter's accuracy for cirrhosis, which was already significantly improved compared to FibroMeter.

Cirrhosis being the main diagnostic target, some might argue that a binary diagnosis with a single-targeted test using a single cutoff would be sufficient. However, this more classical approach has two main limitations. First, if we consider VCTE (Fibroscan) as a reference for noninvasive cirrhosis diagnosis, the commonly employed cutoff of 14 kPa has a positive predictive value (PPV) for cirrhosis of only 57% in CHC⁽²²⁾ (55% in the present CHC population #2 or 47% with the Youden cutoff; results not shown). The interest of the present Multi-FibroMeter classification is that it provides three categories of cirrhosis diagnosis: a firm class for definitive cirrhosis (class 4, cirrhosis PPV of Multi-FibroMeter^{V2G}, 83%; results not shown), a firm class for early cirrhosis (class 3/4, cirrhosis PPV, 64%), and a remaining class for doubtful cirrhosis (class 3 [F3 ± 1], cirrhosis PPV, 25%), leaving only 6% of patients with cirrhosis undetected. In the doubtful class 3, test results will need to be considered in light of other available examinations, such as VCTE alone or combined with a blood test⁽²⁷⁾ and/or imaging and closer follow-up. The second limitation

to binary cirrhosis diagnosis is that non-cirrhosis results leave clinicians with great uncertainties. In particular, they cannot easily distinguish patients with severe fibrosis, who will require close follow-up or more active intervention, from those without it. In that respect, a detailed and effective classification is far more informative.

The classification metric is infrequently evaluated in the literature despite its common use in clinical practice where it offers easy, comprehensive, and precise results for each fibrosis stage (far more eloquent than the Obuchowski index). The fibrosis classification metric used here has been validated in several papers^(21,22,26,27) and independent populations.^(20,28) The "imprecision" of fibrosis classification, e.g., the information furnished by fibrosis class 3 (Metavir F3 ± 1), is specious as it reflects Metavir staging imprecision. Indeed, this classification metric is more accurate than Metavir staging when referenced against objective outcomes.⁽²⁹⁾ It also reduces gray zones and ensuing biopsy requirements.^(7,26) Finally, for all these reasons, factual classifications of noninvasive fibrosis tests can be considered per se as an accurate fibrosis staging metric with its own classes from 0/1 to 4 (Fig. 1).⁽²⁶⁾

The performance gain conferred by Multi-FibroMeters might seem modest (relative gain in correct classification versus FibroMeters is 4.8% for Multi-FibroMeter^{V2G} and 8.3% for Multi-FibroMeter^{V3G}; Table 6), but the baseline performance of FibroMeters is already high for that aspect. Of greater interest is the gain in reduction of misclassification (relative gain in misclassification versus FibroMeters is -22% for Multi-FibroMeter^{V2G} and -32% for Multi-FibroMeter^{V3G}). This means that 1 out of 3 or 5 patients misclassified by FibroMeters is correctly classified by Multi-FibroMeters. Considering that liver biopsy is not a "gold" but only a "best" standard, the Multi-FibroMeter accuracy was nevertheless improved by comparisons not only to pathologic staging as reference but also to liver morphometry in a population without optimism bias. There are at least two ways to address the biopsy limitations. First, biopsy quality could be improved, for example, imposing a size ≥15 mm, but this was not always the case in the subpopulations of the present study. Second, a statistical method without a gold standard could be deployed,⁽³⁰⁾ but that approach is better suited to test comparisons than it is to test development. We also mention that the reliability criteria of noninvasive tests and the impact of the XL probe of VCTE were not considered in the present study but should be in future studies.

Finally, the present results need independent validation, particularly to evaluate the effect of CLD etiology, which will require larger NAFLD and ALD populations.

Although remaining somewhat unknown among clinicians, CirrhoMeter has been validated in large independent populations^(9,14,22,31) and prognostic cohorts.^(29,32) It was the only blood test shown to assess fibrosis progression with sensitivity significantly higher than liver morphometry, considered the most sensitive pathologic technique.⁽³²⁾ Its synergic prognostication with FibroMeter has been shown⁽²⁹⁾ as has its accuracy for ruling large esophageal varices in or out.⁽³³⁾ It remains to be verified that all these advantages are retained by Multi-FibroMeter.

The Multi-FibroMeter calculation might appear somewhat complex, but once computerized it will exist only in abstraction for the clinician. Computerization is furthermore necessary to calculate multivariate fibrosis tests and potentially enable other digitized steps, such as reliability analyses.^(27,34) The new Multi-FibroMeter test is like an automated multiple-speed gearbox that improves the efficiency of what biomarkers can do and makes them better adapted to each clinical condition. Thus, Multi-FibroMeters are better adapted to the patient fibrosis stage.

This new type of test brings three advantages to the clinic: improved cirrhosis diagnosis, an important clinical target, without the limitations inherent to VCTE (failure, unreliable results); improved global staging, translated into a fibrosis classification appreciated by clinicians; and improved reproducibility. Thus, Multi-FibroMeter had the best reproducibility over time, an important element for determining the natural progression of fibrosis.

The multitargeted test induces no additional cost compared to the corresponding single-targeted test because the tests share the same biomarkers. Importantly, the present diagnostic method can be applied to any noninvasive diagnostic test based on a semiquantitative (ordinal) reference, e.g., a severity score in radiology.

In conclusion, the present study demonstrates that the construction of noninvasive tests of liver fibrosis can be improved by multitargeting the biomarkers. This approach provides a new blood test with overall accuracy superior to classical single-targeted blood tests or VCTE.

Acknowledgment: We thank all the participants in the included studies, particularly those of the SNIFF 12,

17, and 87 studies, the VINDIAG 7 study, and the ANRS HCEP 23 Fibrostar Group, coordinated by Jean Pierre Zarski and Vincent Leroy, and Victor de Ledinghen for the Pessac center. We also thank Sandra Girre, Pascal Veillon, Marc de Saint Loup, and Audrey Morrisset for data recording; Julien Chaigneau for morphometric data; and Kevin L. Erwin for writing assistance (English proofreading).

REFERENCES

- 1) European Association for the Study of the Liver, Asociacion Latinoamericana para el Estudio del Hgado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015;63:237-264.
- 2) Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, Aube C, et al. Noninvasive diagnosis of hepatic fibrosis or cirrhosis. *Gastroenterology* 1997;113:1609-1616.
- 3) Chou R, Wasson N. Blood tests to diagnose fibrosis or cirrhosis in patients with chronic hepatitis C virus infection: a systematic review. *Ann Intern Med* 2013;158:807-820.
- 4) Boursier J, Bacq Y, Halfon P, Leroy V, de Ledinghen V, de Muret A, et al. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009;21:28-38.
- 5) Cales P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A, et al. A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005;42:1373-1381.
- 6) Cales P, de Ledinghen V, Halfon P, Bacq Y, Leroy V, Boursier J, et al. Evaluating the accuracy and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C. *Liver Int* 2008;28:1352-1362.
- 7) Boursier J, de Ledinghen V, Zarski JP, Fouchard-Hubert I, Gallois Y, Oberti F, et al. multicentric groups from SNIFF 32, VINDIAG 7, and ANRS/HCEP23 FIBROSTAR studies. Comparison of eight diagnostic algorithms for liver fibrosis in hepatitis C: new algorithms are more precise and entirely noninvasive. *Hepatology* 2012;55:58-67.
- 8) Boursier J, de Ledinghen V, Zarski JP, Rousselet MC, Sturm N, Foucher J, et al. A new combination of blood test and fibroscan for accurate non-invasive diagnosis of liver fibrosis stages in chronic hepatitis C. *Am J Gastroenterol* 2011;106:1255-1263.
- 9) Ducancelle A, Leroy V, Vergniol J, Sturm N, Le Bail B, Zarski J, et al. A single test combining blood markers and elastography is more accurate than other fibrosis tests in the main causes of chronic liver diseases. *J Clin Gastroenterol* 2017;51:639-649.
- 10) Leroy V, Sturm N, Faure P, Trocme C, Marlu A, Hilleret MN, et al. Prospective evaluation of FibroTest(R), FibroMeter(R), and HepaScore(R) for staging liver fibrosis in chronic hepatitis B: comparison with hepatitis C. *J Hepatol* 2014;61:28-34.
- 11) Calès P, Halfon P, Batisse D, Carrat F, Perré P, Penaranda G, et al. Comparison of liver fibrosis blood tests developed for HCV with new specific tests in HIV/HCV co-infection *J Hepatol* 2010;53:238-244.
- 12) Cassinotto C, Boursier J, de Ledinghen V, Lebigot J, Lapuyade B, Cales P, et al. Liver stiffness in nonalcoholic fatty liver disease: a comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. *Hepatology* 2016;63:1817-1827.

- 13) [No authors listed]. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology* 1994;20(Pt. 1):15-20.
- 14) Zarski JP, Sturm N, Guechot J, Paris A, Zafrani ES, Asselah T, et al.; ANRS HCEP 23 Fibrostar Group. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: the ANRS HCEP-23 study. *J Hepatol* 2012;56:55-62.
- 15) Sandrini J, Boursier J, Chaigneau J, Sturm N, Zarski JP, Le Bail B, et al. Quantification of portal-bridging fibrosis area more accurately reflects fibrosis stage and liver stiffness than whole fibrosis or perisinusoidal fibrosis areas in chronic hepatitis C. *Mod Pathol* 2014;27:1035-1045.
- 16) Cales P, Boursier J, Bertrais S, Oberti F, Gallois Y, Fouchard-Hubert I, et al.; multicentric groups (SNIFF 14 & 17, ANRS HC EP 23 Fibrostar). Optimization and robustness of blood tests for liver fibrosis and cirrhosis. *Clin Biochem* 2010;43:1315-1322.
- 17) Cales P, Boursier J, Oberti F, Hubert I, Gallois Y, Rousselet MC, et al. FibroMeters: a family of blood tests for liver fibrosis. *Gastroenterol Clin Biol* 2008;32(Suppl. 1):40-51.
- 18) Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al.; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- 19) Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
- 20) Boursier J, Bertrais S, Oberti F, Gallois Y, Fouchard-Hubert I, Rousselet MC, et al. Comparison of accuracy of fibrosis degree classifications by liver biopsy and non-invasive tests in chronic hepatitis C. *BMC Gastroenterol* 2011;11:132.
- 21) Leroy V, Halfon P, Bacq Y, Boursier J, Rousselet MC, Bourliere M, et al. Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008;41:1368-1376.
- 22) Cales P, Boursier J, Oberti F, Bardou D, Zarski JP, de Ledinghen V.; Multicentric Group. Cirrhosis diagnosis and liver fibrosis staging: transient elastometry versus cirrhosis blood test. *J Clin Gastroenterol* 2015;49:512-519.
- 23) Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008;54:1372-1378.
- 24) Thein HH, Yi Q, Dore GJ, Krahn MD. Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression. *Hepatology* 2008;48:418-431.
- 25) Boursier J, de Ledinghen V, Poynard T, Guechot J, Carrat F, Leroy V, et al.; ARDENT group; AFEF. An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. *J Hepatol* 2015;62:807-815.
- 26) Cales P, Boursier J, Lebigot J, de Ledinghen V, Aube C, Hubert I, et al. Liver fibrosis diagnosis by blood test and elastography in chronic hepatitis C: agreement or combination? *Aliment Pharmacol Ther* 2017;45:991-1013.
- 27) Cales P, Boursier J, Ducancelle A, Oberti F, Hubert I, Hunault G, et al.; ANRS HC EP 23 Fibrostar Study. Improved fibrosis staging by elastometry and blood test in chronic hepatitis C. *Liver Int* 2014;34:907-917.
- 28) Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, 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.
- 29) Boursier J, Brochard C, Bertrais S, Michalak S, Gallois Y, Fouchard-Hubert I, et al. Combination of blood tests for significant fibrosis and cirrhosis improves the assessment of liver-prognosis in chronic hepatitis C. *Aliment Pharmacol Ther* 2014;40:178-188.
- 30) Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11:iii, ix-51.
- 31) Degos F, Perez P, Roche B, Mahmoudi A, Asselineau J, Voitot H, et al.; FIBROSTIC study group. Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study). *J Hepatol* 2010;53:1013-1021.
- 32) Cales P, Zarski JP, Chapplain JM, Bertrais S, Sturm N, Michelet C, et al.; multicentric groups (Fibrosys, ANRS HC 12, Fibrostop). Fibrosis progression under maintenance interferon in hepatitis C is better detected by blood test than liver morphometry. *J Viral Hepat* 2012;19:e143-e153.
- 33) Cales P, Sacher-Huvelin S, Valla D, Bureau C, Olivier A, Oberti F, et al.; multicenter group VO-VCO. Large esophageal varice screening by a sequential algorithm using a cirrhosis blood test and optionally capsule endoscopy. *Liver Int* 2018;38:84-93.
- 34) Boursier J, Zarski JP, de Ledinghen V, Rousselet MC, Sturm N, Le Bail B, et al.; Multicentric Group from ANRS/HC/EP23 FIBROSTAR Studies. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013;57:1182-1191.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1161/full.