ELSEVIER

Journal of Hepatology xxx (2006) xxx–xxx

Journal of Hepatology

www.elsevier.com/locate/jhep

Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C

Philippe Halfon[1,*], Yannick Bacq[2], Anne De Muret[3], Guillaume Penaranda[4], Marc Bourliere[5], Denis Ouzan[6], Albert Tran[7], Danielle Botta[8], Christophe Renou[9], Marie-Claude Bréchot[10], Claude Degott[11], Valérie Paradis[11]

[1]*Alphabio Laboratory, 23 rue Friedland, 13006 Marseilles, France*
[2]*Hepatology Unit, Trousseau Universitary Hospital, Tours, France*
[3]*Pathology Department, Trousseau Universitary Hospital, Tours, France*
[4]*CDL Pharma, Marseilles, France*
[5]*Hepatology Unit, Saint-Joseph Hospital, Marseilles, France*
[6]*Hepatology Unit, Arnault Tzanck Institute, Saint Laurent du Var, France*
[7]*Hepatology Unit, Archet Universitary Hospital, Nice, France*
[8]*Hepatology Unit, La Conception Universitary Hospital, Marseilles, France*
[9]*Hepatology Unit, General Hospital, Hyères, France*
[10]*Biochemistry Laboratory, Trousseau Universitary Hospital, Tours, France*
[11]*Pathology Department, Beaujon Universitary Hospital, Clichy, France*

***Background/Aim*: We evaluated the test performance profile (TPP) of blood tests of liver fibrosis.**

***Methods*: Three hundred and fifty-six patients with C chronic hepatitis were included in two centers. Metavir staging of liver specimens by two independent pathologists and the following tests were evaluated: Fibrotest (FT), APRI, FibroMeter (FM), and Hepascore (HS).**

***Results*: Metavir stages were: F0: 4%, F1: 55%, F2: 26%, F3: 11%, and F4: 4%. The AUROCs were not significantly different, respectively, FT, FM, APRI, HS: $\geqslant$F2: 0.79, 0.78, 0.76, 0.76; $\geqslant$F3: 0.81, 0.85, 0.81, 0.81; and F4: 0.86, 0.94, 0.92, 0.89. The TPP relies on the paired comparison of blood-test misclassification based on liver specimen, e.g. FT vs FM, respectively: F0+1: 18 vs 28% ($p = 0.0003$), $\geqslant$F2: 43 vs 31% ($p = 0.004$). There was no center effect.**

***Conclusions*: In those populations, the four blood tests had a similar performance for significant fibrosis ($F \geqslant 2$), lying in the lower range of published results which is attributable to a low $\geqslant$F2 prevalence, and for $\geqslant$F3 and F4. However, FM and FT had performance profiles significantly different as a function of fibrosis stages or diagnostic target (fibrosis cut-off). This has to be considered during the interpretation process. Moreover, the performance should be reported with different diagnostic targets.**

***Keywords*: Liver fibrosis; Blood test; Metavir staging; Liver biopsy; C chronic hepatitis; Blood marker**

1. Introduction

Non-invasive diagnosis of liver fibrosis currently relies on blood tests and physical techniques like imaging and liver stiffness measurement. Several blood tests have been published [1]. The main diagnostic target criterion is clinically significant fibrosis. We can classify these blood tests into four categories: simple tests made

Received 31 July 2006; received in revised form 28 August 2006; accepted 11 September 2006

* Corresponding author. Tel.: +33 4 91 25 41 00; fax: +33 4 91 79 20 44.
E-mail address: philippe.halfon@alphabio.fr (P. Halfon).

Abbreviations: APRI, aspartate aminotransferase to platelet ratio index; AUROC, area under the receiving–operating-characteristic; DA, diagnostic accuracy; FM, FibroMeter; FT, Fibrotest; HS, Hepascore; PACA, Provence-Côte d'Azur.

0168-8278/$32.00 © 2006 Published by Elsevier B.V. on behalf of the European Association for the Study of the Liver.
doi:10.1016/j.jhep.2006.09.020

DOCTOPIC: Viral Hepatitis

of a few indirect markers without algorithms like the aspartate aminotransferase to platelet ratio index (APRI) [2], indirect scores including indirect markers composite of an algorithm like Fibrotest (FT) [3], direct scores including direct markers – involved in connective tissue – composite of an algorithm like European liver fibrosis score [4] and mixed scores including indirect and direct markers composite of an algorithm like FibroMeter (FM) [5]. The performance of these tests largely depends on three main parameters: the accuracy of the reference method, usually liver biopsy, the intrinsic performance of the test and the prevalence of the diagnostic target. Although some characteristics of liver biopsy interpretation can be controlled (length, observer variability), this is the main limitation of performance measurement. The intrinsic performance mainly depends the blood markers and their measurement. Usually, such tests are implemented within one center with a training set and a validating set of patients. However, the next validating step is considered essential before expanding the tests to widespread routine use. Indeed, independent external validation includes assessing a potential variability of the three main parameters of diagnostic performance. Only, a few tests have been evaluated by external independent analysis with the exception of simple tests like APRI and an indirect test such as FT. However, recently some mixed tests with a high putative performance were published: Hepascore (HS) and FM [5,6]. In addition, the comparison of tests is often limited to simple descriptors whereas a more precise description of performance could be useful for clinical application [7].

The main aim of this study was to perform an external independent evaluation of these mixed tests in chronic hepatitis C by comparison with APRI as a simple test and FT as an indirect test. Direct tests were not evaluated since they include direct markers either with a low performance in chronic hepatitis C [4] or they are not easily available in routine [8–10]. The secondary aim was to depict a new descriptor, the test performance profile.

2. Patients and methods

2.1. Patients

Three-hundred fifty-six patients attending the hepatogastroenterology unit of the University Hospital in Tours, and five units (two University Hospital, two public hospitals, one private clinic) from Provence-Côte d'Azur area, France, were retrospectively enrolled. Patients with chronic viral hepatitis C were included if they had a positive HCV-RNA in the serum and a liver biopsy and an alcohol consumption <30 g/day for the past 5 years. Blood fasting samples were taken at entry and a percutaneous (usually 1.4–1.6 mm diameter needle, suction technique) transcostal liver biopsy was performed within 1 week. Patients were not included if they had liver specimen <15 mm or other cause of liver disease or complicated cirrhosis or were given putative anti-fibrotic treatment (e.g. interferon or sartan) in the past 6 months. The study protocol was approved by a local Ethics committee and conformed to the ethical guidelines of the current Declaration of Helsinki. The dates of inclusion (liver biopsy) ranged from October 1994 to March 2004 in Tours center and from September 2002 to January 2004 in Provence area.

2.2. Methods

2.2.1. Blood parameters

Analyses of blood samples provided the following variables: platelet count, urea, bilirubin, γ-glutamyl transpeptidase, aspartate and alanine aminotransferases, prothrombin index, apolipoprotein A1, haptoglobin, hyaluronic acid, and α-2-macroglobulin. Blood markers were measured either on fresh blood or frozen sample of serum stored at −20 °C. Techniques are detailed in Table 1. Sampling was performed for routine diagnostic aim within 1 week of liver biopsy.

2.2.2. Liver biopsy

Biopsy specimens were fixed in a formalin–alcohol–acetic solution and embedded in paraffin; 5 μm thick sections were stained with haematoxylin–eosin–saffron. Fibrosis was staged by two independent expert pathologists according to the Metavir staging system [11]. Observers were blinded for patient characteristics. When the pathologists did not agree, the specimens were re-examined under a double-headed microscope to analyse discrepancies and reach a consensus.

Table 1
Analytical methods for blood tests: automates and reagents as a function of center

	Provence-Côte d'Azur	Tours
Aspartate aminotransferase	CX 7 (Beckman)	Hitachi/Olympus[a]
Alanine aminotransferase	CX 7 (Beckman)	Hitachi/Olympus
γ-Glutamyl transpeptidase	CX 7 (Beckman)	Hitachi/Olympus
Bilirubin	CX 7 (Beckman)	Hitachi/Olympus
Urea	CX 7 (Beckman)	CX 7 (Beckman)
Platelets	LH 750 (Coulter)	STKS (Beckman-Coulter)/LH750 (Beckman-Coulter[b]
Prothrombin index	STAR	STA/STAR (neoplastine CI Diagnostica Stago)[c]
Apolipoprotein A1	BN Prospec (Dade Behring)	BN Prospec (Dade Behring)
Haptoglobin	BN Prospec (Dade Behring)	Array, Immage (Beckman-Coulter)
α-2-Macroglobulin[d]	BN Prospec (Dade Behring)	BN Prospec (Dade Behring)
Hyaluronate[d]	EIA Corgenix	EIA Corgenix

[a] Hitachi until May 31, 2003.
[b] SKTS until November 11, 2002.
[c] STA until December 13, 2003.
[d] Centralized dosage.

Please cite this article in press as: Halfon P et al., Comparison of test performance profile for blood tests ..., J Hepatol (2006), doi:10.1016/j.jhep.2006.09.020

The pair of expert pathologists was different in each center including a local pathologist and one out of two external pathologists from Beaujon Hospital for each center, i.e. a total of four pathologists. The rate of initial observer disagreement was 38.8% in Provence area and 44.3% in Tours ($p = 0.30$).

2.2.3. Statistical analysis

2.2.3.1. Statistical tests. Quantitative variables were expressed as means ± SD, unless otherwise specified. The prediction (or performance) of each model is expressed either by the diagnostic accuracy (DA), i.e. true positives and negatives, and by the area under the receiving–operating-characteristic (AUROC). AUROCs were compared by the Hanley–McNeil method for paired data [12]. Agreement of qualitative variables was evaluated by the kappa index. Binary logistic regression with forward stepwise inclusion of independent variables was used. The cut-off of regression score was determined according to the highest Youden index (Se + Spe − 1) to provide the best DA. The statistical software used was SAS V8.2 (SAS Institute Inc., Cary, NC, USA).

2.2.3.2. Sample size calculation. The size of population was determined to show a significant difference between FM test and FT [5]. With α risk: 0.05, β risk: 0.2, clinically significant fibrosis prevalence: 0.5, AUROC correlation: 0.70, and bilateral test, the sample size was 360 patients for the following hypothesis of AUROC: FM: 0.85, FT: 0.80.

3. Results

3.1. General characteristics

These characteristics are presented in Table 2. The two populations were significantly different: the population from Tours was younger with less severe hepatitis and had a lower prevalence of significant fibrosis (31 vs 49%, $p = 0.001$) compared to the population from Provence-Côte d'Azur. The date of liver biopsy was significantly more recent in the Provence-Côte d'Azur population.

3.2. Test performance

Indices are presented in Tables 3–5, respectively, for significant fibrosis, severe fibrosis and cirrhosis. Significant fibrosis includes F2 + F3 + F4, severe fibrosis includes F3 + F4, and cirrhosis includes F4. The statistical comparison of performance between tests is classically only available with AUROC and is presented in Table 6. There was no significant difference between the four tests for the three diagnostic targets.

3.3. Pattern of test according to F stage

Fig. 1 depicts the box plots of blood tests for the probability of significant fibrosis as a function of Metavir fibrosis stage. It clearly shows a different pattern of test probability. FT, FM and HS had a similar pattern whereas APRI showed a narrower range for F0 and F1. F2 patterns were similar for the four tests. F3 pattern indicated a trend which was more evident in F4 stage: tests including hyaluronic acid (FM and HS) displayed a higher probability value and a narrower range than others which was especially evident for FM in F4. These observed differences lead us to compare the performance within each Metavir F stage. This was made

Table 2
General characteristics of patient populations

	Whole population	Provence-Côte d'Azur	Tours	*p*
No. of patients	356	198	158	–
Sex (% male)	189 (53)	99 (50)	90 (57)	0.23
Age (year)	44.9 ± 12.9	48.1 ± 13.7	40.9 ± 10.5	0.01
Metavir fibrosis stage				
F0 (%)	15 (4)	9 (4)	6 (4)	0.79
F1 (%)	195 (55)	92 (48)	103 (65)	0.002
F2 (%)	95 (26)	61 (31)	34 (22)	0.08
F3 (%)	38 (11)	25 (12)	13 (8)	0.29
F4 (%)	13 (4)	11 (6)	2 (1)	0.07
Clinically significant fibrosis (%)	146 (41)	97 (49)	49 (31)	0.001
Severe fibrosis (%)	51 (14)	36 (18)	15 (9.5)	0.02
Metavir fibrosis score	1.5 ± 0.9	1.7 ± 0.9	1.4 ± 0.7	0.001
Liver specimen length (mm)	22.0 ± 7.1	23.3 ± 7.3	20.5 ± 6.4	0.04
Median liver biopsy date	15 January 2003	20 May 2003	8 April 1999	<.0001
Platelets (G/L)	211.2 ± 65.0	203.2 ± 60.3	221.2 ± 69.2	0.64
Prothrombin index (%)	92.8 ± 7.6	92.1 ± 7.7	93.7 ± 7.4	0.03
ASAT (UI/L)	49.3 ± 37.3	57.3 ± 42.7	39.2 ± 26.0	<.0001
ALAT (UI/L)	76.5 ± 66.2	77.7 ± 71.3	74.9 ± 59.3	0.99
GGT (UI/L)	67.1 ± 108.7	75.0 ± 117.3	57.3 ± 96.5	0.10
Bilirubin (μmol/L)	10.5 ± 5.3	9.9 ± 5.5	11.0 ± 5.0	0.03
Urea (mmol/L)	5.8 ± 1.6	6.0 ± 1.8	5.6 ± 1.3	0.11
Apolipoprotein A1 (g/L)	1.5 ± 0.4	1.5 ± 0.4	1.4 ± 0.4	0.05
Haptoglobin (g/L)	0.94 ± 0.45	0.93 ± 0.45	0.96 ± 0.46	0.59
α-2-Macroglobulin (mg/dL)	287 ± 111	272 ± 91	305 ± 130	0.10
Hyaluronate (μg/L)	52.4 ± 96.2	64.9 ± 114.7	36.7 ± 63.2	<.0001

Please cite this article in press as: Halfon P et al., Comparison of test performance profile for blood tests ..., J Hepatol (2006), doi:10.1016/j.jhep.2006.09.020

Table 3
Performance of blood tests for significant fibrosis

Test	Cut-off	Se	Spe	Positive PV	Negative PV	Positive LR	Negative LR	DA	AUROC (95% CI)
FibroMeter	0.57	64	81	70	77	3.30	0.44	74	0.78 (0.73;0.82)
Fibrotest	0.44	67	80	70	78	3.36	0.41	74	0.79 (0.75;0.83)
Hepascore	0.32	77	63	59	80	2.09	0.37	69	0.76 (0.71;0.80)
APRI	0.39	77	66	61	80	2.24	0.35	70	0.76 (0.72;0.81)

Se, sensitivity; Spe, specificity; PV, predictive value; LR, likelihood ratio; DA, diagnostic accuracy; AUROC, area under the receiving–operating-characteristic.

Table 4
Performance of blood tests for severe fibrosis

Test	Cut-off	Se	Spe	Positive PV	Negative PV	Positive LR	Negative LR	DA	AUROC (95% CI)
FibroMeter	0.67	82	76	37	96	3.44	0.23	77	0.84 (0.80;0.88)
Fibrotest	0.45	84	69	31	96	2.74	0.23	71	0.81 (0.77;0.85)
Hepascore	0.53	78	72	32	95	2.78	0.30	73	0.81 (0.76;0.85)
APRI	0.58	75	76	34	95	3.11	0.34	76	0.81 (0.76;0.85)

Se, sensitivity; Spe, specificity; PV, predictive value; LR, likelihood ratio; DA, diagnostic accuracy; AUROC, area under the receiving–operating-characteristic.

Table 5
Performance of blood tests for cirrhosis

Test	Cut-off	Se	Spe	Positive PV	Negative PV	Positive LR	Negative LR	DA	AUROC (95% CI)
FibroMeter	0.88	92	87	21	100	7.20	0.09	87	0.94 (0.91;0.96)
Fibrotest	0.56	85	74	11	99	3.19	0.21	74	0.86 (0.82;0.89)
Hepascore	0.61	92	72	11	100	3.30	0.11	73	0.89 (0.86;0.92)
APRI	0.83	100	83	18	100	5.81	0.00	83	0.92 (0.88;0.94)

Se, sensitivity; Spe, specificity; PV, predictive value; LR, likelihood ratio; DA, diagnostic accuracy; AUROC, area under the receiving–operating-characteristic.

Table 6
Comparison of AUROC between blood tests (*p* value of Hanley–McNeil test)

	Significant fibrosis (≥F2)	Severe fibrosis (≥F3)	Cirrhosis (F4)
FM vs FT	0.67	0.32	0.09
FM vs HS	0.28	0.19	0.19
FM vs APRI	0.46	0.23	0.63
FT vs APRI	0.35	0.89	0.40
FT vs HS	0.08	0.91	0.40
HS vs APRI	0.84	0.95	0.73

FM, FibroMeter; FT, Fibrotest; HS, Hepascore.

possible by first calculating the disagreement rate between each blood test and liver specimen for the diagnosis of significant fibrosis. Then, the disagreement rates were compared between a pair of blood tests by the McNemar test for paired comparison. We focused the presentation of test performance profile to the two tests with the nearer and highest AUROCs, i.e. FM and FT. Table 7 shows a fair agreement between both tests. More importantly, it clearly shows a significantly different test performance profile. FT was superior to FM for

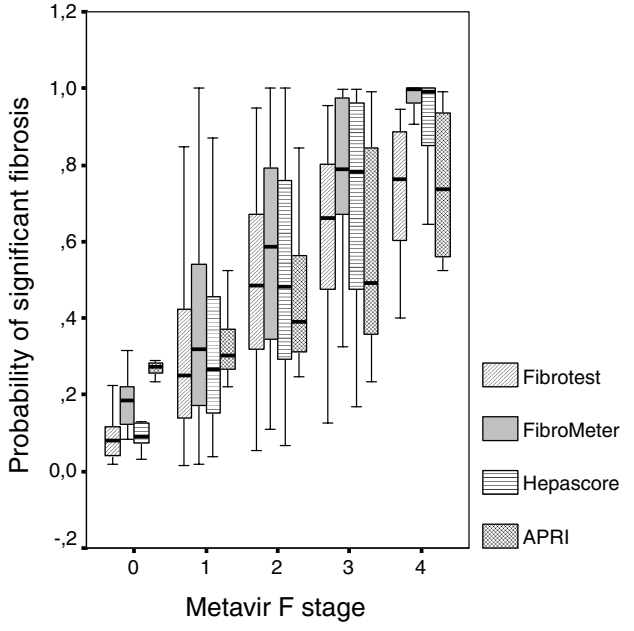

Fig. 1. Box plots (median, quartiles and extremes) of blood tests as a function of Metavir fibrosis stage for probability of significant fibrosis. APRI was normalized using logistic regression.

Table 7
Rate (%) of misclassified patients for significant fibrosis according to Metavir fibrosis stage for FT and FM

	Metavir fibrosis stage								
	0	1	2	3	4	<2	⩾2	⩾3	All
No. of patients	15	195	95	38	13	210	146	51	356
Fibrotest	0	19.0	51.6	28.9	15.4	17.6	42.5	25.5	27.8
FibroMeter	0	29.7	41.1	15.8	0	27.6	30.8	11.8	28.9
p (McNemar)	–[a]	0.0003	0.03	0.18	–[a]	0.0003	0.004	0.065	0.734
Kappa	–[a]	0.383	0.539	0.335	–[a]	0.396	0.520	0.310	0.461

[a] Calculation impossible.

F1 stage whereas FM was superior to FT for F2 or F3 or F4 stages. Finally, FT was significantly lower than FM for the misclassification rate in patients without significant fibrosis (18 vs 28%, respectively, $p = 0.0003$) whereas FM was significantly lower than FT for the misclassification rate in patients with significant fibrosis (31 vs 42.5%, respectively, $p = 0.004$). Fig. 2 clearly shows the different test performance profiles between FT and FM.

3.4. Center effect

The differences of test performances according to each center are shown for AUROC in Table 8. Values were similar except for F4 but only 2 patients had cirrhosis in one center which precluded a reliable comparison. It should be stated that there is no statistical test to compare unpaired AUROC. Therefore, the center effect was tested using forward logistic regression by including the variables significantly different between centers as independent variables and disagreement for significant fibrosis between blood tests and liver specimen as dependent variable. For FM, the only independent variable was age ($p = 0.002$) whereas center had no significant effect ($p = 0.90$). For FT, the only independent variable was Metavir F stage ($p = 0.001$) whereas center had no significant effect ($p = 0.43$) with a borderline effect for age ($p = 0.058$) and sex ($p = 0.088$). So, the center had no independent effect on test performance, the differences between centers being due to demographic differences, age or fibrosis stage, both being significantly related ($p < 10^{-4}$ by ANOVA).

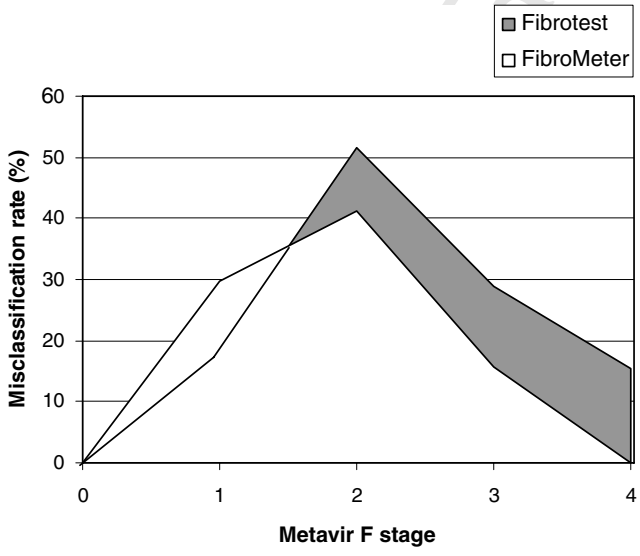

Fig. 2. Profile of test performance: comparison of misclassification rate for significant fibrosis between FT and FM.

3.5. Effect of liver specimen size

There was no correlation between length of liver specimen and Metavir F stage ($r = -0.065$, $p = 0.22$). AUROCs were evaluated as a function of the liver specimen median (21.0 mm) in Table 9. AUROCs were not improved in the largest specimens.

4. Discussion

4.1. Comparison with previous studies

APRI and FT have been evaluated in several independent studies [5,13]. However, truly independent studies mainly devoted to validation are rare [14,15]. In the present study, a lower performance was observed for the 4 blood tests (AUROC for significant fibrosis: 0.76–0.79) than in the original publications where AUROCs were 0.80–0.88 (test-validation) for APRI [2], 0.85 for HS [6], 0.84–0.87 for FT [3] and 0.88-0.91 for FM [5]. But our results fit well with the 0.77 median AUROC reported for 10 different blood tests in validation populations vs 0.81 in training populations [7]. So, the expected difference between FT and FM [5] was not observed. The main cause was a less severe fibrosis stage and a younger age. This was observed in one center compared to the other. In addition, the prevalence of significant fibrosis in both centers was low: 41% here vs 56% in the original FM study [5] and 49% in the initial FT study [3], 44–57% in the original HS study [6] and 47–50% in the original APRI study [2]. The interaction between this prevalence and AUROC, depicted in the present study, shows that the sample size calculation is hindered by this confusing factor.

Please cite this article in press as: Halfon P et al., Comparison of test performance profile for blood tests ..., J Hepatol (2006), doi:10.1016/j.jhep.2006.09.020

Table 8
Comparison of AUROC for significant fibrosis between the two centers

Center	Significant fibrosis (⩾F2)		Severe fibrosis (⩾F3)		Cirrhosis (F4)	
	PACA	Tours	PACA	Tours	PACA	Tours
No. of patients with diagnostic target	97	49	36	15	11	2
FibroMeter	0.767	0.773	0.852	0.800	0.911	1
Fibrotest	0.805	0.789	0.815	0.817	0.844	0.920
Hepascore	0.774	0.758	0.819	0.798	0.893	1
APRI	0.762	0.736	0.804	0.772	0.872	0.926

No statistical comparison available for unpaired data.
PACA, Provence-Côte d'Azur.

In the present study, we have implemented a new method to compare the test performance. The test performance profile relies on the comparison of misclassification rates, based on liver specimens, that are tested by a paired McNemar test. This method, only applied here to FM and FT, clearly shows that the test performance depends on the fibrosis stage (Table 7). Thus, both FT and FM displayed a peak of misclassification rate for F2 (Fig. 2). Moreover, the test performance profile was different since this peak was significantly lower for FM and the values were significantly lower with FM at the right of this peak, i.e. for all stages of clinically significant fibrosis. This explains that the difference in test performance between different blood tests depends on the relative prevalence of fibrosis stages (e.g., a population with a high prevalence of clinically significant fibrosis or severe fibrosis will favour FM compared to FT and *vice versa*).

Finally, the box plots of blood tests against fibrosis stage clearly show the diagnostic robustness of a test for a fibrosis stage, e.g. the range of FM for F4 is very narrow (Fig. 1) that suggests the robustness of this test to correctly classify F4 stage in different populations. In addition, the comparison of box plots shows that the score probabilities of different tests have not a similar prediction for the fibrosis stage, e.g. the median value of FM for F3 is slightly superior to that of FT for F4. Thus, the numerical values of different tests are not comparable.

4.2. Factors influencing the test performance

We have previously shown that the test performance depends on the fibrosis stage. In addition, the performance of tests globally depends on the diagnostic target as determined by Metavir F stage cut-off (Tables 3–5): the greater the diagnostic target, the higher the observed performance. Thus, on one hand, the absolute test performance increased as a function of the diagnostic target. On the other hand, despite a low prevalence of significant fibrosis in the present study, the relative FM performance, compared to that of FT, increased as a function of the diagnostic target. These observations depict, with different shades, a test performance profile as clearly observed in Fig. 2. This profile should be adjusted. Indeed, this profile is based on a liver specimen as reference. However, the reliability of such a reference, as shown by the interobserver disagreement, displays the same V pattern [16] suggesting that a significant part of the misclassification rate is attributable to misclassification by liver biopsy interpretation itself. Nevertheless, this profile is correct in interpreting the differences between blood tests. Thus, both FM and FT are characterized by a 100% accuracy in F0 whereas this figure is attained in F4 only with FM.

This new descriptor of test performance can be applied to compare different populations. Thus, the prevalence of fibrosis stage depends on the cause of chronic liver disease, e.g. in co-infections with HIV, the prevalence of F0 is usually low and that of F4 is high whereas the test performance might depend on specific variation of blood markers in HIV population [17,18] thus confusing the comparison of global test performance between mono-infected and co-infected populations. The use of the test profile contributes to clarify the comparison by deleting this confounding factor and making possible a direct comparison of test performance for each fibrosis stage.

Table 9
AUROC as a function of the liver specimen median according to diagnostic targets

	Significant fibrosis		Severe fibrosis		Cirrhosis	
Liver specimen length (mm)	<21	⩾21	<21	⩾21	<21	⩾21
No. of patients with diagnostic target	67	79	29	22	10	3
FibroMeter	0.798	0.757	0.854	0.830	0.952	0.942
Fibrotest	0.799	0.787	0.851	0.765	0.893	0.750
Hepascore	0.770	0.746	0.828	0.788	0.910	0.864
APRI	0.753	0.775	0.826	0.787	0.940	0.875

Please cite this article in press as: Halfon P et al., Comparison of test performance profile for blood tests ..., J Hepatol (2006), doi:10.1016/j.jhep.2006.09.020

This study also shows that there was no center effect on the AUROC of the four blood tests and no significant center effect on the misclassification rate. In fact, this misclassification rate was independently influenced by patient age or the fibrosis stage which were significantly different as a function of centers. This clearly suggests that the difference between centers was not due to difference in analytical characteristics of blood determinations. The lower prevalence of significant fibrosis in the two centers compared to previous studies [2–5] can be attributed either to the centers themselves, e.g. the Provence-Côte d'Azur area included units closer to the general population of chronic HCV patients cared for by hepato-gastroenterologists. It can be also attributed to the more recent inclusion period compared to original studies since the degree of fibrosis stage decreases as a function of dates [19].

The blood test accuracy is also limited by the sampling error of liver biopsy. This could be partially circumvented by using only the largest liver specimens. However, in the present study, there was no correlation between length of liver specimen and Metavir F stage unlike in other studies [16,20]. Moreover, AUROCs were not improved in largest specimens (Table 8). Finally, recent studies have suggested that blood markers of liver fibrosis combined either in a stepwise algorithm [21] or to elastometry [22] might improve the diagnostic performance.

4.3. What should be the ideal test profile?

Of course an ideal test would have a 100% diagnostic accuracy. However, this will be very difficult to achieve as long as liver biopsy remains the reference. Meanwhile, if we have to choose a blood score among others, the test performance profile described here contributes to facilitate that choice. Indeed, the test profile suggests that a global performance like diagnostic accuracy or AUROC is not sufficient for comparison since it is too dependent on the prevalence of fibrosis stage. The inverse significant differences in misclassification rate between FT and FM for the low or the high-fibrosis stages validate the use of a test performance profile depicting the accuracy (or it inverse: the misclassification rate) as a function of fibrosis stage. In the present case, we have the choice between a test overestimating low stages and better classifying high stages of fibrosis and a test better classifying the early stage and underestimating high stages of fibrosis. We consider that the first profile is convenient for a diagnostic goal, since it is important to correctly classify severe fibrosis or cirrhosis for complication screening, as well in a screening perspective since overestimation means higher sensitivity which is suitable for a fibrosis screening. However, this last goal has to be validated in an adequate population. Recently, improved test reporting was recommended for the description of blood tests including likelihood ratios and diagnostic odds ratio [7]. We suggest that this test performance profile should be added to this description. Indeed, test performance profile allows two new kinds of description of diagnostic accuracy for a given diagnostic target and a test based on a semiquantitative classification: first, detailed accuracy as a function of stages for a single test; second, comparison between two tests using diagnostic accuracy either of all the stages (global test) and/or stage by stage of fibrosis. We suggest that test performance profile should be used in addition to AUROC, positive and negative predictive values, especially as main descriptor in statistical comparisons besides AUROC. But both are limited by paired comparison, i.e. restricted to the same sample.

In conclusion, this study shows a good performance of the four blood tests but lower than in the original publications. We have used a new method to compare the diagnostic accuracy as a function of fibrosis stage. This pattern is called test performance profile and shows significant differences in diagnostic accuracy between blood tests as a function of fibrosis stage. This test performance profile should be added in the description of test performance especially in comparative studies. Finally, the test performance should be reported for different diagnostic targets.

Acknowledgement

We thank other investigators: Christophe Sattonnet, Marie-Christine Saint-Paul, Thierry Benderitter, Stéphane Garcia, Henri Pierre Bonneau, and Romaric Deydier.

References

[1] Sebastiani G, Alberti A. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. World J Gastroenterol 2006;12:3682–3694.
[2] Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38:518–526.
[3] Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poynard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. Lancet 2001;357:1069–1075.
[4] Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, Hubscher S, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004;127:1704–1713.
[5] Cales P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A, et al. A novel panel of blood markers to assess the degree of liver fibrosis. Hepatology 2005;42:1373–1381.
[6] Adams LA, Bulsara M, Rossi E, DeBoer B, Speers D, George J, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. Clin Chem 2005;51:1867–1873.
[7] Parkes J, Guha IN, Roderick P, Rosenberg W. Performance of serum marker panels for liver fibrosis in chronic hepatitis C. J Hepatol 2006;44:462–474.

[8] Leroy V, Monier F, Bottari S, Trocme C, Sturm N, Hilleret MN, et al. Circulating matrix metalloproteinases 1, 2, 9 and their inhibitors TIMP-1 and TIMP-2 as serum markers of liver fibrosis in patients with chronic hepatitis C: comparison with PIIINP and hyaluronic acid. Am J Gastroenterol 2004;99:271–279.
[9] Patel K, Gordon SC, Jacobson I, Hezode C, Oh E, Smith KM, et al. Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients. J Hepatol 2004;41:935–942.
[10] Poordad FF. FIBROSpect II: a potential noninvasive test to assess hepatic fibrosis. Expert Rev Mol Diagn 2004;4:593–597.
[11] The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. Hepatology 1994;20:15–20.
[12] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–843.
[13] Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 2005;128:343–350.
[14] Halfon P, Bourliere M, Deydier R, Botta-Fridlund D, Renou C, Tran A, et al. Independent prospective multicenter validation of biochemical markers (fibrotest–actitest) for the prediction of liver fibrosis and activity in patients with chronic hepatitis C: the fibropaca study. Am J Gastroenterol 2006;101:547–555.
[15] Lackner C, Struber G, Liegl B, Leibl S, Ofner P, Bankuti C, et al. Comparison and validation of simple noninvasive tests for prediction of fibrosis in chronic hepatitis C. Hepatology 2005;41:1376–1382.
[16] Rousselet MC, Michalak S, Dupre F, Croue A, Bedossa P, Saint-Andre JP, et al. Sources of variability in histological scoring of chronic viral hepatitis. Hepatology 2005;41:257–264.
[17] Macias J, Giron-Gonzalez JA, Gonzalez-Serrano M, Merino D, Cano P, Mira JA, et al. Prediction of liver fibrosis in human immunodeficiency virus/hepatitis C virus coinfected patients by simple non-invasive indexes. Gut 2006;55:409–414.
[18] Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317–1325.
[19] Guyader D, Lefeuvre C, Jacquelinet S, Prat M, Baudouard Y, Turlin B, et al. Epidemiology of hepatitis C virus infection in 1,304 HCV positive patients: variations according to the origin of transmission and year of diagnosis. Gastroenterol Clin Biol 1998;22:375–380.
[20] Colloredo G, Guido M, Sonzogni A, Leandro G. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. J Hepatol 2003;39:239–244.
[21] Sebastiani G, Vario A, Guido M, Noventa F, Plebani M, Pistis R, et al. Stepwise combination algorithms of non-invasive markers to diagnose significant fibrosis in chronic hepatitis C. J Hepatol 2006;44:686–693.
[22] Corpechot C, El Naggar A, Poujol-Robert A, Ziol M, Wendum D, Chazouilleres O, et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. Hepatology 2006;43:1118–1124.

Please cite this article in press as: Halfon P et al., Comparison of test performance profile for blood tests ..., J Hepatol (2006), doi:10.1016/j.jhep.2006.09.020