

Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: A multicenter prospective study (the FIBROSTIC study)

Francoise Degos^{1,*}, Paul Perez², Bruno Roche³, Amel Mahmoudi⁴, Julien Asselineau², H el ene Voitot⁵, Pierre Bedossa⁶, for the FIBROSTIC study group[†]

¹APHP, H opital Beaujon, Hepatology Dept., INSERM U 773, Clichy, France; ²CHU de Bordeaux, Clinical Epidemiology Unit & CIC-EC7, Bordeaux, France; ³APHP, CHB, H opital Paul Brousse, INSERM Paris Sud U 785, Villejuif, France; ⁴APHP, H opital Jean Verdier, Bondy, France; ⁵APHP, H opital Beaujon, Biochemistry Dept., Clichy, France; ⁶APHP, H opital Beaujon, Pathology Dept., INSERM U 773, Clichy, France

Background & Aims: The diagnostic accuracy of non-invasive liver fibrosis tests that may replace liver biopsy in patients with chronic hepatitis remains controversial.

We assessed and compared the accuracy of FibroScan[®] and that of the main biomarkers used for predicting cirrhosis and significant fibrosis (METAVIR \geq F2) in patients with chronic viral hepatitis.

Methods: A multicenter prospective cross-sectional diagnostic accuracy study was conducted in the Hepatology departments of 23 French university hospitals. Index tests and reference standard (METAVIR fibrosis score on liver biopsy) were measured on the same day and interpreted blindly. Consecutive patients with chronic viral hepatitis (hepatitis B or C virus, including possible Human Immunodeficiency Virus co-infection) requiring liver biopsy were recruited in the study.

Results: The analysis was first conducted on the total population (1839 patients), and after excluding 532 protocol deviations, on 1307 patients (non-compliant FibroScan[®] examinations). The overall accuracy of FibroScan[®] was high (AUROC 0.89 and 0.90, respectively) and significantly higher than that of biomarkers in predicting cirrhosis (AUROC 0.77–0.86). All non-invasive methods had a moderate accuracy in predicting significant fibrosis (AUROC 0.72–0.78). Based on multilevel likelihood ratios, non-invasive tests provided a relevant gain in the likelihood of diagnosis in 0–60% of patients (cirrhosis) and 9–30% of patients (significant fibrosis).

Conclusions: The diagnostic accuracy of non-invasive tests was high for cirrhosis, but poor for significant fibrosis. A clinically relevant gain in the likelihood of diagnosis was achieved in a low proportion of patients. Although the diagnosis of cirrhosis may rely on non-invasive tests, liver biopsy is warranted to diagnose intermediate stages of fibrosis.

Keywords: Liver fibrosis non-invasive tests; Chronic hepatitis B; Chronic hepatitis C; Liver biopsy.

Received 15 February 2010; received in revised form 5 May 2010; accepted 22 May 2010; available online 14 August 2010

* Corresponding author. Address: Hepatology Dept., H opital Beaujon, 100 bd G en eral Leclerc, 92110 Clichy, France.

E-mail address: francoise.degos@bjn.aphp.fr (F. Degos).

[†] see Appendix 1 for the participating centers.

  2010 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Diagnosis and treatment of patients with chronic hepatitis mostly rely on the staging of liver fibrosis. Antiviral therapy is proposed if moderate to severe (METAVIR stages F2 and F3) [1] fibrosis is present. If cirrhosis is present, specific surveillance is initiated, in particular for the early detection of hepatocellular carcinoma. Despite its limitations, liver biopsy is the usual procedure for staging fibrosis and is recommended by the international guidelines [2,3].

Non-invasive procedures such as transient elastography (FibroScan[®], Echosens, Paris, France) and serum biomarkers (particularly Fibrometre[®], Fibrotest[®], Hepascore and APRI) have been developed in order to avoid biopsy. Transient elastography is a new imaging technique measuring liver stiffness, i.e. its elasticity. As highlighted by recent meta-analyses [4], its diagnostic accuracy is difficult to judge because of the small sample sizes of most published studies and because of the inter-study variability of accuracy estimates and elasticity thresholds, varying in the literature from 8.4 to 18.2 kPa for the diagnosis of cirrhosis and from 5.0 to 11.8 for the diagnosis of significant fibrosis (METAVIR fibrosis score \geq F2) [4,5]. In addition, there are few direct comparisons of transient elastography and of serum biomarkers [6] and the reported accuracy parameters do not always allow clinicians to estimate the probability of a condition in a given patient [7].

We evaluated the diagnostic accuracy and clinical usefulness of FibroScan[®] in predicting two conditions – significant histological liver fibrosis (METAVIR \geq F2) and cirrhosis (F4) – in patients with chronic viral hepatitis, in a multicenter prospective study funded by the French Ministry of Health (the FIBROSTIC study). The secondary objectives were: (i) to assess variations in the diagnostic accuracy of FibroScan[®] by cause of infection and by patient selection criteria for liver biopsy; (ii) to compare the diagnostic accuracy of the most common liver fibrosis serum biomarkers with that of transient elastography; (iii) to assess the gain in likelihood of target conditions provided by non-invasive tests.



Research Article

Patients and methods

Design overview, setting, and participants

This was a multicenter prospective cross-sectional diagnostic accuracy study. From June 15, 2006, to July 15, 2008, the hepatology departments of 23 French university hospitals included all consecutive patients with chronic viral hepatitis due to hepatitis C virus (HCV) or hepatitis B virus (HBV) (with or without Human Immunodeficiency Virus – HIV co-infection) in whom liver fibrosis assessment was indicated. Chronic hepatitis C was defined as the presence of anti-HCV antibodies for more than 6 months and positivity (except in documented successfully treated patients) of HCV RNA (Amplicor HCV Monitor[®]). Chronic hepatitis B was defined as the presence of serum HBsAg (commercially available enzyme immunoassays) for more than 6 months. Each hospital applied their usual criteria (clinical criteria or discordant results in prior non-invasive tests) to determine whether liver biopsy was indicated.

Patients underwent liver biopsy and non-invasive tests the same day or within a maximum of 30 days. Raters of each test were blind to the results of the other tests. The diagnostic accuracy of tests was estimated versus liver biopsy as the reference standard. In the accuracy comparisons, patients with HCV and HBV co-infection were included in the HCV infection subgroup, and those with HIV co-infection in the HIV subgroup. The protocol was approved by the Institutional Review Board (CCPPRB, Hôpital Pitié-Salpêtrière). Participants gave their informed consent.

Liver biopsy

Liver biopsy procedures [8] were performed by experienced hepatologists or radiologists according to current French practice [9], either by the transjugular, or intercostal approach under ultrasound guidance using disposable 1.8 mm diameter Menghini needles. The liver tissue was processed routinely and read on site by expert liver pathologists. The length of each liver specimen, the number of fragments, and portal tracts were recorded. Fibrosis stage and necro-inflammatory activity were evaluated using METAVIR scores [1].

Transient elastography

Transient elastography was performed by physicians or trained technicians with experience of at least 50 transient elastography procedures, as previously recommended [6] (Appendix 2). At least 10 valid measurements, a 60% success rate, an interquartile range of less than 30% of the median elasticity, and a Body Mass Index (BMI) <30 kg/m² were required for eligibility [6]. Elasticity thresholds were chosen on study data, for achieving 90% sensitivity in the prediction of significant fibrosis and 90% specificity in the prediction of cirrhosis. One of the thresholds previously published and commonly proposed was also used [6].

Serum biomarker assays

The usual liver function tests were performed on site. Frozen sera (–80 °C) were collected for other assays to be included in serum biomarkers scores (Appendix 3). The following recommended cut-off levels were used to define a positive test result for significant fibrosis and cirrhosis, respectively: Fibrometre[®] (0.411 and 0.442 [10], Fibrotest[®] (: 0.48 and 0.74 [11], APRI 0.5 and 2 [12], 0.50 and 0.84 [13]. Hepascore and APRI were calculated from biological results, Fibrotest[®], and Fibrometre[®] are licensed (cost 50 euros per test) and were calculated by the courtesy of the manufacturers (Biopredictive, BioLiveScale). Biochemical scores could not be calculated in patients with at least one biological value missing, and when they were considered by the manufacturer as non-interpretable, they were not included in the analysis.

Statistical analysis

The diagnostic accuracy parameters of the non-invasive tests were estimated by comparison with liver biopsy used as the reference standard. Multilevel likelihood ratios (LR) and post-test probabilities of significant fibrosis and cirrhosis were calculated for the deciles of elasticity and biomarker scores. Published reference values of LR [14] were used to establish test result categories. Confidence intervals (95% CI) were calculated by the exact binomial method for proportions, and by the asymptotic Gaussian distribution method for AUCs [15]. Variations in diagnostic accuracy according to patient subgroup (viral etiology, patient selection for liver biopsy) were investigated by comparing AUCs [15], specificities

for the diagnosis of significant fibrosis, and sensitivities for the diagnosis of cirrhosis (Chi-square tests). If there was no statistically significant variation of accuracy according to viral etiology, subsequent analyses were not stratified according to this characteristic. Differences in the diagnostic accuracy of transient elastography and biomarkers were studied by comparing AUCs (DeLong et al.'s method for correlated data) [15]. Agreement between METAVIR scores, as measured by each center and by the central pathology laboratory was assessed using Kappas coefficients (Appendix 4). Type I error was fixed at 0.01 for comparing subgroups or non-invasive methods and at 0.05 for other comparisons. Analyses were conducted with SAS v. 9.1.3 and Stata v. 9.2 softwares.

Results

Flow diagram and patients' characteristics

Among the 1839 potentially eligible patients, 532 (29%) were not eligible, mostly because their transient elastography examinations did not comply with the recommendations for high measurement reproducibility [16] (Fig. 1). The more frequent reason of exclusion was an interquartile range of elasticity >30%

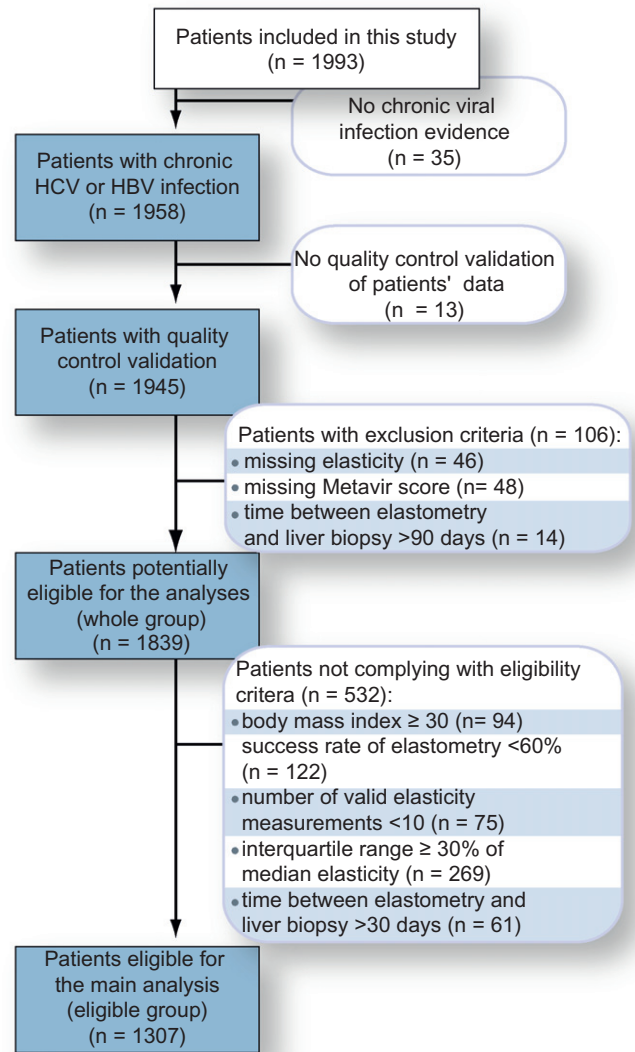


Fig. 1. Flow chart, FIBROSTIC study, France, 2006–2008.

of the median (269 patients, 14%). We conducted the main analysis on the group of patients complying with the adequate technique, but results on the whole group were analyzed in a sensitivity analysis.

The clinical and histopathological characteristics of the 1307 eligible patients are given in Table 1 according to their viral etiology; they were found to have similar characteristics to those of the whole group: median age 47.2 vs. 47.6, male 69.2% vs. 67.2%, median BMI 23.8 vs. 24.3, mean ALT 2.4 ± 3.1 vs. 2.4 ± 3.1 , significant fibrosis 57.1% vs. 55.3%, cirrhosis 13.8% vs. 13.3%.

Liver biopsy

The median size of liver biopsies was 24 mm (1st–3rd quartiles: 18–30) with 71% of the biopsies measuring at least 20 mm. The

median number of portal tracts was 16 (10–22). The reproducibility of the METAVIR fibrosis score is reported in Appendix 4.

Diagnostic accuracy of transient elastography

The results of stiffness according to the five stages of fibrosis in the total group and in the group of eligible patients are given in Fig. 2A and B.

The median number of measurements was 10 (10–10) and the IQR/stiffness ratio was 15% (11–20). Median stiffness was in the same range for F0, F1, and F2 patients and slightly higher in F3 patients (Fig. 2). Patients with cirrhosis had markedly higher stiffness values. The distribution of stiffness values differed according to METAVIR stage ($p < 0.0001$).

Results for the diagnostic accuracy of TE for eligible patients are shown in Fig. 3A–D and Table 2a and b). The AUC in the

Table 1. Clinical, biochemical, and histological data of patients eligible for the main analysis, according to viral infection subgroups, FIBROSTIC study, France, 2006–2008.

	HCV		HBV		HIV		All	
Number of patients	913		284		110		1307	
Age (years)	50.2	(10.8)	38.2	(12.1)	45.3	(6.3)	47.2	(11.8)
Male gender	591	(64.7)	230	(81.0)	83	(75.5)	904	(69.2)
BMI	24.0	(2.9)	23.6	(2.8)	22.2	(2.8)	23.8	(2.9)
Daily alcohol use	138	(15.1)	26	(9.2)	11	(10.0)	175	(13.4)
Active tobacco use	326	(35.7)	47	(16.5)	58	(52.7)	431	(33.0)
Clinical cirrhosis	116	(12.7)	21	(7.4)	17	(15.5)	154	(11.8)
Complicated	36	(31.0)	4	(19.0)	5	(29.4)	45	(29.2)
Transplant recipients	115	(12.6)	7	(2.5)	42	(38.2)	164	(12.5)
HCV or HBV antiviral therapy	416	(45.6)	60	(21.1)	61	(55.5)	537	(41.1)
HCV RNA positive*	841	(92.1)	-	-	84	(84.0)	925	(91.3)
ALT (x ULN)	2.4	(2.2)	2.4	(4.7)	2.8	(4.1)	2.4	(3.1)
AST (x ULN)	2.0	(1.6)	1.7	(2.6)	2.5	(2.6)	2.0	(2.0)
Total bilirubin (μ M)	14.8	(24.6)	13.6	(8.6)	21.5	(18.7)	15.1	(21.7)
Prothrombin (%)	93	(12)	88	(12)	89	(13)	92	(12)
Platelets ($10^3/\text{mm}^3$)	201.8	(69.2)	200.6	(67.7)	166.5	(73.6)	198.6	(69.9)
Albumin (g/L)	41.0	(4.9)	41.5	(4.3)	38.3	(5.7)	40.9	(4.9)
Urea (mM)	5.90	(4.55)	5.16	(2.18)	6.85	(5.66)	5.82	(4.29)
GGT (U/L)	132	(185)	57	(73)	246	(422)	126	(205)
Haptoglobin (g/L)	0.9	(0.5)	0.8	(0.6)	0.8	(0.6)	0.8	(0.5)
Alpha 2 macro-globulin (g/L)	2.99	(1.02)	2.45	(0.88)	2.87	(1.21)	2.86	(1.03)
Apolipoprotein A1 (g/L)	1.53	(0.36)	1.54	(0.33)	1.32	(0.34)	1.51	(0.36)
Hyaluronic acid (μ g/L)	107.7	(408.1)	42.1	(49.9)	153.3	(209.7)	97.7	(349.8)
METAVIR fibrosis								
F0	43	(4.7)	54	(19.0)	16	(14.5)	113	(8.6)
F1	308	(33.7)	112	(39.4)	28	(25.5)	448	(34.3)
F2	277	(30.3)	62	(21.8)	32	(29.1)	371	(28.4)
F3	159	(17.4)	27	(9.5)	8	(7.3)	194	(14.8)
F4	126	(13.8)	29	(10.2)	26	(23.6)	181	(13.8)
METAVIR activity								
A2-A3	407	(44.6)	98	(34.5)	51	(46.4)	556	(42.5)
Steatosis \geq 30%	169	(18.5)	22	(7.7)	14	(12.7)	205	(15.7)

Data are expressed as numbers of patients (percentages) or as means (standard deviation).

BMI, body mass index; ULN, upper limit of normal; GGT, gamma-glutamyl-transferase.

* Percentages of HCV-infected patients.

Research Article

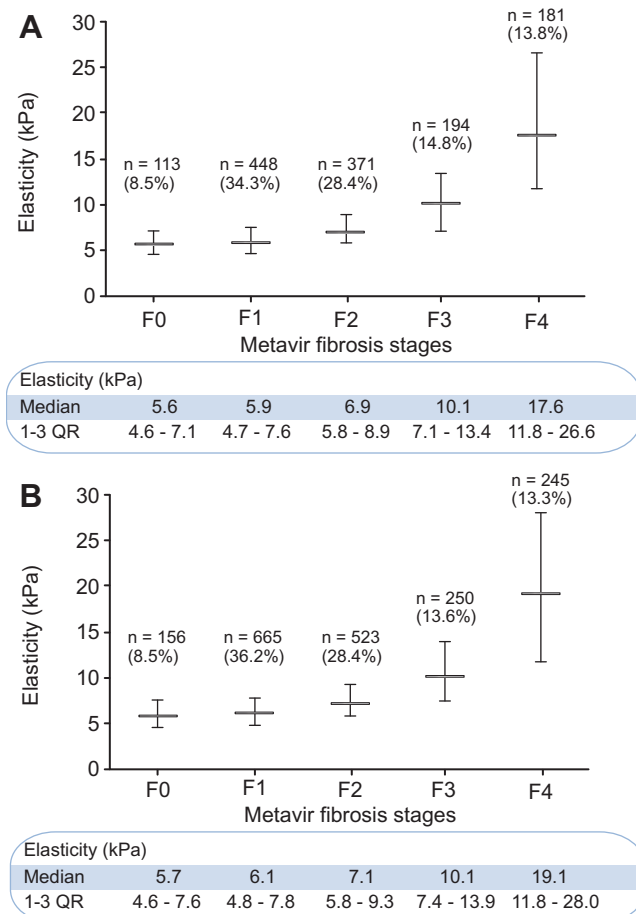


Fig. 2. Distribution of elasticity values according to METAVIR fibrosis stage. (A) In eligible patients ($n = 1307$, complying with technical prerequisites of FibroScan[®]) and (B) in the whole group of patients ($n = 1839$), FIBROSTIC study, France, 2006–2008.

prediction of cirrhosis was 0.90. The desired specificity level of 90% was achieved for a 12.9 kPa threshold, and the sensitivity was 70.2%. AUCs and sensitivity were slightly higher in the HIV subgroup but the difference between subgroups was not statistically significant ($p = 0.04$ for both) (Fig. 3A). The AUC in the prediction of significant fibrosis was 0.76. The desired sensitivity level of 90% was achieved for a 5.2 kPa threshold, and the specificity was 34.0%. AUCs and specificity were similar in the viral infection subgroups (AUC $p = 0.08$, sensitivity $p = 0.45$) (Fig. 3B).

Fourteen centers selected patients for liver biopsy on the basis of prior discordant non-invasive test results (37% of patients) and eight centers used customary clinical criteria (63% of patients). The general characteristics of patients were similar and the overall diagnostic accuracy of TE was the same, but the METAVIR stage distribution differed (F0: 5% vs. 11%, F1: 34% vs. 34%, F2: 28% vs. 28%, F3: 20% vs. 12%, F4: 13% vs. 14%, $p < 0.0001$) (Table 2) and (Appendix 5).

When applying elasticity cut-offs that are previously published, [6] the accuracy of TE in predicting cirrhosis was similar to that obtained in our work (Table 2a). For the diagnosis of significant fibrosis, overall accuracy was the same, sensitivity was lower when using published cut-offs and specificity was higher (Table 2b).

TE accuracy was slightly higher in patients eligible for the primary analysis (AUC 0.90 and 0.76 for cirrhosis and significant fibrosis, respectively) than in those who were excluded (AUC 0.87 and 0.73) (Appendix 5).

Comparative diagnostic accuracy of biomarkers and transient elastography

The comparison of the accuracy of biomarkers and TE is given in Table 2 and Fig. 3C and D. Results for all biomarkers were not available for all patients because of missing biological values and 44 additional Fibrotest[®] scores which were considered non-interpretable by the manufacturer. Therefore, the comparison with FibroScan was based on results of 1197 (Fibrotest), 1204 (Fibrometre), 1272 (APRI), and 1238 patients (Hepascore). The diagnostic accuracy of TE was significantly higher than that of any biomarker except Fibrometre[®] in diagnosing cirrhosis ($p = 0.007$ to $p < 0.0001$). However, there were no differences in the detection of significant fibrosis. The diagnostic accuracy of biomarkers in non eligible patients was similar than in patients included in these paired comparisons with FibroScan[®] (data not shown).

Gain in likelihood of significant fibrosis and cirrhosis

Because better LRs (10.0 and 0.1) were rarely achieved, we used threshold values of 5.0 and 0.2 for defining three test result categories of each test that would indicate a relevant gain in the likelihood of presence (“high” result value) and absence (“low” result value) of the target conditions – or no relevant gain (LR between 0.2 and 5.0, “intermediate” result value).

The post-test probability of significant fibrosis and cirrhosis (See details in Appendix 7) is summarized for each of these three test result categories (Table 3). For example, when the elasticity was >17.1 kPa the probability of cirrhosis was high: 72% of these patients had cirrhosis. The gain in likelihood of cirrhosis was deemed relevant, as the probability of cirrhosis was 13.8% before the test. However, only 10% of patients fell into this class of results. Overall, a high percentage of patients had intermediate values, ranging from 40% to 100% for the diagnosis of cirrhosis and from 70% to 91% for the diagnosis of significant fibrosis (Table 3).

Discussion

The FIBROSTIC study has compared the performance of transient elastography (FibroScan[®]) with that of the main non-invasive biological methods of liver fibrosis assessment in a large representative sample of patients suffering from viral chronic hepatitis and consecutively selected for liver biopsy in routine clinical practice. The performance of FibroScan[®] in predicting cirrhosis was high and higher than that of biomarkers. However, the performance of all the non-invasive methods in predicting significant fibrosis was moderate to poor. There was no difference in the overall accuracy of FibroScan[®]: (i) when studying the whole population recruited, or only the eligible patients who complied with the prerequisites for a valid procedure, (ii) when thresholds were chosen to obtain a targeted specificity (cirrhosis) or sensitivity (significant fibrosis) and when using previously published thresholds [6], (iii) according to the cause of the viral disease, (iv) according to whether patients had, or had not, been selected

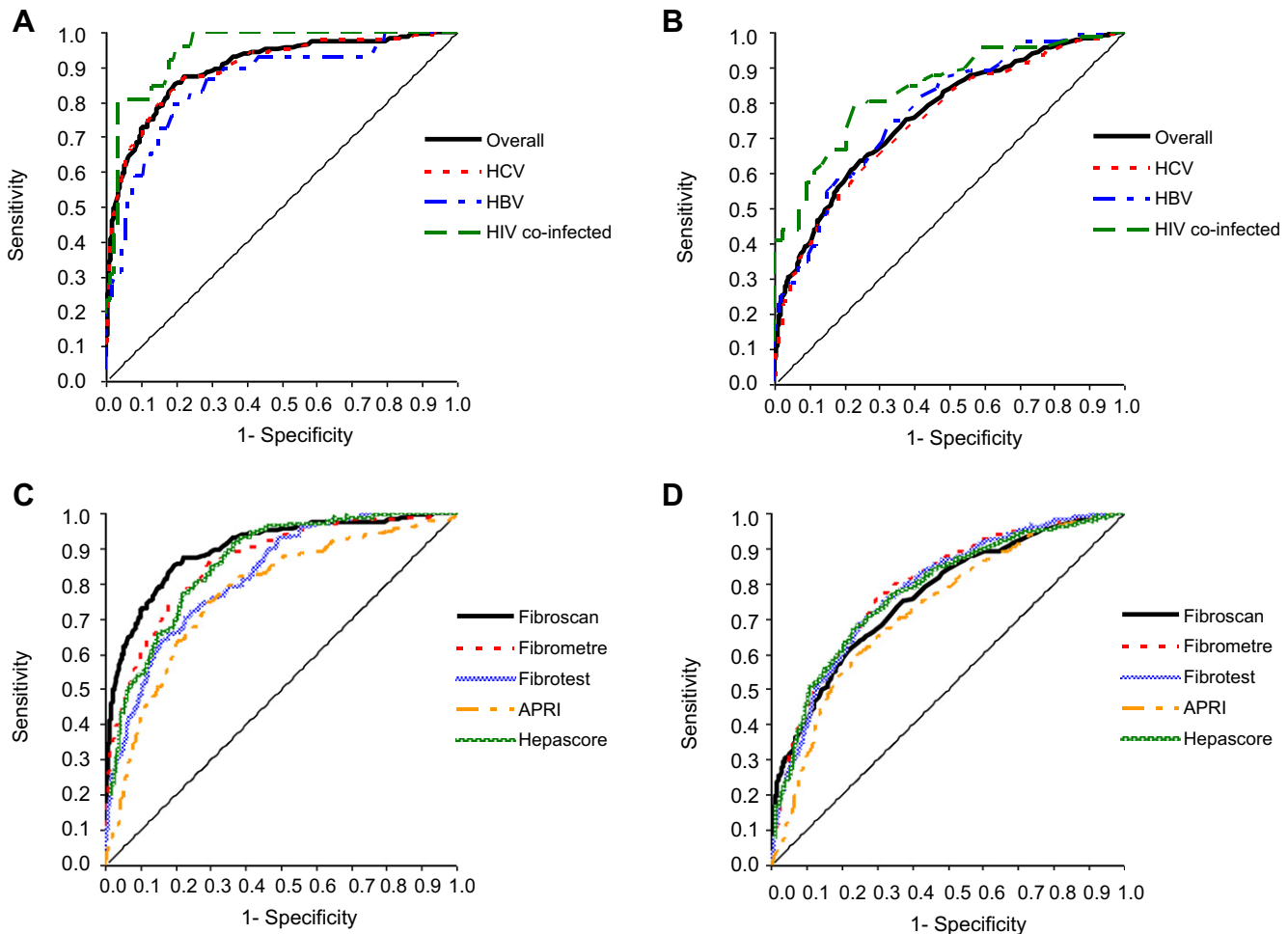


Fig. 3. Receiver-operating-characteristic curves of non-invasive (NI) liver fibrosis tests in the diagnosis of cirrhosis (A and C) and significant fibrosis (B and D) in eligible patients ($n = 1307$). (A and B) Performance of FibroScan[®] according to viral etiology (C and D). Performance of FibroScan[®] compared to that of serum biomarkers. FIBROSTIC study, France, 2006–2008.

for biopsy on the basis of prior abnormal non-invasive test results. Post-test probability estimates indicated that the non-invasive methods are able to rule in or rule out cirrhosis and significant fibrosis in a low proportion of patients.

Liver biopsy has a distinct advantage in that commonly associated liver lesions, such as steatohepatitis and iron overload which can impact on fibrosis progression and treatment response, can be diagnosed and investigated [3]. However, it also has its limitations in staging fibrosis because of the heterogeneous distribution of fibrosis in the liver and the moderate reproducibility of readings [17]. This can constitute a bias when liver biopsy is used as the reference standard in the evaluation of diagnostic tests. This problem is common in diagnostic studies and has attracted much, but not very successful, statistical research [18–20]. The performance can, however, be improved by using an appropriate specimen length and number of portal tracts, as we have done in this study [17]. According to a central review of a random sample of biopsies, our study's METAVIR score reproducibility was acceptable (Appendix 4). Moreover, when biopsies were read by local pathologists or by a single one (at the central laboratory, PB, avoiding the inter-observer variation), the diagnostic accuracy of transient elastography did not differ markedly

(Appendix 5). Finally, as liver biopsy remains the basis of patient management when it is available, it is wise to estimate the accuracy of markers in predicting its results.

The design of our study, which is the first to compare major non-invasive tests for liver fibrosis in a large population [6,21,22], has several strong points and was conducted according to current guidelines [18,23,24]: (i) all patients underwent the reference standard test; (ii) the persons performing transient elastography or interpreting biomarkers were blinded to the results of liver biopsy, and the pathologists were blinded to the results of the non-invasive tests; (iii) to ensure the validity of diagnostic accuracy estimates of non-invasive tests [23,24], the study recruited a large number of consecutive patients in whom liver fibrosis assessment by biopsy was indicated for clinical purposes, and who were thus potential candidates for non-invasive methods in routine practice. All centers applied their usual decision rules for performing a liver biopsy and thus including patients. In some centers, the decision was based on prior results between non-invasive markers [6,21]. This led to a better selection of patients with avoidable biopsies (4% vs. 11% patients with F0) but did not affect the diagnostic accuracy of the tests. Moreover, the comparison between tests could not be biased by the

Research Article

Table 2a. Diagnostic accuracy estimates (95% CI) of FibroScan® and serum biomarkers in the prediction of cirrhosis, FIBROSTIC study, France, 2006–2008.

	Number of patients	Prevalence	Cut-off	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
FibroScan®	1307	13.8	12.9	0.90 (0.87-0.92)	70.2 (62.9-76.7)	90.1 (88.3-91.8)	53.4 (46.8-59.8)	94.9 (93.5-96.2)	7.12 (5.82-8.70)	0.33 (0.26-0.41)
FibroScan®*	1307	13.8	12.5	0.90 (0.87-0.92)	72.4 (65.3-78.7)	89.8 (87.9-91.5)	53.3 (46.8-59.6)	95.3 (93.8-96.5)	7.09 (5.83-8.61)	0.31 (0.24-0.39)
Aetiology										
HCV	913	13.8	12.9	0.90 (0.86-0.93)	72.2 (63.5-79.8)	89.3 (87.0-91.4)	52.0 (44.3-59.6)	95.3 (93.5-96.7)	6.77 (5.38-8.51)	0.31 (0.23-0.41)
HBV	284	10.2	12.9	0.85 (0.78-0.93)	51.7 (32.5-70.6)	92.9 (89.1-95.8)	45.5 (28.1-63.6)	94.4 (90.8-96.9)	7.33 (4.15-12.92)	0.52 (0.36-0.76)
HIV	110	23.6	12.9	0.95 (0.91-0.99)	80.8 (60.6-93.4)	89.3 (80.6-95.0)	70.0 (50.6-85.3)	93.8 (86.0-97.9)	7.54 (3.95-14.37)	0.22 (0.10-0.48)
Liver biopsy										
Prior NI test	479	13.4	12.9	0.86 (0.81-0.92)	62.5 (49.5-74.3)	89.6 (86.3-92.4)	48.2 (37.1-59.4)	93.9 (91.1-96.1)	6.03 (4.29-8.48)	0.42 (0.30-0.57)
No selection	816	14.3	12.9	0.91 (0.88-0.94)	74.4 (65.5-82.0)	90.4 (88.0-92.5)	56.5 (48.3-64.5)	95.5 (93.6-96.9)	7.76 (6.03-9.97)	0.28 (0.21-0.39)
Biomarkers†										
Fibrotest®	1197	14.3	0.74	0.82 (0.79-0.85)†	62.6 (56.1-68.8)	84.4 (82.4-86.2)	40.1 (35.1-45.3)	93.1 (91.6-94.4)	4.01 (3.34-4.82)	0.44 (0.36-0.54)
Fibrometre®	1204	13.6	0.442	0.86 (0.83-0.89)	43.9 (37.4-50.6)	95.0 (93.7-96.1)	58.1 (50.3-65.6)	91.5 (90.0-92.8)	8.78 (6.40-12.05)	0.59 (0.52-0.68)
APRI	1272	13.9	2.0	0.77 (0.73-0.81)†	45.2 (38.9-51.7)	88.4 (86.7-90.0)	38.7 (33.0-44.6)	90.9 (89.3-92.3)	3.90 (3.10-4.91)	0.62 (0.54-0.71)
Hepascore	1238	14.0	0.84	0.86 (0.83-0.88)‡	59.0 (52.4-65.3)	87.4 (85.6-89.1)	43.2 (37.8-48.8)	92.9 (91.4-94.2)	4.69 (3.83-5.73)	0.47 (0.39-0.56)

AUC, Area Under the Receiver-Operator-Characteristic curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value; LR+, Positive Diagnostic Likelihood Ratio; LR-, Negative Diagnostic Likelihood Ratio; Prior NI test, Patients' indication for liver biopsy (and therefore inclusion in the study) based on prior non-invasive tests results.

* Threshold reported by Castera et al. [6].

† Missing data for biomarkers: Fibrotest®: 66 (+44 not interpretable), Fibrometre®: 103, APRI: 35, Hepascore: 69 patients. The overall FibroScan® AUC was compared to each biochemical score (statistically significant if $p < 0.01$). The table reports statistical comparisons for AUCs only: † p -value < 0.0001 . ‡ p -value = 0.0075.

distribution of fibrosis stages as it was done in the same population.

The representativity of the included population is an important feature, as the study population characteristics may influence diagnostic accuracy measures [25–27]. The fact that, in a prospective wide evaluation of the technique, 29% of patients were not eligible, needs particular attention. However, most of these exclusions were due to non compliance with technical specifications of transient elastography [6,28], making the validity of measures uncertain. Complying with these specifications is needed for a correct interpretation of elasticity results by physicians. Therefore, the population that we analyzed will be representative of the real life population in whom the FibroScan® technique will be considered interpretable.

Physicians or technicians performing the technique were previously trained (more than 50 procedures), and the technical and clinical prerequisites were mentioned in the protocol. Pitfalls of liver stiffness measurement were recently reported from a single center with a large experience of over 13,000 examinations, showing that unreliable measures due to non compliance with technical specifications of FibroScan® (fewer than 10 valid shots, success rate < 60 , or IQR/median stiffness $< 30\%$) were reported in 15.8% of the procedures [28]. These figures are comparable to ours, where the transient elastography was performed in 23 different centers, by less trained operators. Our results, as those of Castera, should be used very carefully when recommending a large use of this technique, since even in a very specialized center, pitfalls were frequently observed. Moreover, one must also keep in mind the cost of a non valid technique and the possible

consequences of misinterpretation of the histological liver status. FibroScan® users should comply with these conditions of use in order to exclude undesirable variability and guarantee the excellent reproducibility reported for the technique when used in those conditions [16].

Our underlying objective was to compare the accuracy of the non-invasive tests in informing clinicians about the probability of disease in individual patients. As thresholds for FibroScan® were variable in previous studies, elasticity thresholds were based on study data in the present work and chosen in order (i) to achieve high sensitivity when looking for significant fibrosis, so as not to miss patients needing treatment, and (ii) to achieve high specificity when looking for cirrhosis, so as not to submit patients to unnecessary detection of complications. Despite these choices, there were 30% false-negatives for significant fibrosis and 47% false-positives for cirrhosis. This is partly explained by the prevalence of these conditions in patients undergoing fibrosis assessment. Differences with other reported accuracy estimates may be due to the use of different thresholds, different patient spectrums, or the small number of patients in earlier studies [4,5,29–32]. Variation of accuracy related to the fed state is unlikely to have occurred, as 87% of patients were fasting or had a light meal only [33]. We obtained the same estimates for AUC as the largest previously reported study which enrolled 955 patients with chronic hepatitis C [29]. The overall accuracy of FibroScan® was as good as or better than that of other non-invasive methods. The choice of different elasticity thresholds for FibroScan® did not bias this comparison, as it was based on ROC curves AUC, which take into account every possible

Table 2b. Diagnostic accuracy estimates (95% CI) of FibroScan® and serum biomarkers in the prediction of significant fibrosis, FIBROSTIC study, France, 2006–2008.

	Number of patients	Prevalence	Cut-off	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
FibroScan®	1307	57.1	5.2	0.76 (0.74–0.79)	90.1 (87.7–92.1)	34.0 (30.1–38.1)	64.5 (61.5–67.4)	72.1 (66.3–77.4)	1.37 (1.28–1.46)	0.29 (0.23–0.37)
FibroScan®	1307	57.1	7.1	0.76 (0.74–0.79)	66.5 (63.0–69.9)	70.8 (66.8–74.5)	75.2 (71.7–78.4)	61.4 (57.5–65.1)	2.27 (1.98–2.61)	0.47 (0.42–0.53)
Aetiology										
HCV	913	61.6	5.2	0.75 (0.71–0.78)	89.7 (86.9–92.1)	32.2 (27.3–37.4)	67.9 (64.4–71.3)	66.1 (58.5–73.1)	1.32 (1.22–1.43)	0.32 (0.24–0.43)
HBV	284	41.6	5.2	0.78 (0.72–0.83)	89.0 (81.9–94.0)	38.0 (30.5–45.8)	50.5 (43.5–57.5)	82.9 (72.5–90.6)	1.43 (1.25–1.64)	0.29 (0.17–0.50)
HIV	110	60.0	5.2	0.84 (0.76–0.91)	95.5 (87.3–99.1)	34.1 (20.5–49.9)	68.5 (58.0–77.8)	83.3 (58.6–96.4)	1.45 (1.16–1.80)	0.13 (0.04–0.43)
Liver biopsy										
Prior NI test	479	61.8	5.2	0.76 (0.72–0.80)	88.9 (84.7–92.2)	38.8 (31.7–46.3)	70.1 (65.2–74.7)	68.3 (58.4–77.1)	1.45 (1.28–1.64)	0.29 (0.20–0.42)
No selection	816	54.5	5.2	0.77 (0.73–0.80)	90.8 (87.7–93.3)	31.8 (27.1–36.8)	61.5 (57.6–65.2)	74.2 (66.7–80.8)	1.33 (1.23–1.44)	0.29 (0.21–0.40)
Biomarkers ^{††}										
Fibrotest®	1197	57.5	0.48	0.78 (0.75–0.81)	67.4 (64.4–70.4)	75.3 (71.9–78.4)	78.7 (75.7–81.4)	63.1 (59.8–66.4)	2.72 (2.32–3.20)	0.43 (0.38–0.49)
Fibrometre®	1204	57.1	0.411	0.79 (0.76–0.81)	83.1 (80.6–85.4)	57.1 (53.4–60.7)	72.0 (69.3–74.6)	71.8 (67.9–75.4)	1.94 (1.74–2.15)	0.30 (0.25–0.35)
APRI	1272	57.1	0.5	0.72 (0.69–0.75)	79.9 (77.3–82.3)	48.4 (44.8–52.0)	67.3 (64.6–69.9)	64.4 (60.3–68.3)	1.55 (1.42–1.69)	0.42 (0.35–0.49)
Hepascore	1238	57.1	0.5	0.78 (0.75–0.80)	52.9 (49.7–56.0)	86.3 (83.6–88.7)	83.7 (80.5–86.5)	57.9 (54.9–60.8)	3.85 (3.08–4.81)	0.55 (0.50–0.59)

AUC, Area Under the Receiver–Operator–Characteristic curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value; LR+, Positive Diagnostic Likelihood Ratio; LR–, Negative Diagnostic Likelihood Ratio; Prior NI test, Patients' indication for liver biopsy (and therefore inclusion in the study) based on prior non-invasive tests results.

threshold. The comparisons of diagnostic accuracy between FibroScan® and biomarkers were based on results available and interpretable at the same time for both non-invasive techniques. Paired comparisons are more valid than comparisons between different groups of subjects and we checked that the diagnostic accuracy of biomarkers was not different when assessed in the non eligible group. However, biomarkers will provide results in more patients than with FibroScan®, as they usually do not generate non- interpretable results (with the exception of Fibrotest®: 3.5% of the eligible group) if the results of biological assays are available.

Multilevel diagnostic LRs inform on different classes of measurements and not just on positive and negative results according to a single cut-off level [14]. Their calculation allowed us to define three classes of test results providing a rough guide of the values that determine whether the result rules in or rules out significant fibrosis and cirrhosis. These probabilities enable clinicians to directly interpret the results of patients tested in similar settings. However, non-invasive tests are strongly predictive of the target conditions in a limited proportion of patients only (cirrhosis: 0–60% significant fibrosis: 9–30% of patients, depending on the test). Results will thus not prove useful for ruling in or ruling out the target conditions in most patients.

According to current practice guidelines, significant fibrosis ($\geq F2$) is a frequent selection criterion for antiviral treatment of HCV and HBV chronic hepatitis, but the use of non-invasive tests to stage liver fibrosis remains highly controversial. In summary,

our study has established that non-invasive tests, especially FibroScan®, may be useful in the prediction of cirrhosis. However, it supports guideline conclusions that non-invasive tests should not replace liver biopsy in routine clinical practice for the detection of significant fibrosis that may warrant treatment [2,3].

Funding

French health authorities (National STIC grant).

Conflict of interests

The authors do not have a relationship with the manufacturers of the drugs involved either in the past or present and did not receive funding from the manufacturers to carry out their research.

Acknowledgements

We are indebted to Florence Tubach, Yolande Costa and Virginia Olympio (Unité de Recherche Clinique Paris Nord) for their outstanding contribution to the setting up and monitoring of this study. Calculations of Fibrometre and Fibrotest are a courtesy of Bioliverscale and Biopredictive and we thank them.

We thank Professor Bruno Falissard, President of the Scientific Committee of the study for helpful discussion. We also thank Tiiu Ojasoo for help in editing the manuscript.

Research Article

Table 3. Probability of significant fibrosis and cirrhosis (%) according to three classes of non-invasive test results, 1307 patients included in the main analysis, FIBROSTIC study, France 2006–2008.

	Class of test result value*		
	Low (LR ≤ 0.2)	Intermediate (LR 0.21 to 4.9)	High (LR ≥ 5.0)
Result for diagnosis of cirrhosis (pre-test probability 13.8%)[†]			
FibroScan [®]	≤ 7.0	7.1 to 17.1	> 17.1
Probability of cirrhosis (%)	0 to 3	7 to 27	72
Number of patients (%)	647 (50)	527 (40)	133 (10)
Fibrometre [®]	≤ 0.018	0.019 to 0.450	> 0.450
Probability of cirrhosis (%)	1 to 2	3 to 29	60
Number of patients (%)	353 (29)	730 (61)	121 (10)
Fibrotest [®]	≤ 0.29	0.30 to 0.86	> 0.86
Probability of cirrhosis (%)	0 to 2	7 to 31	50
Number of patients (%)	351 (29)	718 (60)	128 (11)
APRI	-	≤ 0.29 to 2.74	-
Probability of cirrhosis (%)	-	3 to 40	-
Number of patients (%)	0 (0)	1272 (100)	0 (0)
Hepascore	≤ 0.30	0.31 to 0.97	> 0.97
Probability of cirrhosis (%)	0 to 2	12 to 26	58
Number of patients (%)	606 (49)	503 (41)	129 (10)
Result for diagnosis of significant fibrosis (pre-test probability 57.1%)[†]			
FibroScan [®]	≤ 4.3	4.4 to 17.1	> 17.1
Probability of significant fibrosis (%)	20	30 to 82	95
Number of patients (%)	117 (9)	1057 (81)	133 (10)
Fibrometre [®]	≤ 0.196	0.197 to 0.878	> 0.878
Probability of significant fibrosis (%)	14	24 to 78	88 to 89
Number of patients (%)	119 (10)	844 (70)	241 (20)
Fibrotest [®]	≤ 0.12	0.13 to 0.86	> 0.86
Probability of significant fibrosis (%)	11	31 to 83	91
Number of patients (%)	106 (9)	963 (80)	128 (11)
APRI	≤ 0.29	0.30 to > 2.74	-
Probability of significant fibrosis (%)	21	36 to 84	-
Number of patients (%)	120 (9)	1152 (91)	0 (0)
Hepascore	-	≤ 0.07 to 0.97	> 0.97
Probability of significant fibrosis (%)	-	22 to 84	91
Number of patients (%)	0 (0)	1109 (90)	129 (10)

Classes of test results were defined according to multilevel diagnostic likelihood ratios (LR) values: low when LR ≤ 0.2 , intermediate when LR from 0.21 to 4.9, high when LR ≥ 5.0 (see text for details). [†] The pre-test probability (prevalence) of the target condition must be taken into account to interpret the post-test probability according to test results.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jhep.2010.05.035](https://doi.org/10.1016/j.jhep.2010.05.035).

References

- [1] Bedossa P, Poynard T for the METAVIR cooperative study group. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289–293.
- [2] EASL clinical practice guidelines: management of chronic hepatitis B. *J Hepatol* 2009;50:227–242.
- [3] Ghani MG, Strader DB, Thomas DL, Seef LB. Diagnosis, management and treatment of hepatitis C: an update. *Hepatology* 2009;49:1335–1374.
- [4] Friedrich-Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, Zeuzem S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. *Gastroenterology* 2008;134:960–974.
- [5] Pambrun E, Bouteloup V, de Lédínghen V, Asselineau J, Fraquelli M, Brunetto M, et al. Individual patient data meta-analysis of transient elastography diagnostic accuracy in liver fibrosis assessment of chronic hepatitis C patients (TE IPD Study). *J Hepatol* 2009;50:S150.
- [6] 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.
- [7] Jaeschke R, Guyatt GH, Sackett DL. Users' Guides to the Medical Literature: III. How to use an article about a diagnostic Test: B. What are the results and will they help me in caring for my patients? *JAMA* 1994;271:703–707.
- [8] Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. *Hepatology* 2009;49:1017–1044.

- [9] Cadranel JF, Rufat P, Degos F for the Epidemiology group of the French Association for Study of the Liver (AFEF). Practices of liver biopsy in France. Results of a prospective nationwide survey. *Hepatology* 2000;32:477–481.
- [10] Boursier J, Bacq Y, Halfon P, Leroy V, de Ledinghen V, de Muret A, et al. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009;21:28–38.
- [11] Imbert-Bismut F, Radzui V, Pieroni L, Charlotte F, Benhamou Y, Poinard T, et al. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001;357:1069–1075.
- [12] Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple non invasive index can predict both relevant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003;38:518–526.
- [13] 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.
- [14] Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;365:1500–1505.
- [15] DeLong E, DeLong D, Clarke-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
- [16] Fraquelli M, Rigamonti C, Casazza G, Conte D, Donato MF, Ronchi G. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver diseases. *Gut* 2007;56:968–973.
- [17] Bedossa P, Dargère D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;38:1449–1457.
- [18] Knottnerus JA, editor. The evidence base of clinical diagnosis. London: BMJ Books; 2002.
- [19] Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11.
- [20] Mehta SH, Lau B, Afdhal NH, Thomas DL. Exceeding the limits of liver histology markers. *J Hepatol* 2009;50:36–41.
- [21] Sebastiani G, Halfon P, Castera L, Pol S, Thomas DL, Mangia A, et al. SAFE biopsy: a validated method for large scale staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1821–1827.
- [22] Moreno S, Garcia-Samaniego J, Moreno A, Ortega E, Pineda JA, del Romero J, et al. Non invasive diagnosis of liver fibrosis in patients with HIV infection and HCV/HBV coinfection. *J Viral Hepat* 2009;16:249–258.
- [23] Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- [24] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med* 2003;138:40–44.
- [25] Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of non invasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008;54:1372–1378.
- [26] Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.
- [27] Arena U, Vizzuti F, Abraldes JG, Corti G, Stasi C, Moscarella S, et al. Reliability of transient elastography for the diagnosis of advanced fibrosis in chronic hepatitis C. *Gut* 2008;57:1288–1293.
- [28] Castéra L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010;51:828–835.
- [29] Coco B, Oliveri F, Maiana AM, Ciccorossi P, Sacco R, Colombatto P, et al. Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat* 2007;14:360–369.
- [30] Kettaneh A, Marcellin P, Douvin C, Poupon R, Zioli M, Beaugrand M, et al. Features associated with success performance of Fibroscan measurements for the diagnosis of cirrhosis in HCV patients. A Prospective study of 935 patients. *J Hepatol* 2007;46:628–634.
- [31] Lupsor M, Badea R, Stefanescu H, Grigorescu M, Sparchez Z, Serban A, et al. Analysis of histopathological changes that influence liver stiffness in chronic hepatitis C. Results from a cohort of 324 patients. *J Gastrointest Liver Dis* 2008;17:155–163.
- [32] Lucidarme D, Foucher J, Le Bail B, Vergniol J, Castera L, Duburque C, et al. Factors of accuracy of transient elastography (Fibroscan) for the diagnosis of liver fibrosis in chronic hepatitis C. *Hepatology* 2009;49:1083–1089.
- [33] Mederacke I, Wurstthom K, Kirschner J, Rifai K, Manns MP, Wedemeyer H, et al. Food intake increases liver stiffness in patients with chronic or resolved hepatitis infection. *Liver Int* 2009;10:1500–1506.