Performance of liver biomarkers, in patients at risk of nonalcoholic steato-hepatitis, according to presence of type-2 diabetes

Thierry Poynard[a,b], Valentina Peta[b,c], Olivier Deckmyn[c], Raluca Pais[a,b], Yen Ngo[b,c], Frederic Charlotte[d], An Ngo[c], Mona Munteanu[b,c], Françoise Imbert-Bismut[e], Denis Monneret[e], Chantal Housset[b], Dominique Thabut[a,b], Dominique Valla[f], Christian Boitard[g], Laurent Castera[f] and Vlad Ratziu[a]; for the FLIP consortium, the FibroFrance Group, the EPIC-3 program and the QUID-NASH group

Objective There is a controversy about the performance of blood tests for the diagnostic of metabolic liver disease in patients with type-2-diabetes in comparison with patients without type-2-diabetes. These indirect comparisons assumed that the gold-standard is binary, whereas fibrosis stages, steatosis and nonalcoholic-steato-hepatitis (NASH) grades use an ordinal scale. The primary aim was to compare the diagnostic performances of FibroTest in type-2-diabetes vs. controls matched on gender, age, fibrosis stages and obesity, and taking into account the spectrum effect by Obuchowski measure.

Methods Data were retrospectively compared among patients prospectively included, with simultaneous biopsy and blindly assessed FibroTest, SteatoTest-2 and NashTest-2. The secondary aim was to construct an index (SpectrumF3F4-Index) to predict an adjusted-area under the receiver operating curve (AUROC) for F3F4 diagnosis from the prevalences of fibrosis stages, permitting to reduce the spectrum effect when performances of FibroTest, transient elastography and magnetic resonance elastography are indirectly compared.

Results In 505 patients at risk of NASH, the Obuchowski measures [95% confidence interval (CI)] of FibroTest, SteatoTest-2 and NashTest-2 were all equivalent in 136 type-2-diabetes cases vs. 369 matched controls: 0.871 (0.837–0.905), vs. 0.880 (0.879–0.881), 0.835 (0.797–0.873) vs. 0.806 (0.780–0.832) and 0.829 (0.793–0.865) vs. 0.855 (0.829–0.869), respectively. Standard-AUROCs (95% CI) were 0.932 (0.898–0.965), 0.872 (0.837–0.907) and 0.834 (0.699–0.969) and reduced after adjustment by SpectrumF3F4-Index to 0.794 (0.749–0.838), 0.767 (0.750–0.783) and 0.773 (0.725–0.822) for transient, magnetic resonance elastography and FibroTest, respectively.

Conclusions When compared by Obuchowski measures, the performances of tests were not different in patients with T2-diabetes vs. patients without T2-diabetes. When individual data are not available, adjusted-AUROCs reduced the spectrum effect. Eur J Gastroenterol Hepatol 32: 998–1007 Copyright © 2019 The Author(s). Published by Wolters Kluwer Health, Inc.

Introduction

Due to the increase in the liver disease-related mortality and the forthcoming treatment of nonalcoholic-stea-

to-hepatitis (NASH), there is a need to identify performant noninvasive biomarkers for the diagnosis of significant metabolic liver diseases [nonalcoholic fatty liver disease (NAFLD)] [1].

In patients with type-2-diabetes mellitus (T2-diabetes) [2], the performances of patented blood tests such as FibroTest, SteatoTest-2, NashTest-2 for the diagnosis of liver diseases lesions, despite in the range of performances observed in patients at risk of NASH with and without diabetes [3–5], have been interpreted as lower than those usually observed in patients without T2-diabetes. These indirect comparisons, without controls and not taking into account the spectrum effect, assumed that the gold standard is binary, whereas fibrosis stages, steatosis and NASH grades use an ordinal scale.

Indirect comparisons should be adjusted on the liver lesions' prevalence in the context of use and on the spectrum of stages or grades (spectrum effect). These issues have been already studied in patients with chronic hepatitis C for the performance of FibroTest for the diagnosis of fibrosis stages, and the use of Obuchowski measure was recommended as well as adjusted-AUROC according to the prevalences of stages [6,7].

These recommendations were not frequently followed in patients at risk of NASH. In a recent meta-analysis of

64 studies of fibrosis biomarkers [8], only four (6.4%) used Obuchowski measure [3,9–11]. We identify only two other studies published in the same period not included in the meta-analysis [12,13], and two studies published later [14,15], which used Obuchowski measure.

Therefore, the first aim was to analyze the performances of FibroTest (primary endpoint), and the new SteatoTest-2 [4], NashTest-2 [5], in a large group of T2-diabetes compared to patients at risk of NASH without T2-diabetes, and with patients with chronic hepatitis C, with and without T2-diabetes, using Obuchowski measures and matched controls for gender, age, obesity and fibrosis stages. The second aim, in the absence of Obuchowski measure, was to construct an index [index of fibrosis spectrum variability (Spectrum-Index)] in patients at risk of NASH, to predict an adjusted-AUROC for F3F4 diagnosis from the prevalences of biopsy proven fibrosis stages, as previously published for chronic hepatitis C [7]. This index was assessed in all the possible combinations of fibrosis stages using database with and without patients with T2-diabetes. This index permitted to assess the variability of the standard (binary) AUROCs for the diagnosis of fibrosis stages F3F4 of three biomarkers, FibroTest, transient elastography and magnetic resonance elastography, due to the presence of T2-diabetes, after adjustment according to the prevalences of fibrosis stages, from an-overview of 25 published studies [8].

Patients and methods

Common criteria of inclusion

The design was a noninterventional analysis of fresh serum specimen recorded in prospective subsets, two with patients at risk of NASH, and one with hepatitis-C to increase the sample size of cases and controls with biopsy with and without T2-diabetes (Fig. 1).

The inclusion criteria were the same. To standardize the performances' comparisons as much as possible, cases were defined as history of T2-diabetes or treatment of T2-diabetes, together with a fasting glucose of $7.0 \geq$ mmol/l, with a contemporaneous biopsy and blood sample used for liver tests [16]. Patients receiving specific treatment for NAFLD before biopsy or other cause of diabetes were not included. All patients had contemporaneous biopsies with centralized scoring of features, independently to any result of blood tests. Patients were excluded if the blood tests were disqualified according to the company recommendations for reliable tests [17], or if the interval between the biopsy and blood tests was greater than 4 weeks.

We analyzed post-HOC patients with T2-diabetes and at risk of NASH, and patients with chronic hepatitis C, prospectively included in two cohorts (FibroFrance NCT01927133, FLIP) [18–20] and a multicenter trial (EPIC3) [12,13,19,20] described in Supplementary File S2, Supplemental digital content 2, *http://links.lww.com/EJGH/A490*. The inclusion criteria were the presence of

Comparison of blood tests performances in patients at risk of NASH, or with chronic hepatitis C, with and without type-2 diabetes
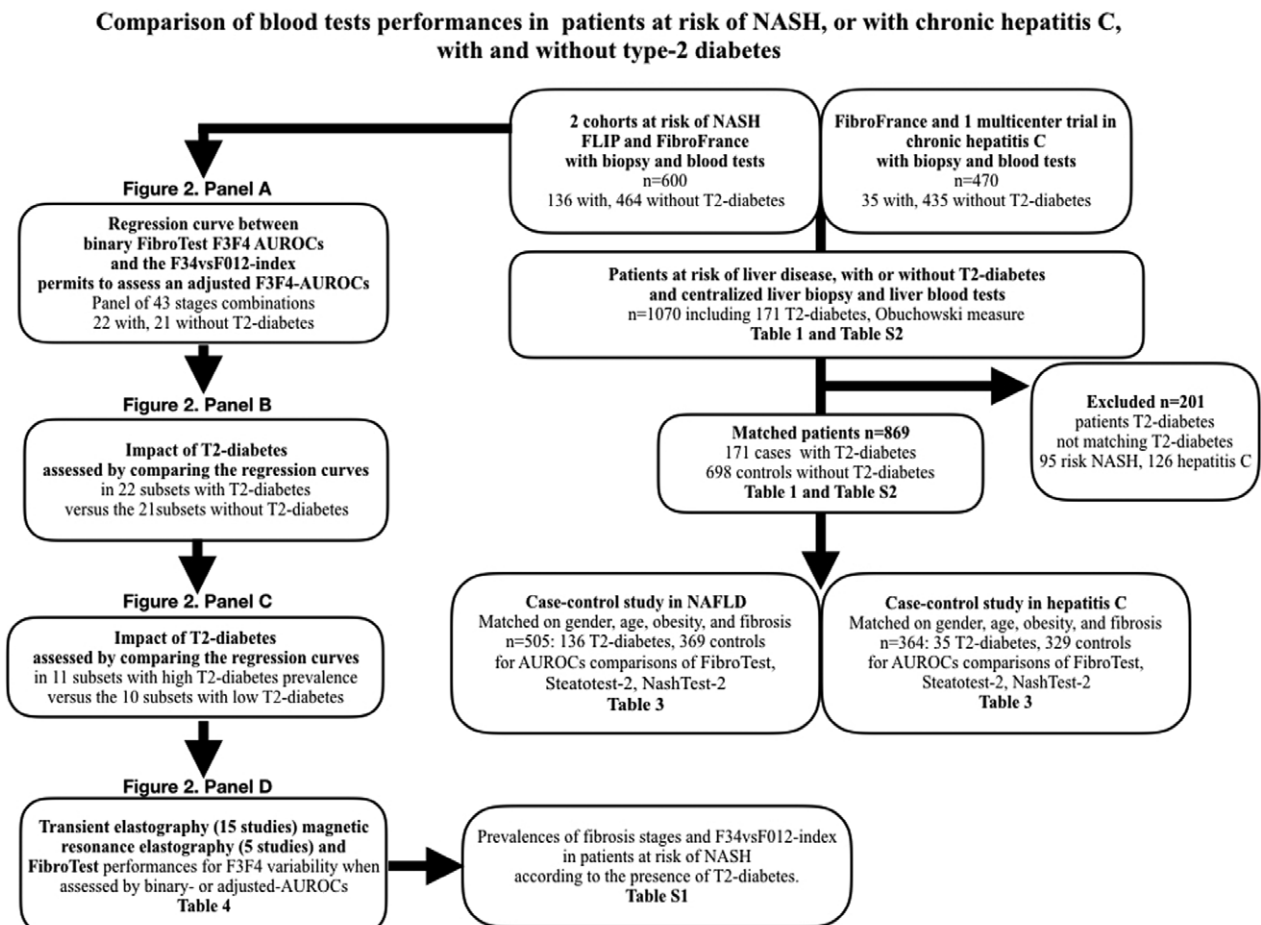
Fig. 1. Populations included.

reliable FibroTest, ActiTest and SteatoTest, as well as the reading of the biopsy using the SAF scoring system. Ethical approvals were obtained for the interventional studies at each participating institution and all patients provided written informed consent [12,13,18–20].

Blood tests

The FibroTest, the updated NashTest-2, the updated SteatoTest-2 are patented tests (BioPredictive, Paris, France) that have been validated extensively to assess the stages of fibrosis, the necro-inflammatory activity and steatosis grades of steatosis, using the SAF and NASH Clinical Research Network (CRN) scoring systems for NAFLD [2–5,12,13,15,18,21], and the meta-analysis of histological data in viral hepatitis (METAVIR) scoring system for hepatitis-C [22]. Details were given in Supplementary file S2, Supplemental digital content 2, *http://links.lww.com/EJGH/A490*.

The FibroTest includes serum α2-macroglobulin, apolipoprotein-A1, haptoglobin, total bilirubin and γ-glutamyl transpeptidase. NashTest-2 combined the seven components of FibroTest, plus cholesterol and triglycerides [5], SteatoTest-2 combined with different weights, 10 components of NashTest-2, without bilirubin, but with fasting glucose [4]. As the presence of at least 5% of steatosis is mandatory for the definition of NASH, but with a risk of false negative due to 'burning steatosis' (NASH without steatosis at biopsy, in cases without any cause of ballooning or inflammation) two NashTest-2 values were assessed by the algorithm. The standard NashTest-2 followed the CRN definition and cases with SteatoTest-2 grade S0 (<0.40 grade S1cutoff) were graded as non-NASH. The second value was the NashTest-2-'raw' value even if the SteatoTest-2 concluded as grade S0 (<0.40 grade S1 cutoff) [4,5,21]. The scores of these biomarkers range from 0 to 1.00, the highest scores being attributed to the most severe lesions. The preanalytical analytical procedures were those recommended by BioPredictive. Exclusion criteria were nonreliable results identified using security control algorithms [17].

Histological references

The SAF scoring system, specific for NAFLD features and permitting simplified construction of blood tests, has been described elsewhere, and detailed in Supplementary file S2, Supplemental digital content 2, *http://links.lww.com/EJGH/A490* [21]. 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) varies from 0 to 3. Activity grade (A, from 0 to 4) is the addition of hepatocyte ballooning (0–2) and lobular inflammation (0–2). Fibrosis stage (F) varies from 0 to 4 [21,22].

Statistical methods

Comparison of Obuchowski measures

Our main hypothesis was that there was no significant difference between the Obuchowski measure of the three blood tests, for the prediction of fibrosis stages, NASH grades and steatosis grades, respectively, in T2-diabetes vs. matched non-TD2M controls.

The primary endpoint was the Obuchowski measure, with five, five and four SAF ordinal classes, for FibroTest, NashTest-2 and SteatoTest-2, respectively [6].

Case-control studies

We matched the controls according to the four main factors associated, besides the presence of T2-diabetes, with the severity of NAFLD (male gender, age ≥ 50 years, liver fibrosis stages ≥ stage 2) and the severity of overweight (BMI ≥ 30 kg/m²) [23–26].

To construct a case-control population with the maximum of power and the minimum of differences between the confounding factors, we classified T2-diabetes and controls in 16 groups according to the 16 possible combinations of the four confounding factors. A nonsignificant difference was defined as a Fisher exact test ≥ 0.05 (Supplementary File S3, Table S1, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*).

Construction and comparisons of adjusted-area under the receiver operating curves in patients with and without diabetes

The second aim, in the absence of Obuchowski measure, was to construct a 'SpectrumF3F4-Index' in patients at risk of NASH, with and without T2-diabetes, to predict an adjusted-AUROC of FibroTest for F3F4 diagnosis from the prevalences of biopsy proven fibrosis stages, as previously published for chronic hepatitis C [7]. Two studies with biopsy and FibroTest were used, 600 subjects at risk of NASH, with and without diabetes [3], and in the only study in cases with T2-diabetes [2]. Details were described in Table 4.

The adjusted-AUROC was predicted by the linear regression linking the observed standard-AUROC to the SpectrumF3F4-Index. In a prospective study including 501 at risk of T2-diabetes of a tertiary center [23], the prevalence of the stages had an almost uniform distribution (each stage prevalence = 20%), and the SpectrumF3F4-Index was 2.59 (Supplementary File S3. Table S2, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*).

The regression curves linking the observed standard-AUROCs to the SpectrumF3F4-Index were compared for the 42 different combinations of prevalences in subsets of patients at risk of NASH with biopsy, in patients with and without T2-diabetes [2,3].

Furthermore, it was possible from published studies, to assess a possible impact of diabetes. For this purpose, we compared the regression lines of the subset of studies with prevalence of T2-diabetes equal or above the median vs. the subset with prevalence lower than the median. These data were those of an overview of the performances of transient and magnetic resonance elastography [8], with the studies of FibroTest [2,3,12,27].

We seek for a minimum of 100 cases of T2-diabetes to reach the median of the number of subjects included in diagnostic studies [28,29]. Comparison between Obuchowski measures used Z-test and equivalence tests of means [21]. Means comparison between several groups used multiple comparison Tukey-Kramer's test. Number Crunching

Statistical System (NCSS12) and numROC-software were used for statistical analyses [30,31].

Results

Patients' characteristics

A total of 1070 subjects were preincluded, 600 at risk of metabolic liver disease, and 470 with chronic hepatitis-C, including 136 and 35 T2-diabetes, respectively (Fig. 1). All the 16 combinations of matching factors were represented (Supplementary File S3, Table S1, Panel A, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*). In control groups, and in comparison, with T2-diabetes, there was a predominance of male under the age of 50 year, with BMI < 30 kg/m², and without clinically significant fibrosis (group 8).

The exclusion of 201 controls permitted to match the remaining controls with all T2-diabetes (Supplementary File S3, Table S1, Panel B, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*), without significant differences in prevalences of gender, age, severe obesity, and advanced fibrosis. A total 505 patients at risk of NASH were included, 369 controls and 136 T2-diabetes cases. A total of 354 hepatitis-C patients were included, 329 controls and 35 T2-diabetes cases, with similar prevalences of clinically significant fibrosis (Table 1).

Comparison of Obuchowski measures in patients with and without diabetes

For the primary endpoints, the Obuchowski measures in patients at risk of NASH, all the diagnostic performances of the three tests, were not different in cases with T2-diabetes vs. matched controls without T2-diabetes (Table 2). The higher difference was observed for NashTest-2, mean [95% confidence interval (CI)], 0.829 (0.793–0.865) in T2-diabetes vs. 0.855 (0.829–0.869) in matched controls, a

Table 1. Characteristics of 869 patients included in the case-control analyses

Characteristics	NASH risk, n = 505		Chronic hepatitis C, n = 364	
	Matched controls	T2-diabetes	Matched controls	T2-diabetes
N	369	136	329	35
Matching factors (%)				
Male gender	211 (57.2)	74 (54.4), P = 0.61	220 (66.9)	26 (74.3), P = 0.37
Age ≥ 50 years	242 (65.6)	101 (74.3), P = 0.07	113 (34.0)	17 (48.6), P = 0.09
Prevalence F2F3F4	204 (53.8)	80 (58.8) P = 0.23	157 (47.7)	17 (48.6), P = 0.92
BMI ≥ 30 kg/m²	194 (52.6)	85 (62.5), P = 0.06	60 (13.8)	11 (31.4), P = 0.06
Prevalence stages		P = 0.02		P = 0.80
F0	66 (17.9)	13 (9.6)	22 (6.7)	4 (11.4)
F1	109 (29.5)	43 (31.6)	150 (45.6)	14 (40.0)
F2	96 (26.0)	26 (19.1)	80 (24.3)	7 (20.0)
F3	76 (20.6)	43 (31.6)	46 (14.0)	6 (17.1)
F4	22 (6.0)	11 (8.1)	31 (9.4)	4 (11.4)
Activity grades		P = 0.55		
A0	32 (8.7)	14 (10.3)	NA	NA
A1	52 (14.1)	13 (9.6)	NA	NA
A2	125 (33.9)	34 (25.0)	NA	NA
A3	102 (27.6)	33 (24.3)	NA	NA
A4	58 (15.7)	42 (30.9)	NA	NA
Steatosis grades				P = 0.55
S0 (0–4%)	14 (3.8)	2 (1.5)	241 (73.3)	15 (42.9)
S1 (5–32%)	119 (32.2)	30 (22.1)	59 (17.9)	7 (20.0)
S2 (33–65%)	130 (35.2)	66 (48.5)	17 (5.2)	7 (20.0)
S3 (66–100%)	106 (28.7)	38 (27.9)	12 (3.7)	6 (17.1)
NAS score				
0	7 (1.9)	2 (1.5)	NA	NA
1	23 (6.2)	7 (5.2)	NA	NA
2	29 (7.9)	11 (8.1)	NA	NA
3	59 (16.0)	9 (6.6)	NA	NA
4	87 (23.6)	32 (23.5)	NA	NA
5	78 (21.1)	25 (18.4)	NA	NA
6	61 (16.5)	35 (25.7)	NA	NA
7–8	25 (6.8)	15 (11.0)	NA	NA
BMI (kg/m²), median	28.8	31.6	24.8	28.0
Age (year)	51.0	58.0	44.9	49.0
Glucose (mmol/l)	5.30	8.70	4.70	8.10
Cholesterol (mmol/l)	5.15	4.63	4.61	4.32
Triglycerides (mmol/l)	1.54	1.75	1.17	1.22
ApoA1 (g/l)	1.40	1.38	1.28	1.23
Haptoglobin (g/l)	1.17	1.40	0.79	0.67
A2M (g/l)	1.73	2.00	2.39	2.53
Bilirubin (µmol/l)	10	9	10	10
GGT (IU/l)	61	83	35	50
ALT (IU/l)	60	60	65	79
AST (IU/l)	41	44	47	63
FibroTest (0–1)	0.24	0.37	0.40	0.64
NashTest-2-raw (0–1)	0.64	0.76	NA	NA
NashTest-2 (0–1)	0.63	0.76	NA	NA
SteatoTest-2 (0–1)	0.61	0.81	0.37	0.68

NA, not applicable in patients with chronic hepatitis C; NASH, nonalcoholic-steato-hepatitis.

Table 2. Diagnostic performances of FibroTest, NashTest-2 and SteatoTest-2 in T2-diabetes vs. controls: Obuchowski measures in 600 patients at risk of nonalcoholic-steato-hepatitis

	NASH risk						
	All		Controls			T2-diabetes	
			Not-matched		Matched		
Tests							
N	600		464		369		136
Obuchowski	m	SE	M	SE	m	SE	m	SE
Fibrotest	0.878	0.007	0.876	0.008	0.880	0.009	0.871	0.017
NashTest-2	0.844	0.008	0.842	0.009	0.850	0.010	0.829	0.019
NashTest-2-raw	0.849	0.009	0.849	0.009	0.855	0.010	0.829	0.018
SteatoTest-2	0.815	0.010	0.798	0.011	0.806	0.013	0.835	0.019

m, mean; NASH, nonalcoholic-steato-hepatitis; SE = standard error.

nonsignificant difference (Z = –1.0) and equivalence (tests of means, $P = 0.01$). For SteatoTest-2, the Obuchowski measure was even higher in T2-diabetes, 0.835 (0.797–0.873) vs. matched-controls, 0.806 (0.780–0.832), nonsignificantly $P = 0.23$, with equivalence (tests of means, $P = 0.01$). For FibroTest, the Obuchowski measure was 0.871 (0.837–0.905) in T2-diabetes, vs. 0.880 (0.879–0.881) in matched-controls nonsignificantly different ($P = 0.60$), with equivalence ($P = 0.01$).

In hepatitis-C, the results were in line with the results observed in patients at risk of NASH, for the variability of standard-AUROCs (Supplementary File S3, Table S3, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*), for the absence of impact of diabetes on Obuchowski measures (Supplementary File S3, Table S4, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*) and for the interest of spectrum-adjusted AUROCs in indirect comparisons (Supplementary File S3, Table S5, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*).

Case-control studies to compare performances using standard-area under the receiver operating curves in patients with and without T2-diabetes

According to the histological endpoint, the standard-AU-ROCs of blood test varied in patients with T2-diabetes and suspected NASH from 0.575 to 0.801, without any significant differences with the performances in matched controls, which also varied from 0.555 to 0.848 (Table 3).

In T2-diabetes, the greatest standard-AUROC difference vs. matched controls was for NashTest-2-raw and the diagnostic of clinically significant NASH A2-5, 0.607 (0.481–0.708) vs. 0.671 (0.601–0.731) and was not-significant ($P = 0.34$), with equivalence ($P = 0.01$). For SteatoTest-2 and the diagnostic of steatosis grade S3, the standard-AUROC was even higher in T2-diabetes, 0.637 (0.506–0.739) vs. matched-controls, 0.555 (0.490–0.614), and was nonsignificant (P = 0.22) with equivalence ($P = 0.01$). In hepatitis C, the results were also in line with the results observed in patients at risk of NASH (Supplementary File S3, Table S3, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*).

Construction and comparisons of adjusted-area under the receiver operating curves in patients with and without diabetes

The linear regression between SpectrumF3F4-Index and standard-AUROCs was assessed among the 43 subsets combining fibrosis stages with and without T2-diabetes. The uniform standardized point is the intersection of the regression line (AUROCs vs. SpectrumF3F4-Index) and the horizontal line, assuming a SpectrumF3F4-Index of 2.5 (black dashed line), which is the index of a population of equal fibrosis stage prevalence (prevalence F0 = F1 = F2 = F3 = F4 = 20%). The adjusted-AUROC for FibroTest was 0.774. The regression curve permitted to calculate the adjusted F34-AUROC as = 0.553 + (0.088 × Spectrum-Index) with a $R^2 = 0.628$ (Fig. 2, Panel A). The red dotted line corresponds to an example of a given study in T2-diabetes [2], which had a FibroTest with standard-AU-ROC = 0.720 and is equivalent to an adjusted AUROC = 0.760 using its Spectrum-Index = 2.34 (green dotted line).

Among all the combinations of fibrosis prevalences, the regression lines (with 95% confidence limits), in 22 subsets with diabetes (in red, 95% confidence interval, $R^2 = 0.72$), and in 21 without T2-diabetes (in green, $R^2 = 0.53$), were similar. There was no significant curve-inequality between the two curves (F-ratio-test = 0.03, $P = 0.97$) (Fig. 2, Panel B).

Among 21 different studies of patients at risk of NASH, there was no significant curve-inequality between the two regression lines according to a prevalence of T2-diabetes equal or above the median (≥31.9%, in red with 95% CI, n = 11, $R^2 = 0.36$) vs. those with prevalence <median (in green, n = 10, $R^2 = 0.34$) F-ratio-test = 0.36, $P = 0.70$) (Fig. 2, Panel C).

### Pilot study for comparing the impact of spectrum effect on biomarkers performances

For FibroTest, there was a significant difference between standard-AUROCs lower in patients at risk of NASH (0.741; 95% CI: 0.691–0.784) vs. CHC (0.821; 0.766–0.864; $P = 0.02$). When adjusted-AUROCs were compared, the significance disappeared (0.745; 0.698–0.792) vs. 0.754 (0.705–0.803; $P = 0.80$) (Table 4).

Standard-AUROCs were 0.932 (0.898–0.965), 0.872 (0.837–0.907) and 0.834 (0.699–0.969) and reduced after adjustment to 0.794 (0.749–0.838), 0.767 (0.750–0.783) and 0.773 (0.725–0.822) for magnetic resonance, transient elastography and FibroTest respectively, without significant difference (Table 5).

Regression lines in publications assessing stages with transient elastography (n = 15), magnetic resonance elastography (n = 5) or FibroTest (n = 5), permitted to assess a possible variability between these three biomarkers. Indeed, a significant correlation (R-Pearson = 0.81; $P = 0.005$) was reached by FibroTest (green line; $R^2 = 0.50$) or magnetic resonance (red line $R^2 = 0.79$) but not by transient elastography (blue line, $R^2 = 0.01$) (Fig. 2, Panel D).

Discussion

Ninety percent of studies designed to validate noninvasive tests for the diagnosis of metabolic liver disease continue to use the standard-AUROC as a summary measure of diagnostic accuracy [1,2,8,23,24], even if the risk of spectrum effect and the type-one error of the tests comparing these measures in two samples with different distributions of stages or grades, have been identified [3,4]. In patients at risk of NASH, several articles discussed these limitations,
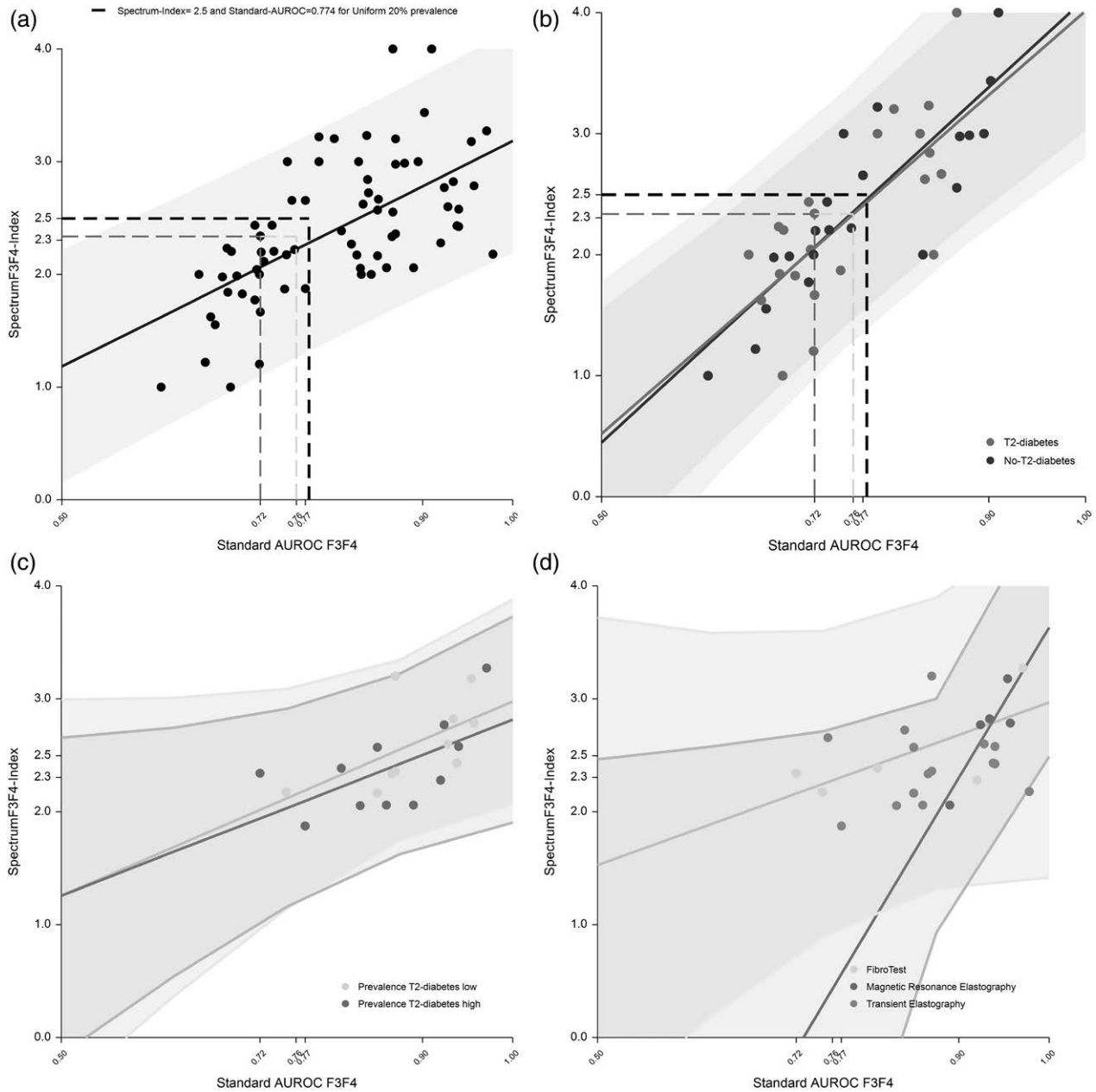
Fig. 2. Correlation between the standard-area under the receiver operating curve (AUROC) of FibroTest for the diagnosis of stage F3F4 according to the different spectrum of stages prevalences (summarized by the SpectrumF3F4-index), in patients with or without T2-diabetes. SpectrumF3F4-index is the difference between the mean of F3-F4 stages (from 3 if all patients were F3 to 4 if all were F4) and the mean of F0-F1-F2 stages (from 0 if all patients were F0 to 2 if all were F2). The regression curve (black line) between SpectrumF3F4-index and observed FibroTest' F3F4-AUROCs permitted to adjust the standard AUROC to various distribution of stages. If there was a uniform distribution between all fibrosis stages (uniform prevalence F0 = F1 = F2 = F3 = F4 = 20%), the SpectrumF3F4-index is equal to 2.5 {(3 + 4)/2 = 3.5 − (2 + 1 + 0)/3 = 1}. The uniform standardized point is the intersection (black dashed line) of the regression curve (between SpectrumF3F4-index and observed FibroTest' F3F4-AUROCs) and the horizontal line assuming a SpectrumF3F4-index = 2.5. Panel A: F3F4 AUROC adjusted using SpectrumF3F4-index (dashed green line) or not (dashed red line). A total of 43 combinations of fibrosis stages, with and without T2-diabetes. R^2 = 0.628. Regression curve (black line with 95%CI): predicted FibroTest adjusted F34-AUROC = 0.554 + (0.088 × SpectrumF3F4-index). For FibroTest, the adjusted F34-AUROC for a uniform distribution was 0.774 (black dashed line) with SpectrumF3F4-Index = 2.5 (dashed black line). The red dotted line corresponds to an example of a given study in T2-diabetes, the reference 2 (Bril *et al.*), which had a FibroTest with standard-AUROC = 0.720, which is equivalent to an adjusted AUROC = 0.760 (dashed dotted line) using its SpectrumF3F4-index which is 2.34. Panel B: Regression curves in 22 subsets with diabetes (in red, 95% confidence interval, R^2 = 0.72), and in 21 without T2-diabetes (in blue, R^2 = 0.53) were similar. There was no significant curve-inequality between the two curves (F-ratio-test = 0.03, P = 0.97). Panel C: Regression curves in 21 different publications in patients at risk of nonalcoholic-steato-hepatitis (NASH) giving the prevalence of diabetes, permitted to assess a possible impact according to a prevalence of T2-diabetes equal or above the median (≥31.9%, in red with 95% CI, n = 11, R^2 = 0.36) vs. those with prevalence < median (in blue, n = 10, R^2 = 0.34). There was no significant curve-inequality between these two regression lines (F-ratio-test = 0.36, P = 0.70). Panel D: Regression curves in 25 different publications assessing fibrosis stages assessed with transient elastography (TE, n = 15 in grey), magnetic resonance elastography (MRE, n = 5, in red) or FibroTest (n = 5, in blue), permitted to assess a possible variability between these three biomarkers. Indeed, a significant correlation (R-Pearson = 0.81; P = 0.005) was reached by the 10 studies using FibroTest (blue line; R^2 = 0.50) or magnetic resonance elastography (red line R^2 = 0.79) but not by the 15 studies using transient elastography (red curve, R-Pearson = 0.16; R^2 = 0.01).

Table 3. Standard-area under the receiver operating curves variability according to the histological feature's endpoint and choices of controls in 600 patients at risk of nonalcoholic-steato-hepatitis

Feature	Tests	NASH risk							
CRN score		All		Controls without T2-diabetes				T2-diabetes	
				Not-matched		Matched			
N		600		464		369		136	
Fibrosis		m/P-value	95% CI	m/P-value	95% CI	m/P-value	95% CI	m/P-value	95% CI
F1–F4	Fibrotest	0.666 < 0.001	0.613–0.714	0.644 < 0.001	0.585–0.697	0.659 < 0.001	0.584 0.723	0.706 0.004	0.516–0.830
F2–F4	Fibrotest	0.697 < 0.001	0.653–0.736	0.702 < 0.001	0.651–0.746	0.712 < 0.001	0.656–0.760	0.650, 0.001	0.546–0.734
F3–F4	Fibrotest	0.741 < 0.001	0.691–0.784	0.735 < 0.001	0.672–0.787	0.724 < 0.001	0.659–0.778	0.716 < 0.001	0.616–0.793
F4	FibroTest	0.845 < 0.001	0.744–0.909	0.861 < 0.001	0.746–0.926	0.848 < 0.001	0.730–0.917	0.801, 0.001	0.562–0.917
NASH									
A2–A5	NashTest-2	0.636 < 0.001	0.584–0.682	0.631 < 0.001	0.574–0.683	0.659 < 0.001	0.592–0.718	0.614, 0.03	0.486–0.716
A2–A5	NashTest-2raw	0.643 < 0.001	0.591–0.690	0.647 < 0.001	0.588–0.699	0.671 < 0.001	0.601–0.731	0.607, 0.03	0.481–0.708
A4–A5	NashTest-2	0.705 < 0.001	0.646–0.757	0.687 < 0.001	0.607–0.753	0.688, 0.001	0.607–0.754	0.669 < 0.001	0.559–0.757
A4–A5	NashTest-2raw	0.700 < 0.001	0.641–0.751	0.683, 0.001	0.605–0.748	0.673 < 0.001	0.591–0.742	0.672 < 0.001	0.562–0.758
Steatosis									
S1–S3	ST-2	0.632, 0.03	0.478–0.748	0.622, 0.05	0.454–0.747	Not applicable only 2 cases S0	Not applicable only 2 cases S0	Not applicable only 2 cases S0	Not applicable only 2 cases S0
S2–S3	ST-2	0.586 < 0.001	0.536–0.631	0.546, 0.04	0.490–0.598	0.555, 0.04	0.490–0.614	0.637, 0.01	506–0.739

CI, confidence interval; CRN, NASH Clinical Research Network; NASH, nonalcoholic-steato-hepatitis.

Table 4. Comparison between standard-F3F4-area under the receiver operating curves of FibroTest vs. adjusted-area under the receiver operating curves and Spectrum-index, according to presence of T2-diabetes

Disease (n)	F0-F1-F2 stages			F3F4 stages		Mean F012 spectrum	Mean F34 spectrum	F34 vs. F012-Index	Adjusted-AUROC	Standard AUROC
	%F0	%F1	%F2	%F3	%F4	(%F1 × 1 + %F2 × 2)/ (%F0 + %F1 + %F2)	(%F3 × 3 + %F4 × 4)/(%F3 + %F4)	Mean-F3F4- Mean-F0F1F2	Mean (95% CI)	
Uniform fibrosis prevalence	20.0	20.0	20.0	20.0	20.0	1.00	3.50	2.50	0.774	0.774
Subsets (number cases)										
All NASH risk (600)[a]	20.3	30.6	23.3	20.1	5.50	1.04	3.21	2.17	0.744	0.741
									0.694	0.691
									0.787	0.784
No-T2-diabetes (464)	23.4	30.3	24.5	16.8	4.7	1.01	3.22	2.21	0.747	0.735
									0.684	0.672
									0.799	0.787
Controls										
Non-T2-diabetes matched (369)	17.9	29.5	26.0	20.6	0.06	1.11	3.23	2.12	0.740	0.712
									0.684	0.656
									0.798	0.760
T2-diabetes (136)	9.5	31.6	19.1	31.6	8.0	1.16	3.20	2.04	0.733	0.716
									0.633	0.616
									0.585	0.568
T2-diabetes (151)[b]	25.1	41.7	16.5	12.5	3.9	0.90	3.24	2.34	0.760	0.720
									0.650	0.610
									0.870	0.830

Adjusted AUROC = 0.553 + (0.088 × Spectrum Index).
AUROC, area under the receiver operating curve; CI, confidence interval; NASH, nonalcoholic-steato-hepatitis.
[a]Present study; [b]Bril et al.

but less than 10% used the recommended Obuchowski measure [2,8,24].

The present results demonstrate that using appropriate methods, Obuchowski measures, adjusted AUROCs or case-controls comparisons, there was no significant difference between the performances of FibroTest, NashTest-2 or SteatoTest-2, in patients with or without T2-diabetes. A simple index of spectrum effect, SpectrumF3F4-Index, has been constructed, which permits to assess adjusted-AUROCs with decreased variability in comparison with standard-AUROCs. Finally, a pilot study suggests that this SpectrumF3F4-Index could identify biomarkers more impacted than others by spectrum index.

Strengths

First, we reproduce the same results for estimating the accuracy of FibroTest for fibrosis staging in patients at-risk of NASH than in hepatitis-C[7] and extend the analyses for estimating the accuracy of NashTest-2 for NASH grading and SteatoTest-2 for steatosis grading. Indeed, the Obuchowski measures were equivalent in T2-diabetes vs. nonT2-diabetes, as well than the standard-AUROCs when adjusted on the prevalence of fibrosis stages and the main covariables.

The strength of the present study was the consistent concordance between all the performances of FibroTest,

Table 5. Comparisons between 25 standard-area under the receiver operating curves of transient elastography, magnetic resonance elastography, and FibroTest after adjustment on the SpectrumF3F4-Index

Disease (n)	F0-F1-F2 stages			F3F4 stages		Mean F0-F1-F2 (%F1 × 1 + %F2 × 2)/%F0 + %F1 + %F2)	Mean F3-F4 (%F3×3 + %F4x4)/(%F3 + %F4)	Spectrum index Mean-F3-F4-Mean-F0-F1-F2	Adjusted AUROC Based on stages prevalence	Standard AUROC M (SD)	%T2 diabetes
	%F0	%F1	%F2	%F3	%F4						
Uniform fibrosis prevalence (n)	20.0	20.0	20.0	20.0	20.0	1.00	3.50	2.50	0.774		
Transient elastography M or XL probes (15) mean; 95% CI	25.3; 18.4–32.1	30.3; 25.8–34.8	19.2; 15.1–23.4	15.3; 12.3–18.2	9.9; 5.6–14.2	0.93; 0.79–1.07	3.34; 3.24–3.45	2.41; 2.23–2.60	0.767; 0.750–0.783	0.872; 0.837–0.907	31.5; 21.5–41.9
1 Wong GL 2008 (34)	50.0	20.6	5.9	8.8	14.7	0.42	3.63	3.20	0.836	0.870	19.0
2 Wong VW 2010 (246)	28.5	30.5	18.3	12.6	10.2	0.87	3.45	2.58	0.781	0.940	36.2
3 Lupsor 2010 (72)	34.7	40.3	18.1	6.9	0.0	0.82	3.00	2.18	0.746	0.978	NA
4 Petta 2011 (120)	26.0	27.4	24.0	15.1	7.5	0.97	3.33	2.36	0.762	0.870	14.0
5 Gaia 2011 (72)	31.9	22.2	22.2	11.1	12.5	0.87	3.53	2.66	0.788	0.755	NA
6 Kumar 2013 (120)	26.7	28.3	22.5	14.2	8.3	0.95	3.37	2.42	0.767	0.940	NA
7 Mahadeva 2013 (131)	3.8	38.9	35.1	16.0	6.1	1.40	3.28	1.87	0.719	0.770	47.3
8 Frossard 2013 (30)	13.3	20.0	16.7	16.7	33.3	1.07	3.67	2.60	0.783	0.928	23.3
9 Aykut 2014 (88)	26.1	23.9	19.3	20.5	10.2	0.90	3.33	2.43	0.768	0.938	23.9
10 Petta 2015 (253)	24.9	30.4	23.7	14.2	6.7	0.98	3.32	2.34	0.760	0.866	19.8
12 Imajo 2016 (142)	9.9	35.9	22.5	23.9	7.7	1.19	3.24	2.06	0.735	.860	50.0
13 Boursier 2016 (452)	8.6	27.2	26.1	25.2	12.8	1.28	3.34	2.05	0.735	0.831	46.7
14 Friedrich-Rust 2010 (50)	42.0	28.0	6.0	18.0	6.0	0.53	3.25	2.72	0.794	0.840	NA
15 Wong VW 2012 (193)	24.4	31.1	15.0	16.6	13.0	0.87	3.44	2.57	0.780	0.850	51.0
16 Naveau 2014 (100)	28.0	50.0	13.0	9.0	0.0	0.84	3.00	2.16	.745	0.850	15.0
Studies with magnetic resonance elastography (5) mean; 95% CI	34.3; 16.3–52.2	30.1; 23.6–36.3	12.1; 4.8–19.5	12.7; 4.5–20.9	10.9; 0.9–21.0	0.72; 0.39–1.05	3.45; 3.19–3.70	2.72; 2.22–3.23	0.794; 0.749–0.838	932; 0.898–0.965	32.6; 19.8–45.4
17 Kim 2013 (142)	35.2	23.9	8.5	7.0	25.4	0.60	3.78	3.18	0.834	0.954	27.5
18 Loomba 2014 (117)	36.8	33.3	11.1	10.3	8.5	0.68	3.45	2.77	0.798	0.924	34.2
19 Cui 2015 (125)	47.1	25.5	8.8	12.7	5.9	0.53	3.32	2.79	0.799	0.957	25.5
20 Imajo 2016 (142)	9.9	35.9	22.5	23.9	7.7	1.19	3.24	2.06	0.735	0.890	50.0
21 Cui 2016	42.4	31.2	9.6	9.6	7.2	0.61	3.43	2.82	0.803	0.934	26.0
Studies with FibroTest (5) mean; 95% CI	35.2; 15.1–55.3	35.9; 29.4–42.4	15.6; 7.1–24.2	10.2; 0.9–19.5	3.1; 0.4–5.8	0.79; 0.51–1.07	3.28; 2.96–3.61	2.49; 1.94–3.04	0.773; 0.725–0.822	0.834; 0.699–0.969	44.5; 5.5–83.5
50 Ratziu 2006 (170)	44.7	31.8	18.2	5.3	0	0.72	3.00	2.28	0.755	0.920	36.0
51 Ratziu 2006 (97)	26.8	41.2	15.5	12.4	4.1	0.86	3.25	2.39	0.764	0.810	32.0
52 Lassailly 2011 (288)	59.0	34.0	4.5	0.7	1.7	0.44	3.71	3.27	0.842	0.971	31.9
53 Munteanu 2016 (600)	20.3	30.7	23.3	20.2	5.5	1.04	3.21	2.17	0.745	0.749	22.7
54 Bril 2018 (151)	25.2	41.8	16.6	12.5	3.9	0.90	3.24	2.34	0.760	0.720	100.0

There was no significant difference between the means of standard-AUROCs of the three tests: multiple Tukey-Kramer $P > 0.08$, between adjusted-AUROCs of the three tests: multiple Tukey-Kramer $P > 0.26$, and between prevalences of T2-diabetes ($P > 0.43$). AUROCs, area under the receiver operating curves.

SteatoTest-2, in the two cases-controls population at-risk of NASH or with hepatitis-C, whatever the methodology used, Obuchowski measure, or adjusted-AUROCs.

The number of cases and controls together with the centralized assessment of both the histological endpoint and the components of the tests were also a strength of the study. The number of controls was sufficient to find those who matched according to the main covariables, gender, age, obesity and fibrosis. Indeed, cases with diabetes had twice more women, older than 50 years of age and with severe obesity (25.7%) than in the non-T2-diabetes of the population at risk of NASH (11.3%). Matching was mandatory as male younger than 50 years of age, without severe obesity, and without significant fibrosis was 10 times more frequent (17.2%) than among T2-diabetes (1.5%) (Supplementary file S4, Table S2, Supplemental digital content 3, *http://links.lww.com/EJGH/A491*).

The SpectrumF3F4-Index will never replace the Obuchowski measure to assess the complete performance of diagnostic tests for the prediction of an ordinal endpoint such as fibrosis stages. However, this method needs to have all the individual data. The interest of a spectrum index is that it can be simply assessed if the details of stages are given, and the adjusted-AUROCs can be directly compared to the standard-AUROCs.

Our study has revealed several other limitations of recent overviews or meta-analyses of biomarkers in patients at risk of NASH [8,32,33]. Ideally, comparisons between biomarkers should be performed by direct comparisons in the same patients and analyzed in intention to diagnose, the failure and the nonreliable results being considered as a misclassification of the biomarker [29]. Therefore, the AUROCs of transient elastography with the M-probe in patients at risk of NASH should be reduced by at least 20%. If patients with nonreliable/failure results had different prevalences than those with reliable results, that is more reliable in F4 and less reliable in obese F0, a major spectrum effect is possible. Frequently, the prevalences of F0 were not separately given from F1 [8]. Here, the heterogeneity of histological F0 prevalence varied from 3.8 to 50% in transient-elastography studies, much more than for FibroTest (26.8–59.0%).

Limitations

We acknowledge several limitations. We defined T2-diabetes as fasting glucose ≥7 mmol/l. This definition permitted to homogenize the matching with controls, but we have not analyzed the previous history of diabetes, the glycosylated hemoglobin-A1C, and the treatments received for the diabetes as well as lipid-lowering and blood pressure medications. Due to the sample size and to the four covariables already included in the matching, we were not able to assess several other factors potentially associated with variability of the tests in T2-diabetes, that is, African-American ethnicities [34].

Another limitation was the absence of recognized 'natural' prevalences of fibrosis stages, NASH grades and steatosis grades, in large population of T2-diabetes and matched controls, as available in hepatitis-C [3]. Ideally, the tests should be compared by Obuchowski measures in T2-diabetes vs. not-T2-diabetes, using the 'natural' prevalences of stages or grades in the different context usual

context of use. However, according to the larger study published, we used the prevalences assessed in diabetology open clinics, which was close to the uniform prevalence of the five fibrosis stages [23].

The regression curves and the spectrum index were assessed on a relatively small number of combinations (n = 43) and published studies (n = 33). However, the graphical concordance of the curves with and without T2-diabetes was useful for excluding a significant impact of T2-diabetes. For transient-elastography, contrarily to magnetic resonance and FibroTest curves, there was no significant correlation between the standard-AUROCs and the spectrum index. In the absence of individual data, it was not possible to explain these differences, such as a heterogeneity between covariables or between cutoffs. We also acknowledge that the power to exclude a significant impact of diabetes was limited by the sample size of patients with T2-diabetes. However, it is the first time that the risk of false interpretations of indirect comparisons between NASH tests performances was underlined.

In conclusion, when compared by Obuchowski measures, the performances of FibroTest, SteatoTest-2 and NashTest-2 were not different in patients with T2-diabetes vs. patients without T2-diabetes. When individual data are not available, adjusted-AUROCs reduced the variability due to spectrum effect.

Conflicts of interest

Thierry Poynard is the inventor of FibroTest, SteatoTest-2, and NashTest-2, founder of BioPredictive, the patents belong to the public organization Assistance Publique-Hôpitaux de Paris. Valentina Peta, Olivier Deckmyn, Mona Munteanu, Yen Ngo, and An Ngo are full employees of BioPredictive. The other authors have nothing to declare, Denis Monneret, Frederic Charlotte, Olivier Lucidarme, Françoise Imbert-Bismut, Chantal Housset, Dominique Thabut, Dominique Valla, Christian Boitard, Laurent Castera and Vlad Ratziu.

References

1   Younossi ZM, Loomba R, Anstee QM, Rinella ME, Bugianesi E, Marchesini G, *et al*. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. *Hepatology* 2018; 68:349–360.

2   Bril F, McPhaul MJ, Caulfield MP, Castille JM, Poynard T, Soldevila-Pico C, *et al*. Performance of the steatotest, actitest, nashtest and

fibrotest in a multiethnic cohort of patients with type 2 diabetes mellitus. *J Investig Med* 2019; 67:303–311.

3　Munteanu M, Tiniakos D, Anstee Q, Charlotte F, Marchesini G, Bugianesi E, *et al.*; FLIP Consortium and the FibroFrance Group. Diagnostic performance of fibrotest, steatotest and actitest in patients with NAFLD using the SAF score as histological reference. *Aliment Pharmacol Ther* 2016; 44:877–889.

4　Poynard T, Peta V, Munteanu M, Charlotte F, Ngo Y, Ngo A, *et al.* The diagnostic performance of a simplified blood test (SteatoTest-2) for the prediction of liver steatosis. *Eur J Gastroenterol Hepatol* 2019; 31:393–402.

5　Poynard T, Munteanu M, Charlotte F, Perazzo H, Ngo Y, Deckmyn O, *et al.*; FLIP Consortium, the FibroFrance-CPAM Group, the FibroFrance-Obese Group. Diagnostic performance of a new noninvasive test for nonalcoholic steatohepatitis using a simplified histological reference. *Eur J Gastroenterol Hepatol* 2018; 30:569–577.

6　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.

7　Poynard T, Halfon P, Castera L, Munteanu M, Imbert-Bismut F, Ratziu V, *et al.*; FibroPaca Group. Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; 53:1615–1622.

8　Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. *Hepatology* 2017; 66:1486–1501.

9　Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. *Radiology* 2013; 268:411–419.

10　Naveau S, Lamouri K, Pourcher G, Njiké-Nakseu M, Ferretti S, Courie R, *et al.* The diagnostic accuracy of transient elastography for the diagnosis of liver fibrosis in bariatric surgery candidates with suspected NAFLD. *Obes Surg* 2014; 24:1693–1701.

11　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.

12　Lassailly G, Caiazzo R, Hollebecque A, Buob D, Leteurtre E, Arnalsteen L, *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.

13　Poynard T, Lassailly G, Diaz E, Clement K, Caïazzo R, Tordjman J, *et al.*; FLIP consortium. Performance of biomarkers fibrotest, actitest, steatotest, and nashtest in patients with severe obesity: meta analysis of individual patient data. *Plos One* 2012; 7:e30325.

14　Loong TC, Wei JL, Leung JC, Wong GL, Shu SS, Chim AM, *et al.* Application of the combined fibrometer vibration-controlled transient elastography algorithm in chinese patients with non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 2017; 32:1363–1369.

15　Ducancelle A, Leroy V, Vergniol J, Sturm N, Le Bail B, Zarski JP, *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.

16　Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 1998; 15:539–553.

17　Poynard T, Munteanu M, Deckmyn O, Ngo Y, Drane F, Messous D, *et al.* Applicability and precautions of use of liver injury biomarker fibrotest. A reappraisal at 7 years of age. *BMC Gastroenterol* 2011; 11:39.

18　Munteanu M, Pais R, Peta V, Deckmyn O, Moussalli J, Ngo Y, *et al.*; FibroFrance Group. Long-term prognostic value of the fibrotest in patients with non-alcoholic fatty liver disease, compared to chronic hepatitis C, B, and alcoholic liver disease. *Aliment Pharmacol Ther* 2018; 48:1117–1127.

19　Poynard T, Munteanu M, Colombo M, Bruix J, Schiff E, Terg R, *et al.* Fibrotest is an independent predictor of virologic response in chronic hepatitis C patients retreated with pegylated interferon alfa-2b and ribavirin in the EPIC[3] program. *J Hepatol* 2011; 54:227–235.

20　Poynard T, Ratziu V, McHutchison J, Manns M, Goodman Z, Zeuzem S, *et al.* Effect of treatment with peginterferon or interferon alfa-2b and ribavirin on steatosis in patients infected with hepatitis C. *Hepatology* 2003; 38:75–85.

21　Poynard T, Munteanu M, Charlotte F, Perazzo H, Ngo Y, Deckmyn O, *et al.*; FLIP consortium, the FibroFrance-CPAM group; FibroFrance-Obese group. Impact of steatosis and inflammation definitions on the performance of NASH tests. *Eur J Gastroenterol Hepatol* 2018; 30:384–391.

22　Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. *Hepatology* 1996;24:289–293.

23　Goh GB, Pagadala MR, Dasarathy J, Unalp-Arida A, Sargent R, Hawkins C, *et al.* Clinical spectrum of non-alcoholic fatty liver disease in diabetic and non-diabetic patients. *BBA Clin* 2015; 3: 141–145.

24　Perazzo H, Munteanu M, Ngo Y, Lebray P, Seurat N, Rutka F, *et al.*; FLIP Consortium. Prognostic value of liver fibrosis and steatosis biomarkers in type-2 diabetes and dyslipidaemia. *Aliment Pharmacol Ther* 2014; 40:1081–1093.

25　Loomba R, Abraham M, Unalp A, Wilson L, Lavine J, Doo E, Bass NM; Nonalcoholic Steatohepatitis Clinical Research Network. Association between diabetes, family history of diabetes, and risk of nonalcoholic steatohepatitis and fibrosis. *Hepatology* 2012; 56:943–951.

26　McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, *et al.* Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *Am J Gastroenterol* 2017; 112:740–751.

27　Ratziu V, Massard J, Charlotte F, Messous D, Imbert-Bismut F, Bonyhay L, *et al.*; LIDO Study Group; CYTOL study group. 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.

28　Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs. nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. *Clin Gastroenterol Hepatol* 2015; 13:643–654.e1.

29　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.

30　Hintze JL. *NCSS 2012 User Guide. Number Cruncher Statistical Systems Software NCSS*. Kaysville, UT: NCSS; 2012.

31　R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. https://www.R-project.org/.

32　European Association for the Study of the Liver. Clinical practice guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol*. 2016; 64:1388–1402.

33　Bazerbachi F, Haffar S, Wang Z, Cabezas J, Arias-Loste MT, Crespo J, *et al.* Range of normal liver stiffness and factors associated with increased stiffness measurements in apparently healthy individuals. *Clin Gastroenterol Hepatol* 2019; 17:54–64.e1.

34　Bril F, Portillo-Sanchez P, Liu IC, Kalavalapalli S, Dayton K, Cusi K. Clinical and histologic characterization of nonalcoholic steatohepatitis in african american patients. *Diabetes Care* 2018; 41: 187–192.