AP&T Alimentary Pharmacology and Therapeutics

Diagnostic performance of FibroTest, SteatoTest and ActiTest in patients with NAFLD using the SAF score as histological reference

M. Munteanu*, D. Tiniakos[†,‡], Q. Anstee[†], F. Charlotte[§], G. Marchesini[¶], E. Bugianesi[**], M. Trauner[††], M. Romero Gomez[‡‡], C. Oliveira[§§], C. Day[†], J.-F. Dufour[¶¶], S. Bellentani[***], Y. Ngo*, S. Traussnig[††], H. Perazzo[†††], O. Deckmyn*, P. Bedossa[‡‡‡], V. Ratziu[§], T. Poynard[§] & the FLIP Consortium and the FibroFrance Group[a]

*BioPredictive, Paris, France.
[†]Liver Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle-upon-Tyne, UK.
[‡]Laboratory of Histology & Embryology, Medical School, National & Kapodistrian University of Athens, Greece.
[§]Groupe Hospitalier Pitié Salpêtrière APHP, Sorbonne Universités, UPMC Univ Paris 06, INSERM, UMR_S 938 & Institute of Cardiometabolism and Nutrition (ICAN), Paris, France.
[¶]Universita di Bologna, Bologna, Italy.
[**]Universita Degli Studi di Torino, Torino, Italy.
[††]Medizinischen Universitaet Wien, Vienna, Austria.
[‡‡]Valme University Hospital, University of Seville, Sevilla, Spain.
[§§]Department of Gastroenterology (LIM-07), University of São Paulo School of Medicine, São Paulo, Brazil.
[¶¶]Université de Berne, Berne, Switzerland.
[***]Gastroenterologia, Azienda USL di Modena Reggio Emilia, Modena, Italy.
[†††]APHP UPMC Liver Center, Paris, France.
[‡‡‡]Assistance Publique-Hôpitaux de Paris, hôpital Beaujon, University Paris-Diderot, Paris, France.

Correspondence to:
Prof. T. Poynard, Hepato-Gastroenterology Department, Groupe Hospitalier Pitié Salpêtrière, Paris 75013, France.
E-mail: thierry@poynard.com

[a]Co-authors members of the FLIP Consortium and the FibroFrance Group are listed in Appendix 1.

Publication data
Submitted 17 June 2016
First decision 19 July 2016
Resubmitted 26 July 2016
Accepted 27 July 2016
EV Pub Online 23 August 2016

The Handling Editor for this article was Professor Peter Hayes, and it was accepted for publication after full peer-review.

SUMMARY

Background
Blood tests of liver injury are less well validated in non-alcoholic fatty liver disease (NAFLD) than in patients with chronic viral hepatitis.

Aims
To improve the validation of three blood tests used in NAFLD patients, FibroTest for fibrosis staging, SteatoTest for steatosis grading and ActiTest for inflammation activity grading.

Methods
We pre-included new NAFLD patients with biopsy and blood tests from a single-centre cohort (FibroFrance) and from the multicentre FLIP consortium. Contemporaneous biopsies were blindly assessed using the new steatosis, activity and fibrosis (SAF) score, which provides a reliable and reproducible diagnosis and grading/staging of the three elementary features of NAFLD (steatosis, inflammatory activity) and fibrosis with reduced interobserver variability. We used nonbinary-ROC (NonBinAUROC) as the main endpoint to prevent spectrum effect and multiple testing.

Results
A total of 600 patients with reliable tests and biopsies were included. The mean NonBinAUROCs (95% CI) of tests were all significant ($P < 0.0001$): 0.878 (0.864–0.892) for FibroTest and fibrosis stages, 0.846 (0.830–0.862) for ActiTest and activity grades, and 0.822 (0.804–0.840) for SteatoTest and steatosis grades. FibroTest had a higher NonBinAUROC than BARD (0.836; 0.820–0.852; $P = 0.0001$), FIB4 (0.845; 0.829–0.861; $P = 0.007$) but not significantly different than the NAFLD score (0.866; 0.850–0.882; $P = 0.26$). FibroTest had a significant difference in median values between adjacent stage F2 and stage F1 contrarily to BARD, FIB4 and NAFLD scores (Bonferroni test $P < 0.05$).

Conclusions
In patients with NAFLD, SteatoTest, ActiTest and FibroTest are non-invasive tests that offer an alternative to biopsy, and they correlate with the simple grading/staging of the SAF scoring system across the three elementary features of NAFLD: steatosis, inflammatory activity and fibrosis.

doi:10.1111/apt.13770

INTRODUCTION

As stated in recent guidelines and overviews, serum biomarkers of liver injury have been less well validated in non-alcoholic fatty liver disease (NAFLD) than in chronic viral hepatitis.[1, 2] One reason was that the histological systems previously used as references when assessing biomarker performance in NAFLD were less consensual, mixing different elementary histological features, than with chronic viral hepatitis in which fibrosis, activity and steatosis were separately scored.[3–5] Examples of such dependence between elementary features in NAFLD diagnosis were the exclusion of non-alcoholic steato-hepatitis (NASH) in cases with activity but without steatosis, or the diagnostic of NASH in cases with steatosis and fibrosis but without activity.[3, 4]

In viral hepatitis, the METAVIR (META-analysis of histological data in VIRal hepatitis) scoring system, which has been widely validated since 1996, is used to independently classify fibrosis based on five classes, and the necro-inflammatory histological activity grade based on four classes.[5] Since 2012, the SAF score (Steatosis, inflammatory Activity and Fibrosis) has provided reliable and reproducible diagnoses and grading/staging of the three elementary features of NAFLD with reduced inter-observer variability.[3, 4]

The diagnostic performance of specific blood tests of SAF in NAFLD has previously been validated using a range of nonconsensual scoring systems for NAFLD: six studies for FibroTest (Nash-FibroSure in USA) using the METAVIR fibrosis score,[6–11] three studies for ActiTest using the METAVIR activity score [7, 10, 12] and three studies for SteatoTest using the Goodman steatosis score.[9, 10, 13] FibroTest and SteatoTest have been already used as secondary efficacy endpoints in clinical trials of patients with presumed NAFLD or NASH.[14–16] In this report, we aimed to improve the validation of these three blood tests in patients with NAFLD using the new SAF scoring system as the histological reference.

PATIENTS AND METHODS

We included a large group of new patients from the ongoing FibroFrance project (USA-NCT01927133), and the prospective population of the FLIP consortium (http://www.flip-fp7.eu/). Instead of the binary area under the operating characteristic curves (AUROC), the nonbinary area under the operating characteristic curves (NonBinAUROC or Obuchowski measure) was the primary endpoint for assessing the nonbinary performances of these tests and as recommended for preventing the spectrum effect and multiple testing.[17–19]

The FLIP project (Fatty Liver: Inhibition of Progression EU-241762) was initiated in 2010. Two specific aims of the project were the validation of improved diagnostic and prognostic markers, and the validation of a consensual histological classification of NAFLD and NASH. The project was based on a large prospective European cohort of patients with histologically diagnosed NAFLD, with standardised inclusion criteria and histologically proven NAFLD, with and without NASH. (http://www.flip-fp7.eu/) Written patient consent for liver biopsy and data collection was obtained from each subject prior to inclusion.

The FibroFrance project was initiated in 1996. The aim of this project was to assess the natural history of liver fibrosis in chronic liver diseases and the impact of treatments.[20–23] This epidemiological, non-interventional study was exempt from IRB review (Ethical committee of 'Comité de Protection des Personnes of Paris- Ile-de-France', FIBROFRANCE project. CPP-IDF-VI, 10-1996-DR-964, DR-2012-222 and USA-NCT01927133). No consent was given, as all data were analysed anonymously. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki. All authors had access to the study data and reviewed and approved the final manuscript.

Patients

The inclusion criteria in the prospective FLIP cohort were adult patients at risk of NAFLD, i.e. presenting with ultrasound defined steatosis, and/or increased liver function test values, and/or metabolic risk factors (overweight, visceral adiposity, type 2 diabetes, arterial hypertension, dyslipidemia) without well identified chronic liver diseases including alcohol consumption of <50 g/day. All patients had to have had a liver biopsy. The nine participating centres were located in Paris, Seville, Newcastle, Bologna, Turin, Modena, Sao Paulo, Bern and Vienna. A central e-CRF was created, and data quality was enforced by a central data manager. The inclusion criteria in this retrospective analysis were the presence of reliable FibroTest, ActiTest and SteatoTest (FibroMax package), as well as the reading of the biopsy using the SAF scoring system by one of the experts from the FLIP Pathology Consortium. The validation of the patented FibroMax package was pre-determined in the FLIP protocol aims before patient inclusion (http://www.flip-fp7.eu/).

The inclusion criteria of the FibroFrance-NAFLD population were the same as in FLIP cohort, and with measurements of SteatoTest, ActiTest and FibroTest (USA-NCT01927133). The criterion for retrospective inclusion in the FibroFrance-FibroMax subpopulation

was the reading of a liver biopsy using the SAF scoring system by one of the experts from the FLIP Pathology Consortium. Patients from both cohorts were excluded if the blood test was disqualified according to the company recommendations for reliable tests,[24] or if the interval between the biopsy and blood tests was greater than 180 days. Patients receiving specific treatment for NAFLD before biopsy, were not included.

Histological references

The SAF scoring system, specific for NAFLD features, has been described elsewhere.[3, 4] The goal of the SAF test was to find a compromise between the development of a simple, easily applied system for making a firm diagnosis in individual patients, even when applied by nonspecialists, and of a more reliable and discriminating system for therapeutic trials or for the assessment of biomarker diagnostic performance. A FLIP histopathology consortium of eight members developed the FLIP algorithm, a diagnostic tool for the diagnosis and staging of severe forms of NAFLD.[3, 4] According to the combination of each semi-quantification of the three elementary features of NAFLD using the SAF score for steatosis, inflammatory activity and fibrosis respectively. The steatosis score (S) assesses the quantities of large or medium-sized lipid droplets, with the exception of foamy microvesicles, and rates them from 0 to 3 (S0: <5%; S1: 5–33%, mild; S2: 34–66%, moderate; S3: >66%, marked). Activity grade (A, from 0 to 4) is the unweighted addition of hepatocyte ballooning (0–2) and lobular inflammation (0–2). Cases with A0 ($A = 0$) had no activity; A1 ($A = 1$) had mild activity; A2 ($A = 2$) moderate activity; A3 ($A = 3$) severe activity and A4 ($A = 4$) had very severe activity. Fibrosis stage (F) was assessed using the score described by[25] as follows: stage 0 (F0) = none; stage 1 (F1) = 1a or 1b perisinusoidal zone 3 or 1c portal fibrosis; stage 2 (F2) = perisinusoidal and periportal fibrosis without bridging; stage 3 (F3) = bridging fibrosis and stage 4 (F4) = cirrhosis (File S1).

To reduce interobserver variability and homogenise the reading using the new SAF-FLIP histological classification, we used only reports reviewed by members of the FLIP Pathology Consortium (DT and PB for the FLIP subpopulation and FC for the FibroFrance subpopulation).

Blood tests

FibroTest, ActiTest and SteatoTest were patented as the 'In Vitro Diagnostic Multivariate Index Assay' for the diagnosis of METAVIR fibrosis stages, including cirrhosis, for METAVIR activity grades and for SAF-equivalent steatosis grades.[10, 22–24] FibroTest included serum a2-macroglobulin, apolipoprotein A1, haptoglobin, total bilirubin and gamma-glutamyltranspeptidase, adjusted for age and gender. ActiTest included the same components plus alanine-aminotransferase (ALT). SteatoTest included the same six components of FibroTest-ActiTest plus body mass index, serum cholesterol, triglycerides and glucose, adjusted for age and gender.[13] These tests are exclusively available online, including security algorithms. Modelling of fibrosis progression or regression and prognostic performances were similar when FibroTest was compared to liver biopsy, regardless of the cause of liver disease,[8, 22, 26–30] including the same limitations for discriminating between intermediate stages of fibrosis.[31] The recommended cutoffs were the same whatever the chronic liver diseases (File S1).

The following three blood tests were used as comparators for FibroTest, as recommended in the NAFLD guidelines[3]: (i) NAFLD Fibrosis Score (NFS) = $(-1.675 + 0.037 \times \text{age (year)} + 0.094 \times \text{BMI (kg/m}^2) + 1.13 \times \text{IFG/diabetes (yes = 1, no = 0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelet count } (\times 10^9/\text{L}) - 0.66 \times \text{albumin [g/dL]})$; (ii) BARD score (BMI ≥28 = 1; AST/ALT ratio ≥0.8 = 2; diabetes = 1; score ≥2, odds ratio for advanced fibrosis = 17); and (iii) FIB4 score = FIB-4 = age (year) × AST [U/L]/(platelets [10^9/L] × (ALT [U/L])). No guidelines proposed recognised blood tests as comparators of SteatoTest for steatosis grades and of ActiTest for activity grades.[1]

Statistical analysis

The protocol and the analyses followed the FibroSTARD recommendations, which are detailed in File S2.[19] In summary, we checked that the study populations had not been previously used and published for the evaluation of the studied tests. In the FLIP group, all the cases were prospective, and the tests, as well as the contemporaneous biopsies, had not been previously published. In the FibroFrance cases, no patients had been published in previous validations of FibroTest, ActiTest or SteatoTest in NAFLD.[6, 10, 12, 13]

The primary endpoint for the diagnostic performance for each quantitative test, for the diagnostic of each histological class (SAF scoring system), was the NonBinAUROC: FibroTest for the five SAF ordinal stages of fibrosis (F0–F4), ActiTest for the five ordinal SAF grades (A0–A4) and SteatoTest, for the four ordinal grades (S0–S3) (File S1). The penalty function was related to the number of classes difference: for five classes, 0.25 for one class difference (adjacent classes), 0.50 for two for and

0.75 for three. We also graphically represented the medians and the interquartile distribution of all tests according to histological scores. The medians were compared using the Kruskal–Wallis multiple comparisons *Z*-Value test, with the Bonferroni correction due to the number of comparisons.

The secondary endpoint was the comparisons between fibrosis blood tests. The performances of FibroTest and ActiTest were compared to those of NAFLD score, BARD and FIB4, also used for assessing the severity of fibrosis and activity in subjects with presumed NAFLD.

Sensitivity analyses of the primary endpoint (NonBinAUROC) used stratifications according to the following known factors of variability in the tests' performances: biopsy length (≤10 mm, 10 mm to <20 mm and ≥20 mm),[32] morbid obesity (>35 kg/m^2)[10] and presence of diabetes (fasting glucose ≥7.0 μmol). We also compared the NonBinAUROC between the FibroFrance-NAFLD cohort and the non-French FLIP cohort.

The statistical softwares NCSS-2013 [33] and R-nonBinROC were used.[34]

RESULTS

From March 2005 to December 2014, a total of 956 patients with suspected NAFLD were pre-included, and 600 patients were included after exclusion of 356 patients due to the absence of blood tests (n = 305) or of biopsy as assessed by the SAF score (n = 14) or with more than 180 days between biopsy and tests (n = 44) (Figure 1). The characteristics of included and excluded patients were very similar, with a median age of 50 years, 60% male, BMI 30 kg/m^2, 23% type 2 diabetics, and 5% histological cirrhosis, despite a lower prevalence of stage F0 in included (20.3%) than in excluded patients (29.9%) (Table 1).

Primary endpoint (Table 2)

The mean (95% CI) NonBinAUROC of tests were all significant (P < 0.0001): 0.878 (0.864–0.892) for FibroTest and the prediction of five SAF fibrosis stages, 0.846 (0.830–0.862) for ActiTest and five SAF-activity grades, and 0.822 (0.804–0.840) for SteatoTest and four SAF steatosis grades. The highest performances between adjacent stages/grades were observed for F3 vs. F4, A3 vs. A4 and S0 vs. S1; the lowest performances were observed for F0 vs. F1, A2 vs. A3 and S1 vs. S2 (File S3).

Comparison between fibrosis tests (Table 3)

In 574 patients with all blood tests, for the prediction of fibrosis stages, FibroTest had a higher NonBinAUROC (0.877; 0.862–0.892) than BARD (0.836; 0.820–0.852; P = 0.0001), FIB4 (0.845; 0.829–0.861; P = 0.007) but not significantly higher than NAFLD score (0.866; 0.850–0.882; P = 0.26).

ActiTest for the prediction of necro-inflammatory activity grades, had a higher NonBinAUROC (0.846; 0.830–0.862) than BARD (0.810; 0.792–0.828; P = 0.0003), FIB4 (0.798; 0.780–0.816; P < 0.0001) and than NAFLD score (0.815; 0.805–0.825; P = 0.005).

Box plots of test values according to biopsy scores are given in Figure 2a for FibroTest. The median FibroTest values increased (P < 0001) steadily with fibrosis stages after F1, ranging from 0.18 in F0, 0.21 in F1, 0.28 in F2, 0.41 in F3 and 0.71 in cirrhosis (Table S4B in File S4). Only FibroTest had a significant difference in median values (Bonferroni test) between adjacent stage F2 and stage F1 (Table S4A in File S4) when compared to FIB4 (Figure S4B and Table S4B in File S4), to BARD (Figure S4C and Table S4C in File S4) and to NAFLD score (Figure S4D and Table S4D in File S4).

ActiTest and SAF-activity score

The median value of ActiTest increased steadily (P < 0.001) from 0.21 for A0, 0.28 for A1, 0.35 for A2 and 0.38 for A3 and 0.46 for the last grade of SAF scoring system (Figure 2b, Table S5A in File S5); differences however were only significant between 2 grades using the multiple comparisons rules. The ActiTest median values increased (P < 0.001) for the two features of the SAF-activity score, ballooning and lobular inflammation. For ballooning, the value was 0.22 if absence (n = 116), 0.36 if moderate (n = 253) and 0.39 if severe (n = 231), with a significant difference between grades 0 and 1 (Figure S5B and Table S5B in File S5). For lobular inflammation, the value was 0.23 if absence (n = 118), 0.35 if moderate (n = 331), and 0.43 if severe (n = 151), all with significant differences between adjacent stages (Figure S5C and Table S5C in File S5).

SteatoTest and SAF steatosis score

Only 20 patients had an S0 grade, which was expected due to the inclusion criteria of NAFLD patients. The median value of SteatoTest increased (P = 0.002) from 0.52 for S0 (n = 20), 0.62 for S1 (n = 188), 0.66 for S2 (n = 228), to 0.71 for S3 (n = 163), marked steatosis the last grade. The only significant difference observed between grades was between S3 and S1 (Figure 2c, Figure S6A and Table S6A in File S6).

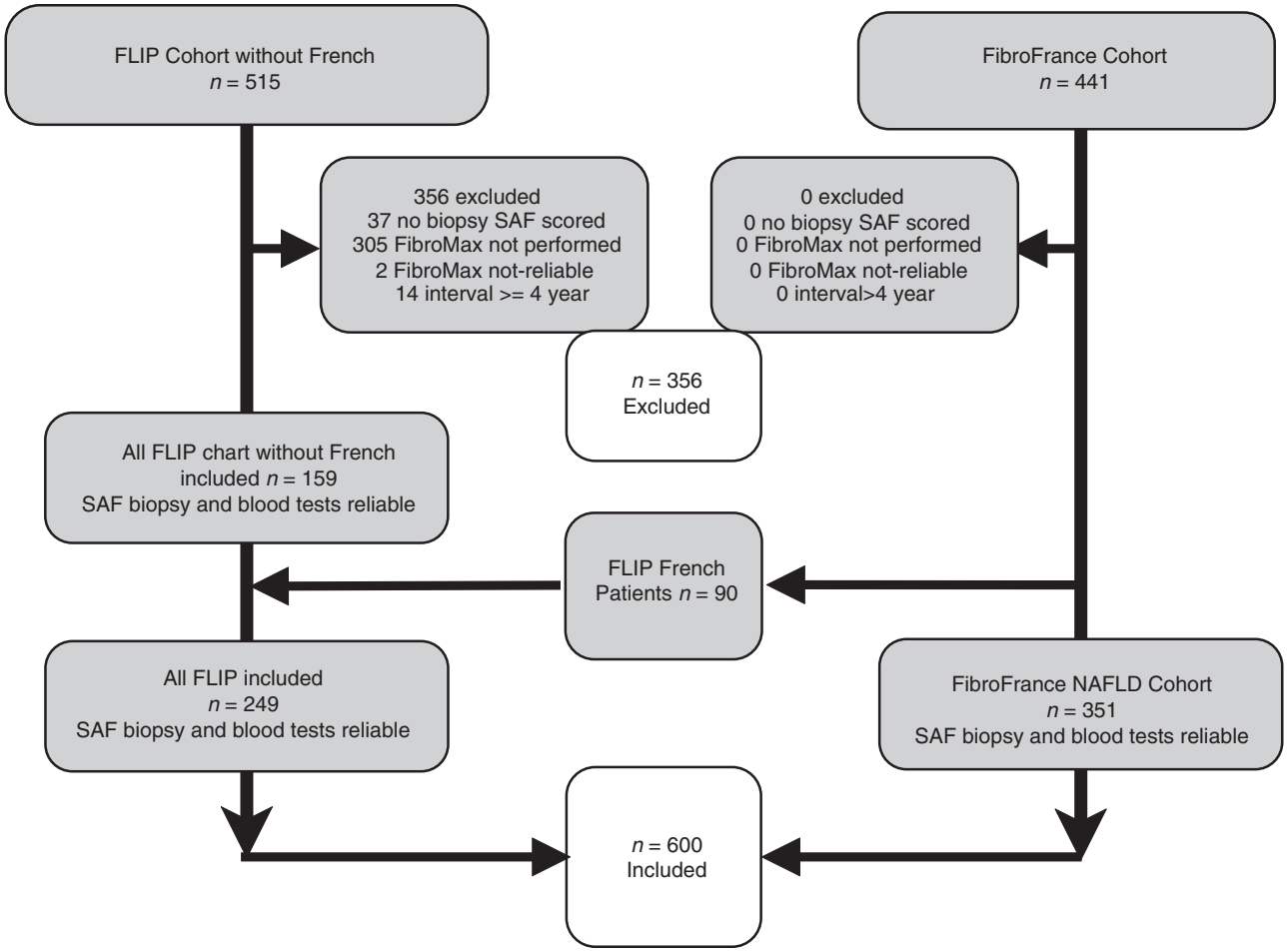

Figure 1 | Flow charts of patient inclusions.

Sensitivity analyses

File S7 shows the sensitivity analyses of the test performances on the primary outcome (NonBinAUROC) according to gender, cohorts, biopsy length, diabetes and severe obesity. There were no significant differences between the NonBinAUROC stratified according the variability factors for FibroTest (Table S7A in File S7), ActiTest (Table S7B in File S7) or SteatoTest (Table S7C in File S7).

DISCUSSION

The results of this large sample of NAFLD patients confirmed the significant performance previously observed for FibroTest,[6–11] ActiTest[7, 10, 12] and SteatoTest.[9, 10, 13] The 2015 EASL guidelines had not yet reviewed the recent validations by external researchers ('independent of inventor's team') for the diagnosis of NAFLD features, including FibroTest or ActiTest, or SteatoTest.[8, 9, 11] Despite many advantages of these tests, as in patients with chronic viral hepatitis, several limitations must be acknowledged.

Limitations

The patients included for possible NAFLD were mostly those from tertiary centres whose selection was based on abnormal tests, which may not be representative of NAFLD in the general population. Indeed, the spectra of fibrosis stages, activity and steatosis grades were much more severe in this study than those of NAFLD (n = 3969) as screened in the French general population[27]: 6.2% cirrhosis as presumed by FibroTest vs. 0.4% (0.2–0.6%); 20.0% severe activity (grades A3 and A4) as presumed by ActiTest vs. 0.81% (32/3969); and 48.3% moderate/marked steatosis (S2/S3) as presumed by SteatoTest vs. 13.4% (525/3934) respectively. The mean BMI of 30 kg/m^2 indicated a high prevalence of obese with a low prevalence of fatty liver occurring in slightly overweight individuals. Median age was 50 years, again

Table 1 | Characteristics of included and excluded patients

Characteristics	Included (n = 600)	Excluded (n = 356)	P-value
FibroFrance NAFLD	441/600 (73.5%)	356/356 (100%)	
Non-French FLIP	159/600 (26.6%)	0/355 (0.0%)	<0.0001
Gender male	380/600 (63.3%)	218/356 (61.4%)	0.52
Diabetes type 2	136/600 (22.7%)	81/347 (23.3%)	0.87
Age (years)	53.2 (51.4–54.4)	51.3 (49.7–53.0)	0.04
BMI (weight/height^2)	29.7 (29.2–30.2)	29.9 (29.2–30.7)	0.62
Biopsy length (mm)*	25 (22–25)	15 (15–18)	0.0001
Number fragments†	2 (1–2)	1 (1–1)	0.0001
Number portal tracts‡	16 (15–17)	NA	NA
Interval between biopsy and test (days)	0 (0–0.015)	0 (0–4.1)	0.0001
SAF F biopsy			
F0 no fibrosis	122/600 (20.3%)	102/342 (29.9%)	0.02
F1 perisinusoidal or portal	184/600 (30.8%)	83/342 (24.0%)	
F2 sinusoidal or periportal without bridging	140/600 (23.3%)	77/342 (22.6%)	
F3 bridging fibrosis	121/600 (20.2%)	61/342 (17.9%)	
F4 cirrhosis	33/600 (5.5%)	19/342 (5.6%)	
Grade of activity (SAF A biopsy)			
A0 no activity	64/600 (10.7%)	32/320 (10.0%)	0.09
A1 mild	86/600 (14.3%)	65/320 (20.3%)	
A2 moderate	191/600 (31.8%)	81/320 (25.3%)	
A3 severe	156/600 (26.0%)	81/320 (25.3%)	
A4 very severe	103/600 (17.2%)	81/320 (19.1%)	
SAF S biopsy			
S0 no steatosis <5%	20/600 (3.3%)	7/339) (2.1%)	0.33
S1 mild 5–33%	204/600 (34.0%)	115/339 (33.9%)	
S2 moderate 34–66%	229/600 (38.2%)	125/339 (36.9%)	
S3 marked >66%	147/600 (24.5%)	92/339 (27.1%)	
Presumed Fibrosis SAF stage (FibroTest range)			
F0 no fibrosis (0.00–0.27)	314/600 (52.3%)	23/50 (46.0%)	0.17
F1 (>0.27–0.48)	149/600 (24.8%)	12/50 (24.0%)	
F2 (>0.48–0.58)	39/600 (6.5%)	6/50 (12.0%)	
F3 (>0.48–0.74)	64/600 (10.7%)	3/50 (6.0%)	
F4 (>0.74–1.00)	34/600 (5.7%)	6/50 (12.0%)	
Presumed SAF-activity grade (ActiTest range)*			
A0 no activity (0.00–0.29)	262/600 (43.7%)	27/50 (54.0%)	0.04
A1 mild (>0.29–0.52)	185/600 (30.8%)	10/50 (20.4%)	
A2 moderate>0.52–0.62)	62/600 (10.3%)	3/50 (6.0%)	
A3 severe grades (>0.62–0.72)	33/600 (5.5%)	7/50 (14.0%)	
A4 very severe grades (>0.72–1.00)	58/600 (9.1%)	3/50 (6.0%)	
Presumed SAF Steatosis (SteatoTest range)			
S0 no steatosis 0–<5% (0.00–0.57)	228/600 (15.7%)	10/45 (6.7%)	0.10
S1 mild >5%–≤33% (>0.57–0.69)	116/600 (19.3%)	12/45 (26.7%)	
S2S3 > moderate-marked >33% (>0.69–1.00)	256/600 (42.7%)	23/45 (51.1%)	
Presumed steatosis including minimal grade			
S0 no steatosis 0% (0.00–0.38)	94/600 (15.7%)	3/45 (6.7%)	0.17
S1 minimal 1–5% (>0.38–0.57)	134/600 (22.3.0%)	7/45 (15.6%)	
S2 mild >5% to ≤33% (>0.57–0.69)	116/600 (19.3%)	12/45 (26.7%)	
S3S4 > moderate-marked >33% (>0.69–1.00)	256/600 (42.7%)	23/45 (51.1%)	

* Data available in 510 included and 325 excluded patients.

† Data available in 346 included and 0 excluded patients.

‡ Data available in 268 included and 188 excluded patients.

indicating a previous selection of patients. Thus individuals, who might have had NAFLD for a significant duration to have progressed, limiting the findings in relation to early detection of NAFLD. In addition, another limitation was the predominance of Caucasians. The advantages of these blood tests were the broad spectrum of

Table 2 | Performance (non binary AUROC) of FibroTest, ActiTest and SteatoTest for the prediction of histological SAF scores of fibrosis, necro-inflammatory activity and steatosis

n = 600 Blood test	Primary endpoint SAF score	Non binary AUROC Mean	 95% CI	Significance vs. 0.500 *P*-value
FibroTest	Fibrosis F0–F4	0.878	0.864–0.892	<0.0001
ActiTest	Activity A0–A4	0.846	0.830–0.862	<0.0001
SteatoTest	SteatoTest S0–S3	0.822	0.804–0.840	<0.0001

Table 3 | Comparison between FibroTest-ActiTest and BARD, FIB4 and NAFLD score for the diagnostic of histological fibrosis and activity estimated by SAF scoring system

n = 574 Blood test	Primary endpoint SAF score	Non binary AUROC Mean	 95% CI	Significance *P*-value vs. FibroTest
FibroTest	Fibrosis F0–F4	0.877	0.862–0.892	1
BARD	Fibrosis F0–F4	0.836	0.820–0.852	0.0001
FIB4	Fibrosis F0–F4	0.845	0.829–0.861	0.007
NAFLD score	Fibrosis F0–F4	0.866	0.850–0.882	0.26
Blood test	**SAF score**	**Mean**	**95% CI**	***P*-value vs. ActiTest**
ActiTest	Activity A0–A4	0.846	0.830–0.862	1
BARD	Activity A0–A4	0.810	0.792–0.828	0.0003
FIB4	Activity A0–A4	0.798	0.780–0.816	<0.0001
NAFLD score	Activity A0–A4	0.815	0.805–0.825	0.005

validations in a variety of populations, from controls, blood donors and cured chronic hepatitis C (CHC) to populations at high risk of NAFLD including diabetes, severe obesity and dyslipidemia.[13, 26, 30, 35–37]

FibroTest had limitations in NAFLD patients, as it was not initially designed to discriminate between the zonal distribution of fibrosis, which is different than the portal distribution of fibrosis in chronic viral hepatitis. However, the prognosis of fibrosis staged using the 'METAVIR fibrosis stages' based on few septa (F2) and many septa (F3) was validated for the most frequent chronic liver diseases,[21] including CHC,[38] chronic hepatitis B (CHB),[28] alcoholic liver disease[29] and also NAFLD.[30] It was expected that FibroTest had a higher performance for discriminating F3 vs. F4 than between the other adjacent pairwise comparisons. This was true also for liver biopsy,[31, 39] even that 25 mm in length, due to the greater differences in the area of fibrosis between F3 vs. F4 than between the other adjacent stages, as demonstrated when large surgical biopsies were used as a true 'gold standard'.[31] Relative to FibroTest, biopsy was affected even more in the discrimination between F1 vs. F2, as the inter- and intra-observer variability was greater than for other adjacent stage comparisons, such as F3 vs. F4.[31, 40] Despite the significant NonBinAUROC of FibroTest and similar estimates for cirrhosis (5.7% vs. 5.5%), we observed that the prevalence of presumed F2 (6.5%) and F3 (10.7%) fibrosis stages was lower than that observed using biopsy (23.3% and 20.2%), respectively. It is not necessarily correct to conclude that FibroTest underestimated these two stages. It could also be related to a lower quantity of fibrosis in perisinusoidal fibrosis than in periportal fibrosis. The F2 METAVIR is based on 'few septa' and F3 on 'many septa' vs. 'bridging' for the SAF fibrosis score F3. A standardisation, such as the area of collagen,[41] on large biopsies could be useful in determining whether FibroTest underestimated or the SAF score overestimated the 'severity' of fibrosis in NAFLD before proposing new cutoffs.

ActiTest could have limitations for NAFLD, as it was originally designed for grading the necrotico-inflammatory histological activity (four grades only) in patients with CHC or CHB. These patients presented more hepatocyte necrosis than NAFLD, with higher levels of ActiTest mostly related to higher ALT values and lower apoA1 values.[42] However, we confirmed here that ActiTest NonBinAUROC in NAFLD also increased very significantly with ballooning and intralobular infiltrates grades[7, 10] but

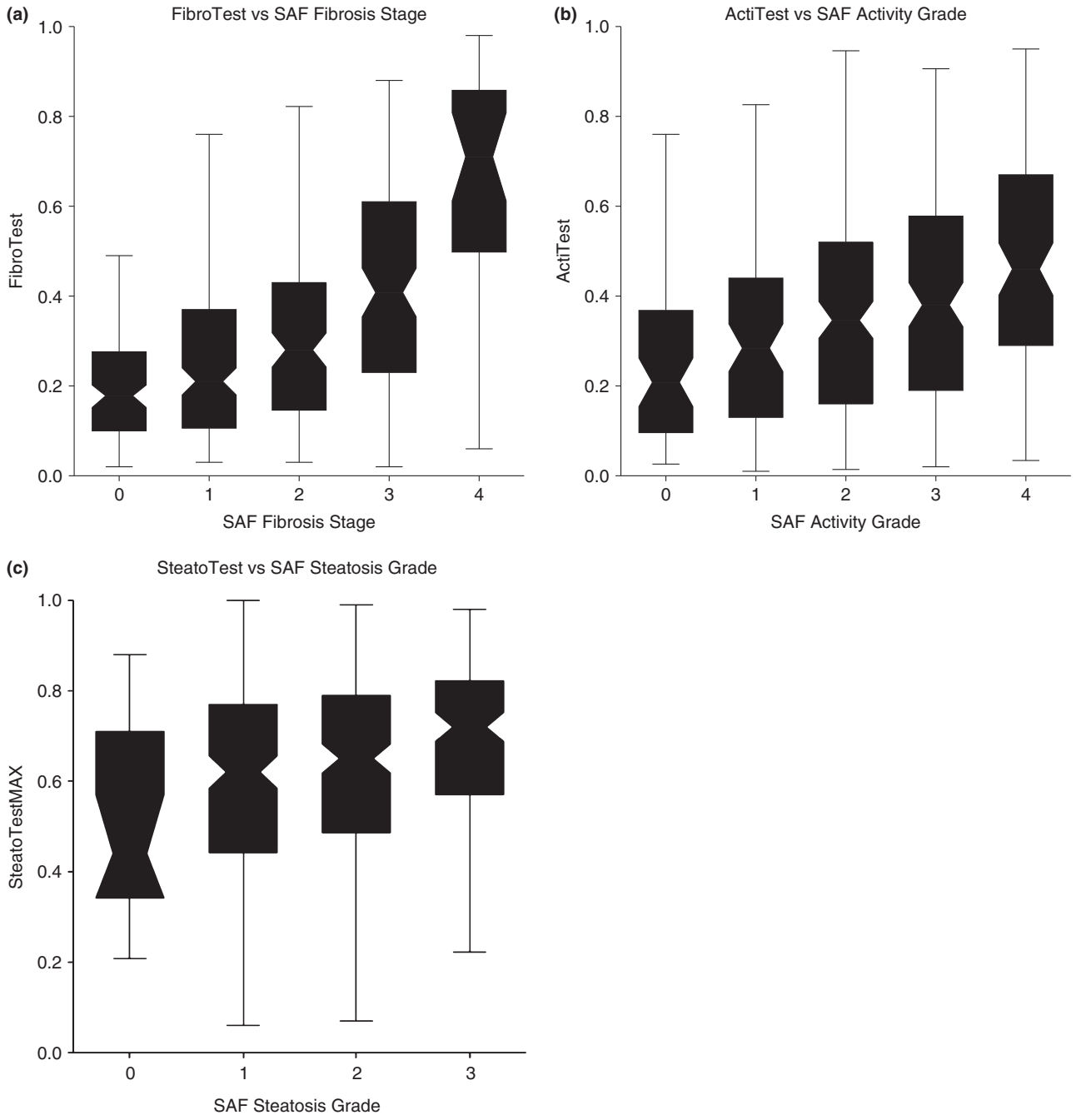

Figure 2 | Box plots of FibroTest (a), ActiTest (b) and SteatoTest (c) according to the respective SAF scoring systems. 600 cases were included. Significance between stages/grades were estimated using NonBinROC measures and detailed in Table 2. The median FibroTest values increased steadily (NonBinROC $P < 0.001$) with fibrosis stages after F1, ranging from 0.18 in F0 ($n = 122$), 0.21 in F1 ($n = 184$), 0.28 in F2 ($n = 140$), 0.41 in F3 ($n = 121$) and 0.71 in cirrhosis ($n = 33$), all differences between adjacent stages were significant (Bonferroni test $P < 0.05$). The median value of ActiTest increased steadily (NonBinROC $P < 0.001$) from 0.21 for A0 ($n = 64$), 0.28 for A1 ($n = 86$), 0.35 for A2 ($n = 191$) and 0.38 for A3 ($n = 156$) and 0.46 for A4 ($n = 103$). Only the differences between 2 grades were significant (Bonferroni test $P < 0.05$). The median value of SteatoTest increased (NonBinROC $P = 0.002$) from 0.52 for S0 ($n = 20$), 0.62 for S1 ($n = 188$), 0.66 for S2 ($n = 228$), to 0.71 for S3 ($n = 163$). The only significant difference was observed between grades was between S3 and S1 (Bonferroni test $P < 0.05$).

in lower ranges than in viral hepatitis.[42] The lower prevalence of presumed A2 (10.3%), A3 (5.5%) and A4 (9.1%) grades vs. those observed using biopsy (31.8%, 26.0% and 17.2%, respectively) suggested, therefore, that cutoffs could be modified. In these tertiary NAFLD populations with a low prevalence of A0 and A1, a simpler classification focusing on the most severe NASH features, as suggested by the FLIP algorithm, would be more appropriate.[4] However, in a primary general population, the sensitivity of ActiTest is an advantage for detecting early grades of ballooning or intralobular inflammation.

SteatoTest had similar limitations. It was originally designed to quantify steatosis using four grades, up to 66% hepatocytes, and with a high sensitivity, as validated with an S1 category defined as fewer than 5% hepatocytes with steatosis. SteatoTest was not designed to discriminate between severe steatosis (greater than 66%) vs. marked steatosis (between 33% and 66%). Therefore, due to the spectrum effect, it was expected that the NonBinAUROC would be low in a tertiary centre population of NAFLD, as steatosis was suspected in almost all patients. Indeed, more than 5% steatosis was observed at biopsy in 94% of patients, with only 20 patients classified as S0. The potential advantages of SteatoTest are its ability to identify subjects with steatosis from those without steatosis in the general population[27] or in patients with chronic viral hepatitis C,[20] and its prognostic value in patients with metabolic disorders.[30]

This study focused on three tests developed by several co-authors of the article, who have an obvious conflict of interest as inventor or employee of the company marketing these tests. However, the other co-authors were totally independent, and they recruited the patients and performed the assay independent of the company and had full access to all data and analyses.

We compared directly FibroTest and ActiTest to popular nonpatented NITs, BARD, FIB4 and NAFLD score, but not several others, such as APRI, fatty liver index, ELF, Fibrospect, Fibrometer and transient elastography (TE) for fibrosis, cytokeratin 18, ultrasonography, magnetic resonance imaging and spectroscopy for steatosis.[1] Other components with different rational such as adiponectin, a hepatoprotective adipocytokine with insulin sensitising properties could also be interesting for the diagnosis of NAFLD features.[41] A recent retrospective comparison of nine NITs in 452 cases with presumed NAFLD, used the biopsy NASH-CRN scoring system as reference.[11] Similar NonBinAUROC were observed, 0.698, 0.722, 0.730, 0.748, 0.763, for BARD, FibroTest, NAFLD score, FIB4 and FibroMeter^{NAFLD}, respectively. Only the Fibrometer designed for viral hepatitis had a small increase in NonBinAUROC vs. the 8 other NITs ($P = 0.04$). TE had 14% of failure and the NonBinROC was not assessed in intention to diagnose.

Here, FibroTest had significantly higher NonBinAUROC than BARD and FIB4 scores for fibrosis prediction. Furthermore, FibroTest had significantly higher median values between adjacent stage F2 and stage F1, contrarily to BARD, FIB4 and NAFLD score (Bonferroni test $P < 0.05$).

We acknowledge that NAFLD score, a noncommercial inexpensive test, despite lower performance for discriminating stage-F2 vs. F1, was not inferior to FibroTest for the main endpoint (NonBinAUROC).

These differences with other direct comparisons[11] could be due to variability in the histological endpoint, SAF being more reproducible than CREN scoring system, but also to differences in fibrosis spectrum between studies. The obvious advantages of nonpatented NITs were their lower price than patented NITs, but their applicability as well as their risk of false positives and false negatives have been less investigated than for FibroTest.[1]

One advantage of FibroTest over the other NITs, including TE,[43] is its assessment together with ActiTest a validated marker of necro-inflammatory histological activity.[44] Therefore, in patients with presumed NAFLD and contrarily to any other NITs, it is possible to assess the severity of these two independent features, defining the overall severity of NAFLD.[4] The results of the present study showed a higher prediction for activity (ballooning and lobular inflammatory) vs. the BARD, FIB4 and NAFLD scores. An advantage of FibroTest is the absence of ALT and AST as components of a fibrosis panel. ALT elevations decrease with age and cirrhosis and therefore should not be used as a component of fibrosis panels.[1, 23, 44, 45]

Methodology

We used a methodology in accordance with the specific guidelines for testing non-invasive biomarkers in chronic liver diseases.[19] When performances of tests were not compared directly in the same patients, the spectrum effect must be taken into account.[2, 17, 18] The performances of tests expressed by standard binary AUROCs or predictive values with standard cutoffs are dependent upon the prevalence of each stage/grade. The indirect comparisons of non-invasive test performances in NAFLD patients vs. CHC patients are therefore

potentially misleading, and so may not be fairly discussed in guidelines and reviews.[1]

A superficial binary analysis of the present results (File S3) for the prediction of cirrhosis (F4) could have concluded that the performance of FibroTest observed in NAFLD was 'excellent', with the standard (binary) AUROC for the prediction of F4 vs. F0 = 0.903 (95% CI, 0.831–0.975), even better to CHC in which a median of 0.860 (range 0.710–0.920) was observed in 11 studies.[46] For the prediction of fibrosis (F1 vs. F2), the performance of FibroTest in NAFLD could be classified as 'poor', with the standard AUROC = 0.592 (95% CI, 0.530–0.654) lower than a 'fair' classification in CHC with a median of 0.790 (range 0.700–0.890) as observed in 25 studies.[46]

If the spectrum effect is not taken into account, the diagnostic performance of FibroTest in NAFLD could be overestimated due a much higher prevalence of F0 (24%) compared with CHC, in which the prevalence of F0 was 2.4% in the larger database of CHC with biopsy published [supplement S1 in Ref. 23]. Comparisons between the present results in NAFLD vs. a large population of CHC (n = 1289)[47] found that the AUROCs between adjacent stages were lower only for F1 vs. F2 and for F2 vs. F3.

Advantages of the SAF scoring system as a reference

The validation of the SAF scoring system [3,4] has been a major breakthrough for NAFLD research, just as the METAVIR scoring system for patients with chronic viral hepatitis was 20 years ago.[5] By the late 1990s, the 'old and confusing' definitions of active or inactive hepatitis had become obsolete. It was necessary to analyse the severity of fibrosis and the severity necro-inflammatory activity separately. This enabled identification of the prognostic value of fibrosis progression to cirrhosis, the development of non-invasive biomarkers (which are now recommended at baseline in CHC and CHB), acceleration of the inclusion of expensive direct acting antivirals (DAA) in phase-3 trials in CHC, and most recently, prioritisation of the reimbursement of these DAA according to fibrosis severity.

In comparison with CHC, the definitions of NAFL vs. NASH in NAFLD are potentially confusing as they combine SAF. While a single lesion might be highly reproducible, lower reproducibility is expected in a composite diagnosis based on a combination of features such as NASH. The SAF scoring system now facilitates the separation of these three elementary features and so should accelerate the validation of non-invasive biomarkers.

Using the SAF scoring system and expressed by NonBinROC, these tests had similar performances as when using the previous references, including the consensual METAVIR scoring system for fibrosis and activity in CHC and CHB[2] or the less consensual adapted METAVIR scores in patients with NAFLD for fibrosis, and activity.[10] This new validation, employing the SAF scoring system specifically designed for NAFLD, clarifies the utility of these blood tests as non-invasive alternatives to biopsy.

Advantages of FLIP and FibroFrance-NAFLD cohorts

The overall population was very homogenous and was clearly distinct from the population of alcohol drinkers, as two-thirds of patients never drank other than very occasionally. The FLIP project allowed us to validate the diagnostic value of the non-invasive biomarkers (FibroTest, SteatoTest and ActiTest) in a large, nationally diverse cohort independent from the inventor of the tests. FibroFrance also had the advantage of prospectively following all types of chronic liver diseases since 1996, with routine non-invasive biomarkers introduced in 2002.

The aim of the present study was not to claim for a near perfect panel of NITs, but to confirm the significant performances of these tests in non-NAFLD patient. Our aim was in line with the recommendations of scientific societies such as EASL. In patients with presumed NAFLD, we improve the validation of NITs already recommended for viral hepatitis.

In conclusion, this study confirmed that FibroTest, ActiTest and SteatoTest are non-invasive tests that may offer an alternative to biopsy and correlate with grade and stage of the three elementary features of NAFLD: fibrosis, inflammatory activity and steatosis.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

File S1. SAF, METAVIR scoring systems and predetermined cutoffs.

File S2. Liver-FibroSTARD checklist.

File S3. Overall accuracy (NonBinROC) of FibroTest, ActiTest and SteatoTest performances for the histological diagnosis of fibrosis stage (SAF scoring system).

File S4. Box plots and statistical analyses (ANOVA) of FibroTest and FIB4, BARD and NAFLD-Fibrosis Score.

File S5. ActiTest and activity scores.

File S6. Box plots and statistical analyses (ANOVA) of SteatoTest according to SAF-Steatosis score in four grades.

File S7. Sensitivity analyses.

AUTHORSHIP

Guarantor of the article: Thierry Poynard.

Author contributions: TP: study concept and design; analysis and interpretation of data; statistical analysis; drafting; study supervision. MM, YN, OD: acquisition of data; analysis and interpretation of data; statistical analysis; drafting; critical revision of the manuscript. DT, QA, FC, VR: acquisition of data, drafting, critical revision of the manuscript; drafting. GM, EB, MT, MRG, CO, CD, JFD, SB, PB: acquisition of data, drafting.

All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

Declaration of personal interests: TP is the inventor of FibroTest/SteatoTest and the founder of BioPredictive, the company that markets these tests. Patents belong to the French Public Organization Assistance Publique-Hôpitaux de Paris (APHP). MM, YN, OD are BioPredictive employees. Others co-authors have no conflict of interest.

Declaration of funding interests: The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement HEALTH-F2-2009-241762 for the project FLIP.

REFERENCES

1. European Association for Study of Liver; Asociacion Latinoamericana para el Estudio del Higado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015; **63**: 237–64.
2. Houot M, Ngo Y, Munteanu M, Marque S, Poynard T. Systematic review with meta-analysis: direct comparisons of biomarkers for the diagnosis of fibrosis in chronic hepatitis C and B. *Aliment Pharmacol Ther* 2016; **43**: 16–29.
3. Bedossa P, Poitou C, Veyrie N, *et al.* Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012; **56**: 1751–9.
4. Bedossa P; FLIP Pathology Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* 2014; **60**: 565–75.
5. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; **24**: 289–93.
6. Ratziu V, Massard J, Charlotte F, *et al.* Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2006; **6**: 6.
7. Lassailly G, Caiazzo R, Hollebecque A, *et al.* Validation of noninvasive biomarkers (FibroTest, SteatoTest, and NashTest) for prediction of liver injury in patients with morbid obesity. *Eur J Gastroenterol Hepatol* 2011; **23**: 499–506.
8. Adams LA, George J, Bugianesi E, *et al.* Complex non-invasive fibrosis models are more accurate than simple models in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 2011; **26**: 1536–43.
9. Sebastiani G, Castera L, Halfon P, *et al.* The impact of liver disease aetiology and the stages of hepatic fibrosis on the performance of non-invasive fibrosis biomarkers: an international study of 2411 cases. *Aliment Pharmacol Ther* 2011; **34**: 1202–16.
10. Poynard T, Lassailly G, Diaz E, *et al.* Performance of biomarkers FibroTest, ActiTest, SteatoTest, and NashTest in patients with severe obesity: meta analysis of individual patient data. *PLoS ONE* 2012; **7**: e30325.
11. 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; doi: 10.1016/j.jhep.2016.04.023.
12. Poynard T, Ratziu V, Charlotte F, *et al.* Diagnostic value of biochemical markers(NashTest) for the prediction of non alcoholo-steato-hepatitis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2006; **6**: 34.
13. Poynard T, Ratziu V, Naveau S, *et al.* The diagnostic value of biomarkers (SteatoTest) for the prediction of liver steatosis. *Comp Hepatol* 2005; **4**: 10.
14. Ratziu V, de Ledinghen V, Oberti F, *et al.* A randomized controlled trial of high-dose ursodesoxycholic acid for nonalcoholic steatohepatitis. *J Hepatol* 2011; **54**: 1011–9.
15. Ratziu V, Harrison SA, Francque S, *et al.* Elafibranor, an Agonist of the Peroxisome Proliferator-Activated Receptor-α and -δ, Induces resolution of Nonalcoholic Steatohepatitis without fibrosis worsening. *Gastroenterology* 2016; **150**: 1147–59.
16. Friedman S, Sanyal A, Goodman Z, *et al.* Efficacy and safety study of cenicriviroc for the treatment of non-alcoholic steatohepatitis in adult subjects with liver fibrosis: CENTAUR Phase 2b study design. *Contemp Clin Trials* 2016; **47**: 356–65.
17. Poynard T, Halfon P, Castera L, *et al.* Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; **53**: 1615–22.
18. 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**: 1372–8.
19. Boursier J, de Ledinghen V, Poynard T, *et al.* An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. *J Hepatol* 2015; **62**: 807–15.
20. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. *Lancet* 1997; **349**: 825–32.
21. Poynard T, Mathurin P, Lai CL, *et al.* A comparison of fibrosis progression in chronic liver diseases. *J Hepatol* 2003; **38**: 257–65.
22. Poynard T, Munteanu M, Deckmyn O, *et al.* Validation of liver fibrosis biomarker (FibroTest) for assessing liver fibrosis progression: proof of concept and first application in a large population. *J Hepatol* 2012; **57**: 541–8.
23. Poynard T, Deckmyn O, Munteanu M, *et al.* Awareness of the severity of liver disease re-examined using software-combined biomarkers of liver fibrosis and necroinflammatory activity. *BMJ Open* 2015; **23**: e010017.
24. Poynard T, Munteanu M, Deckmyn O, *et al.* Applicability and precautions of use of liver injury biomarker FT. A reappraisal at 7 years of age. *BMC Gastroenterol* 2011; **11**: 39.
25. 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**: 1313–132120.
26. Poynard T, Morra R, Halfon P, *et al.* Meta-analyses of FibroTest diagnostic value in chronic liver disease. *BMC Gastroenterol* 2007; **7**: 40.
27. Poynard T, Lebray P, Ingiliz P, *et al.* Prevalence of liver fibrosis and risk factors in a general population using non-invasive biomarkers. *BMC Gastroenterol* 2010; **10**: 40.
28. Poynard T, Vergniol J, Ngo Y, *et al.* Staging chronic hepatitis B into seven categories, defining inactive carriers and assessing treatment impact using a fibrosis biomarker (FT®) and elastography (FibroScan®). *J Hepatol* 2014; **61**: 994–1003.
29. Naveau S, Gaudé G, Asnacios A, *et al.* Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. *Hepatology* 2009; **49**: 97–105.
30. Perazzo H, Munteanu M, Ngo Y, *et al.* Prognostic value of liver fibrosis and steatosis biomarkers in type-2 diabetes and dyslipidaemia. *Aliment Pharmacol Ther* 2014; **40**: 1081–93.
31. Poynard T, Lenaour G, Vaillant JC, *et al.* Liver biopsy analysis has a low level of performance for diagnosis of intermediate stages of fibrosis. *Clin Gastroenterol Hepatol* 2012; **10**: 657–63.
32. Poynard T, Halfon P, Castera L, *et al.* Variability of the area under the receiver operating characteristic curves in the diagnostic evaluation of liver fibrosis markers: impact of biopsy length and fragmentation. *Aliment Pharmacol Ther* 2007; **25**: 733–9.
33. Hintze J. NCSS 2013, LCC. Kaysville, Utah, USA, 2012. Available at: www.ncss.com.
34. Nguyen P.. nonbinROC: Software for evaluating diagnostic accuracies with non-binary gold standards. *J Stat Softw* 2007; **21**: 1–10.
35. Imbert-Bismut F, Messous D, Thibault V, *et al.* Intra-laboratory analytical variability of biochemical markers of fibrosis (Fibrotest) and activity (Actitest) and reference ranges in healthy blood donors. *Clin Chem Lab Med* 2004; **42**: 323–33.
36. Poynard T, Imbert-Bismut F, Munteanu M, *et al.* Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004; **23**: 3–8.
37. Ratziu V, Giral P, Munteanu M, *et al.* Screening for liver disease using non-invasive biomarkers (FibroTest, SteatoTest and NashTest) in patients with hyperlipidaemia. *Aliment Pharmacol Ther* 2007; **25**: 207–18.
38. Poynard T, Vergniol J, Ngo Y, *et al.* Staging chronic hepatitis C in seven categories using fibrosis biomarker (FT™) and transient elastography (FibroScan®). *J Hepatol* 2014; **60**: 706–14.
39. Ratziu V, Charlotte F, Heurtier A, *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005; **128**: 1898–906.
40. Rousselet MC, Michalak S, Dupré F, *et al.* Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005; **41**: 257–64.
41. Ding H, Ma JJ, Wang WP, *et al.* Assessment of liver fibrosis: the relationship between point shear wave elastography and quantitative histological analysis. *J Gastroenterol Hepatol* 2015; **30**: 553–8.
42. Kälsch J, Bechmann LP, Heider D, *et al.* Normal liver enzymes are correlated with severity of metabolic syndrome in a large population based cohort. *Sci Rep* 2015; **5**: 13058.
43. Friedrich-Rust M, Poynard T, Castera L. Critical comparison of elastography methods to assess chronic liver disease. *Nat Rev Gastroenterol Hepatol* 2016; **13**: 402–11.
44. Poynard T, Munteanu M, Ngo Y, *et al.* ActiTest accuracy for the assessment of histological activity grades in patients with chronic hepatitis C, an overview using Obuchowski measure. *Gastroenterol Clin Biol* 2010; **34**: 388–96.
45. Chao DT, Lim JK, Ayoub WS, *et al.* Systematic review with meta-analysis: the proportion of chronic hepatitis B patients with normal alanine transaminase ≤40 IU/L and significant hepatic fibrosis. *Aliment Pharmacol Ther* 2014; **39**: 349–58.
46. 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**: 807–20.
47. Poynard T, de Ledinghen V, Zarski JP, *et al.* FibroTest and Fibroscan performances revisited in patients with chronic hepatitis C. Impact of the spectrum effect and the applicability rate. *Clin Res Hepatol Gastroenterol* 2011; **35**: 720–30.

APPENDIX 1

The FLIP partners' consortium

Vlad Ratziu, Thierry Poynard: Assistance Publique Hôpitaux de Paris, Groupe Hospitalier Pitie Salpetriere, Universite Pierre et Marie Curie, Paris, France; Jean-Marie Castille, Yen Ngo: Biopredictive SAS, Paris, France; Tania Langon: Alma Consulting Group SAS, Paris, France; Chris Day, Dina Tiniakos: University of Newcastle, Newcastle, UK; Debbie Lawlor: University of Bristol, Bristol, UK; Giulio Marchesini: Universita di Bologna, Bologna, Italy; Fabio Marra: University di Firenze, Firenze, Italy; Elisabetta Bugianesi: Universita degli studi di Torino, Torino, Italy; Stefano Bellentani: Universita di Modena e Reggio Emilia, Modena, Italy; Jean-François Dufour: Université de Berne, Berne, Switzerland; Manuel Romero Gomez: Servicio Andaluz de Salud, Spain; Thorkild Sørensen: Bispebjerg hospital, Region Hovedstaden, Bispebjerg, Dutch; Claudio Tribelli: Fondazione Italiana Fegato-Fonlus, Milano, Italy; Samuele De Minicis: Universita di Ancona, Ancona, Italy; Michael Trauner: Medizinischen Universitaet, Wien, Austria; Claudia Oliveira: University of São Paulo School of Medicine, São Paulo, Brazil.

The FLIP pathology consortium

Pierre Bedossa: Assistance Publique-Hopitaux de Paris, hopital Beaujon, University Paris-Diderot, Paris, France; Alastair D. Burt: School of Medicine, University of Adelaide, Adelaide, Australia; Annette S.H. Gouw: Department of Pathology & Medical Biology, University Medical Center Groningen, Groningen, The Netherlands; Carolin Lackner: Institute of Pathology, Medical

University of Graz, Graz, Austria; Peter Schirmacher: Institut Fur Pathologie, Universitätsklinikum Heidelberg, Heidelberg, Germany; Luigi Terracciano: Institute of Pathology, University Hospital, Basel, Switzerland; Dina Tiniakos: Medical School, National and Kapodistrian University of Athens, Greece and Medical School, Newcastle University, Newcastle-upon-Tyne, UK; J. Brain: Medical School, Newcastle University, Newcastle-upon-Tyne, UK; Yvonne Bury: Royal Victoria Infirmary, Hospitals Foundation Trust, Newcastle upon Tyne, UK; Daniela Cabibi: University of Palermo, Palermo, Italy; Frederic Charlotte: Assistance Publique Hopitaux de Paris, Groupe Hospitalier Pitie-Salpetriere, Universite Pierre et Marie Curie, Paris, France; Ezio David: University of Torino, Ospedale Molinette, Torino, Italy; Luisa Losi: Modena University Hospital, Modena, Italy; Matteo Montani: Institute of Pathology, University of Bern, Bern, Switzerland; Maria Jesus Pareja: Hospital Universitario de Valme, Sevilla, Spain; Dominique Wendum: Assistance Publique-Hopitaux de Paris, Hopital St Antoine Universite Pierre et Marie Curie; Fritz Wrba: Medical University of Vienna, Vienna, Austria; Marianne Ziol: Assistance Publique-Hôpitaux de Paris, Hopital Jean Verdier, Université Paris 13, Bobigny, France; Vlad Ratziu: Assistance Publique Hopitaux de Paris, Hopital Pitie Salpetriere, Universite Pierre et Marie Curie, Paris, France.

The FibroFrance group

Thierry Poynard, Vlad Ratziu, Dominique Thabut, Joseph Moussalli, Pascal Lebray, Marika Rudler, Françoise Imbert Bismuth, Frederic Charlotte, Olivier Rosmorduc, Yvon Calmus: Anti-Fibrosis Center, Assistance Publique Hopitaux de Paris, Groupe Hospitalier Pitie-Salpetriere, Universite Pierre et Marie Curie, Paris, France; Agnes Hartemann, Sophie Jacqueminet: Diabetes Center, Assistance Publique Hopitaux de Paris, Groupe Hospitalier Pitie-Salpetriere, Universite Pierre et Marie Curie, Paris, France; Eric Bruckert, Philippe Giral: Lipid Center, Assistance Publique Hopitaux de Paris, Groupe Hospitalier Pitie-Salpetriere, Universite Pierre et Marie Curie, Paris, France; Sylvie Naveau, Gabriel Perlemuter: Alcoholic liver Disease Center, Assistance Publique Hopitaux de Paris, Antoine Beclere Hospital, Clamart, France; Brigitte Varsat: Prevention Centers of Caisse Primaire Assurance Maladie, Paris, France; Anne Mercadier: Transfusion Unit, Assistance Publique Hopitaux de Paris, Groupe Hospitalier Pitie-Salpetriere, Universite Pierre et Marie Curie, Paris, France.