Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY



# Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study

Jean-Pierre Zarski[1,2,*], Nathalie Sturm[2,3], Jérôme Guechot[4], Adeline Paris[5], Elie-Serge Zafrani[6], Tarik Asselah[7], Renée-Claude Boisson[8], Jean-Luc Bosson[5], Dominique Guyader[9], Jean-Charles Renversez[10], Jean-Pierre Bronowicki[11], Marie-Christine Gelineau[12], Albert Tran[13], Candice Trocme[14], Victor De Ledinghen[15], Elisabeth Lasnier[4], Armelle Poujol-Robert[16], Frédéric Ziegler[17], Marc Bourliere[18], Hélène Voitot[19], Dominique Larrey[20], Maria Alessandra Rosenthal-Allieri[21], Isabelle Fouchard Hubert[22], François Bailly[23], Michel Vaubourdolle[4], The ANRS HCEP 23 Fibrostar Group

[1]*Clinique Universitaire d'Hépato-Gastroentérologie – Pôle DIGIDUNE-CHU de Grenoble, France;* [2]*Unité INSERM/UJF U823 IAPC Institut Albert Bonniot-Grenoble, France;* [3]*Département d'Anatomie et Cytologie Pathologiques-Pôle de Biologie, CHU de Grenoble, France;* [4]*Laboratoire de Biochimie A – Hôpital Saint Antoine, AP-HP, Paris, France;* [5]*Centre d'Investigation Clinique INSERM CIC003, CHU Grenoble, Grenoble F-38043, France;* [6]*Département de Pathologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Université Paris Est, Créteil, Val de Marne, France;* [7]*Service d'Hépato-Gastroentérologie, Hôpital Beaujon, AP-HP, Clichy, France;* [8]*Laboratoire de Biochimie, Centre hospitalier Lyon Sud, Hospices Civils de Lyon, France;* [9]*Service d'Hépatologie CHU de Rennes, France;* [10]*Département de Biologie Intégrée, Pôle de Biologie, CHU de Grenoble, France;* [11]*Service d'Hépato-Gastroentérologie, CHU de Nancy, France;* [12]*Laboratoire de Biochimie, Hôtel Dieu, Hospices Civils de Lyon, France;* [13]*Hépatologie, Hôpital l'Archet 2, Nice, France;* [14]*GREPI TIMC-IMAG UMR CNRS 5525, Université de Grenoble Laboratoire d'Enzymologie, Dépt. de Biologie et Pathologie de la Cellule, Pôle de Biologie, CHU de Grenoble, France;* [15]*Service d'Hépato-Gastroentérologie, CHU de Bordeaux, France;* [16]*Service d'Hépato-Gastroentérologie, Hôpital St Antoine, AP-HP, Paris, France;* [17]*Laboratoire de Biochimie, Institut de Biologie Clinique, EA4311, IHU, IFR23/CHU, Rouen, France;* [18]*Service d'Hépato-Gastroentérologie, Hôpital St Joseph, Marseille, France;* [19]*Laboratoire de Biochimie, Hôpital Beaujon, AP-HP, Clichy, France;* [20]*Service d'Hépato-Gastroentérologie, CHU de Montpellier, France;* [21]*Laboratoire d'Immunologie, CHU de Nice, France;* [22]*Service d'Hépato-Gastroentérologie, CHU d'Angers, France;* [23]*Service d'Hépato-Gastroentérologie, Hotel Dieu-Lyon, France*

**Background & Aims**: Blood tests and transient elastography (Fibroscan™) have been developed as alternatives to liver biopsy. This ANRS HCEP-23 study compared the diagnostic accuracy of nine blood tests and transient elastography (Fibroscan™) to assess liver fibrosis, *vs.* liver biopsy, in untreated patients with chronic hepatitis C (CHC).

**Methods**: This was a multicentre prospective independent study in 19 French University hospitals of consecutive adult patients having simultaneous liver biopsy, biochemical blood tests (performed in a centralized laboratory) and Fibroscan™. Two experienced pathologists independently reviewed the liver biopsies (mean length = 25 ± 8.4 mm). Performance was assessed using ROC curves corrected by Obuchowski's method.

**Results**: Fibroscan™ was not interpretable in 113 (22%) patients. In the 382 patients having both blood tests and interpretable Fibroscan™, Fibroscan™ performed similarly to the best blood tests for the diagnosis of significant fibrosis and cirrhosis. Obuchowski's measure showed Fibrometer® (0.86), Fibrotest® (0.84), Hepascore® (0.84), and interpretable Fibroscan™ (0.84) to be the most accurate tests. The combination of Fibrotest®, Fibrometer®, or Hepascore® with Fibroscan™ or Apri increases the percentage of well classified patients from 70–73% to 80–83% for significant fibrosis, but for cirrhosis a combination offers no improvement. For the 436 patients having all the blood tests, AUROC's ranged from 0.82 (Fibrometer®) to 0.75 (Hyaluronate) for significant fibrosis, and from 0.89 (Fibrometer®) and Hepascore® to 0.83 (FIB-4) for cirrhosis.

**Conclusions**: Contrarily to blood tests, performance of Fibroscan™ was reduced due to the uninterpretable results. Fibrotest®, interpretable Fibroscan™, Fibrometer®, and Hepascore® perform best and similarly for diagnosis of significant fibrosis and cirrhosis.



Keywords: Chronic hepatitis C; Liver fibrosis; Surrogate markers; Transient elastography; Blood tests.
*Received 21 October 2010; received in revised form 13 April 2011; accepted 3 May 2011*

* Corresponding author. Address: Clinique Universitaire d'Hépato-Gastroentérologie – Pôle DIGIDUNE-CHU de Grenoble, BP 217, 38043 Grenoble Cedex 9, France. Tel.: +33 476 76 54 41; fax: +33 476 76 51 79.
*E-mail address:* JPZarski@chu-grenoble.fr (J.-P. Zarski).

*Abbreviations:* CHC, chronic hepatitis C; ROC, receiver operating characteristic curves; AUROC, area under receiver operating curve; HCV, Hepatitis C virus; LSM, liver stiffness measurement; LB, liver biopsy; BMI, body mass index; NPV, negative predictive value; PPV, positive predictive value; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma glutamyltranspeptidase.





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

# Research Article

## Introduction

Liver biopsy is the method of reference to assess the fibrosis stage in chronic hepatitis C (CHC). However, it is an invasive procedure with severe complications in about 0.5% of cases [1] and its accuracy is limited by sampling heterogeneity [2] and inter-observer and intra-observer variation [3,4]. Biopsy specimens less than 15 mm in length appear poorly reliable [3]. Semi-quantitative evaluation of fibrosis has high variability especially among non-expert pathologists [4,5]. Several blood tests with or without scores calculated from statistical models have been developed to evaluate fibrosis. Hyaluronate was proposed as a non-invasive marker [6]. Fibrotest® was the first score combining several variables proposed for patients with CHC [7]. Apri [8], Fibrometer® [9], and Hepascore® [10] were then validated in these patients. Other fibrosis scores have been recently proposed but are not often performed in practice, FIB-4 [11], Forns's score [12], MP3 [13,14], and the European Liver Fibrosis Group or ELFG score [15]. However, all these tests have limitations. Blood test results can be influenced by other associated diseases, comorbidities or different dosage techniques.

Another alternative, transient elastography (Fibroscan™; Echosens, Paris, France) is based on liver stiffness measurement. Its diagnostic performance is similar to that of serological markers [16–20]. However Fibroscan™ has some limitations (failure and unreliability) particularly in obese patients or in circumstances of limited operator experience, as recently discussed by Castera *et al.* [21].

The aim of this study was to perform a prospective independent multicenter comparative evaluation of most of the currently best evaluated non-invasive markers i.e. blood tests and transient elastography, *vs.* liver biopsy in an etiologically homogenous study group (CHC), with an appropriate number of patients, appropriate histological analysis and using well standardized biological tests.

## Patients and methods

### *Patients*

Consecutive adult patients with chronic hepatitis C were prospectively considered for inclusion if they were naïve of treatment or had no treatment during the last 6 months, interpretable liver biopsy with delay between biopsy and blood tests of <3 months. All patients had been referred for tests in order to make a decision on treatment strategy. CHC was confirmed by HCV–RNA polymerase chain reaction analysis of serum. Cirrhotic patients were compensated and asymptomatic at the time of inclusion. Patients with co-existing liver diseases attributed to alcohol, hepatitis B, auto-immune hepatitis, primary biliary cirrhosis, hemochromatosis, alpha-1-antitrypsine deficiency, or Wilson's disease were excluded by history and clinical, laboratory, imaging, and histological data. Human immunodeficiency virus co-infected and post-transplant patients were also excluded. The protocol was approved by the ethics committee "CPP Sud-Est 5". All patients gave written informed consent. Liver biopsies were performed as part of normal clinical care for staging and grading of liver disease before antiviral treatment. Demographic data were recorded at the time of the liver biopsy.

### *Biological scores of liver fibrosis*

Fasting blood samples were collected by venipuncture. The same batches of tubes were used for all patients (BD Vacutainer®, type 9NC, K2E and Z, Becton–Dickinson, Plymouth, UK).

Cholesterol, platelet count, and prothrombin time were immediately measured in each center. All other biological parameters were measured in a centralized laboratory using serum samples immediately fractioned into 0.5 ml fractions in 1.5 ml screw cap micro tubes (Sarstedt, Nümbrecht, Germany), then frozen and stored at −80 °C until assayed. Samples were transported in dry-ice by a specialized transporter (AreaTime Logistics, Cergy Pontoise, France). All the tests were performed blind of clinical and histological data.

The following blood tests were evaluated: Fibrotest®, Fibrometer®, Forns score, Apri, MP3, ELFG, Hepascore®, FIB-4, Hyaluronate. Blood test scores were calculated according to the most recent published formulae [8,10–15], or patent for Fibrotest® [7] and Hepascore® [10], or by the courtesy of the manufacturer (BioLivescale) for Fibrometer® [9]. The list of variables included in each test and the measurement techniques are detailed in the Supplementary data.

### *Liver stiffness measurement by transient elastography (Fibroscan™)*

Measurements were made on the right lobe of the liver, through the intercostal spaces with the patient lying in dorsal decubitus with the right arm in maximal abduction by the operator who performed the liver biopsy. The tip of the transducer probe was covered with coupling gel and placed on the skin, between two ribs at the level of the right lobe. Liver stiffness measurement (Fibroscan™) failure was defined as zero valid shots (after at least 10 attempts) and "unreliable examinations" were defined as fewer than 10 valid shots or an interquartile range (IQR)/LSM greater than 30% or a success rate less than 60% [16–19].

### *Liver biopsy*

Liver biopsies (LB) were performed using Menghini's technique with a 1.6 mm needle (Hepafix, Brown, Melsungen, Germany), formalin-fixed in the centers and paraffin embedded. Sections (4 mm) were stained with hematoxylin-eosin-saffron, and picrosirius red. The liver fibrosis stage was evaluated according to the METAVIR scoring system [5], independently by two senior liver pathologists (NS, ESZ) blind to clinical and biological data. In cases of disagreement, slides were simultaneously reviewed using a multi-pipe microscope to reach a consensus. Inter-observer agreement was evaluated using the kappa index, called κ, which excludes chance-expected agreement and the weighted κ index according to a linear evolution of the METAVIR score [4]. The length of biopsy and the number of portal tracts were recorded. To be considered for scoring, LB less than 20 mm had to measure at least 15 mm and/or contain at least 11 portal tracts, except for cirrhosis.

### *Statistical analysis*

Due to the inherent difficulty in the interpretability of Fibroscan™ we defined two populations, the first including patients with all the available blood tests (436 patients), and the second population including patients having both interpretable Fibroscan™ (excluding cases in which Fibroscan™ was not possible, failures and unreliable tests) and all blood tests (382 patients).

Descriptive results were expressed as the mean ± standard deviation or as the number (percentage) of patients. The diagnostic performance of the non-invasive methods was assessed using AUROCs, considering liver biopsy as a "gold standard", albeit imperfect, and its 95% confidence intervals. We used cut-offs corresponding to the score associated with $p < 0.05$ in the corresponding logistic regression model. Comparison of AUROCS was performed using a $Chi^2$ test associated with the procedure of "ROCGOLD" (Stata™). Due to the multiple comparisons between scores, the method of Sidak was used to exclude the risk of concluding wrongly, with an alpha risk of $p_{(Sidak)} \leqslant 0.05$ for statistical significance.

Since AUROC assumes a binary gold standard while histological fibrosis staging is based on an ordinal scale we used another estimator of diagnostic test accuracy which does not require dichotomization of the gold standard. The Obuchowski measure [22], was recently recommended as a multinomial version of the AUC. With N (= 5) categories of the gold standard outcome and AUCst, it estimates the AUC of diagnostic tests differentiating between categories s and t. The Obuchowski measure is a weighted average of the $N(N-1)/2$ (= 10) different AUCst corresponding to all the pair-wise comparisons between two of the N categories. All these paired comparisons are also weighted using a penalty function proportional to the difference in METAVIR units. In our study the penalty function was 1 for each different METAVIR unit. As proposed by Lambert *et al.* [23] we thus defined a penalty function proportional to the difference in METAVIR units between stages (the penalty function was 0.25 when the difference was 1, 0.5 when the difference was 2, 0.75 when the difference was 3, and 1 when the difference was 4).

We combined the main tests pair-wise, calculating the % of concordant well classified patients given by the tests and the number of avoided biopsies (assuming biopsy to be the gold standard).





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

All statistical tests were 2-tailed, with a type I error of 5%. Statistical analysis was performed at the Grenoble Clinical Research Centre using STATATM Mac OS X.

## Results

### *Patient characteristics*

Between November 2006 and July 2008, 590 patients with chronic hepatitis C and liver biopsy were enrolled in 19 French academic centres. METAVIR fibrosis stages in our population were F0: 6.6%, F1: 47.5%, F2: 15.6%, F3: 16.3%, and F4: 14.0%. Fig. 1 gives the reasons for 78 patients being excluded from all analyses. Several patients were excluded from blood test analyses due to the missing data. Fibroscan™ was not interpretable in 113 (22%) patients: 56 failures (11%) and 57 (11%) unreliable. Some statistically significant differences were observed between patients with or with failed Fibroscan™ (see Supplementary results).

We analysed separately the 436 patients who had all the available blood tests and the 382 patients who had both all blood tests and an interpretable Fibroscan™. No difference was observed between the two groups regarding the main demographic, laboratory, and histological features (Table 1). Indeed no significant difference was observed between the 512 non-excluded patients and the 436 patients having all blood tests (Supplementary Table S5).

The median delay between biopsy and test measurements was 5 days (0–65). Only 13 patients (2.5%) had a length of biopsy

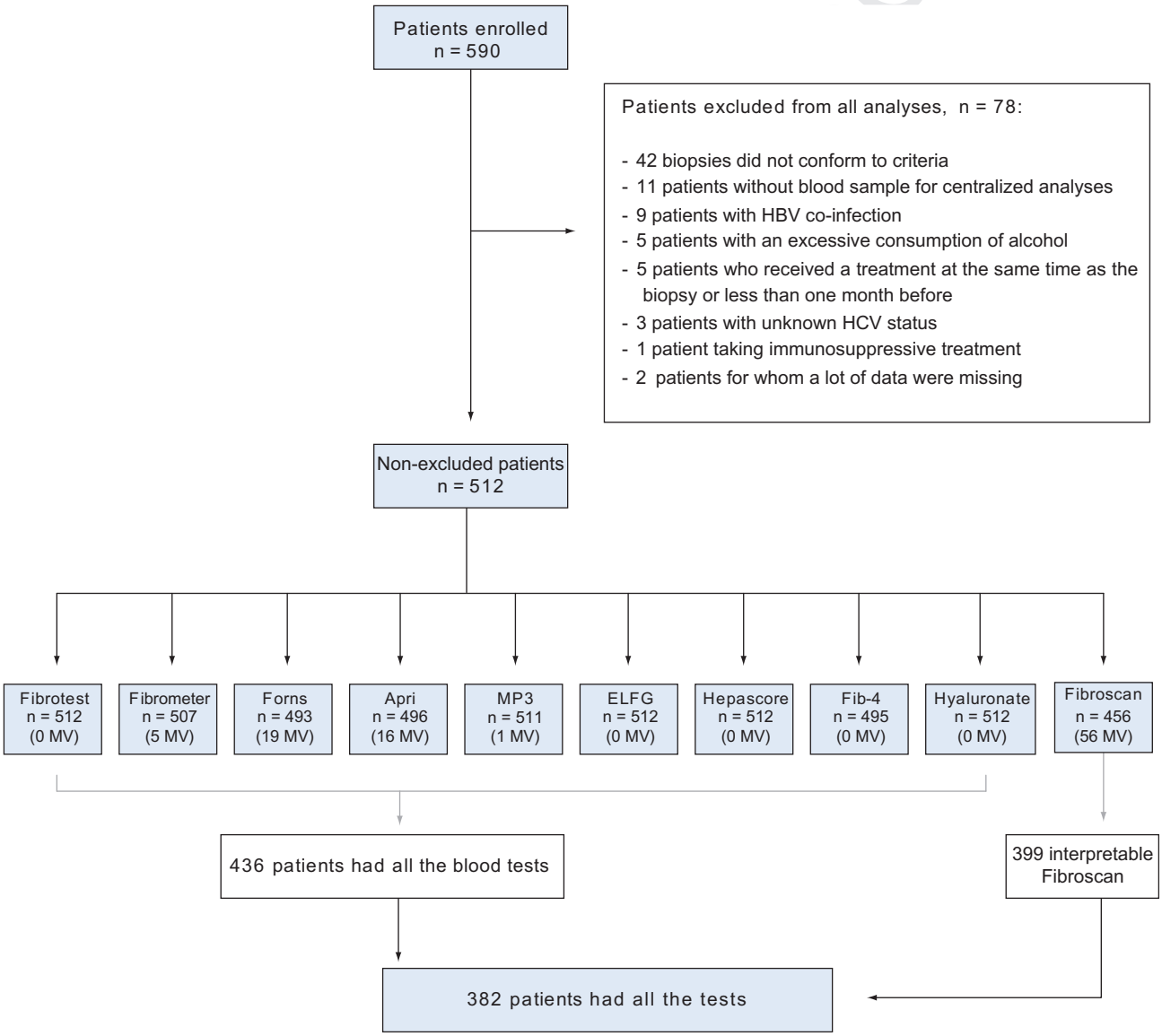


**Fig. 1.** Flow chart. MV: missing value.



Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

# Research Article

**Table 1. Demographic, laboratory, and histological features for the 436 CHC patients having all the blood tests and the 382 CHC patients with all the blood tests and interpretable Fibroscan™.**

| Characteristics | n = 436 | n = 382 |
|---|---|---|
| Age (years) | 51.2 ± 10.9* | 50.9 ± 10.6* |
| Gender (N,%) | | |
| Males | 268 (61.5%) | 232 (60.7%) |
| Females | 168 (38.5%) | 150 (39.3%) |
| BMI (kg/m$^2$) | 24.5 ± 3.5 | 24.3 ± 3.4 |
| TP (%) | 94.4 ± 7.8 | 94.6 ± 7.9 |
| Cholesterol (mmol/L) | 4.7 ± 1.1 | 4.7 ± 1.0 |
| Bilirubin (μmol/L) | 12.5 ± 6.8 | 12.4 ± 6.8 |
| AST (IU/L) | 62.5 ± 42.1 | 62.9 ± 43.2 |
| ALT (IU/L) | 88.0 ± 64.9 | 87.9 ± 65.4 |
| GGT (IU/L) | 93.4 ± 96.8 | 96.6 ± 99.8 |
| Urea (mmol/L) | 5.3 ± 2.6 | 5.3 ± 2.7 |
| Platelet count (Giga/L) | 215.6 ± 64.2 | 215.9 ± 65.4 |
| Length of biopsy (mm) | 25 ± 8.3 | 25.5 ± 8.4 |
| Number of portal tracts | 21 ± 8.4 | 20.8 ± 8.3 |
| Liver fibrosis according to METAVIR (%) | | |
| F0 | 29 (6.6%) | 25 (6.5%) |
| F1 | 207 (47.5%) | 179 (46.9%) |
| F2 | 68 (15.6%) | 57 (14.9%) |
| F3 | 71 (16.3%) | 65 (17.0%) |
| F4 | 61 (14.0%) | 56 (14.7%) |

*Results are expressed as mean ± one standard deviation.
AST = aspartate aminotransferase; ALT = alanine aminotransferase; GGT = gamma glutamyltranspeptidase.

of less than 15 mm and in 259 patients (49.8%) the length of biopsy was greater than 25 mm. The inter-observer κ agreement was 0.48 and the weighted κ agreement was 0.75.

*Test performances*

For the diagnosis of significant fibrosis (Table 2) in the 436 patients having all the tests, no significant difference was observed between Fibrometer®, Hepascore®, Fibrotest®, and ELFG. Fibrometer® was significantly more accurate than Forns's score, APRI, MP3, FIB-4, and Hyaluronate. Adjusted AUROCs (Obuchowski) showed that Fibrometer® and Hepascore® performed equivalently and were significantly superior to all the other tests. In the 382 patients with both blood tests and interpretable Fibroscan™ observed-and adjusted-AUROCS were not statistically different between Fibrometer®, Fibrotest®, Hepascore®, and Fibroscan™.

For the diagnosis of cirrhosis, we compared only tests designed for this diagnosis. All tests (except Fib-4) performed equivalently in both the studied populations (Table 3).

To differentiate F1 and F2 (Supplementary Table S6) all tests performed equivalently with the exception of Hyaluronate, where Fibrometer® was significantly better ($p_{Sidak}$ = 0.002).

In addition, we looked at the percentage of well-classified patients using the previously published cut-offs for the main blood tests and Fibroscan™ (Table 4). This percentage varied between 63.6% and 73.8% for the diagnosis of significant fibrosis and between 79.6% and 87.7% for cirrhosis in the 382 patients having all tests.

*Combinations of tests*

As shown in Table 5 the number of well-classified patients for the diagnosis of significant fibrosis increases from 70–73% for the

**Table 2. Observed AUROCs and adjusted AUROCs (Obuchowski) of blood tests and Fibroscan™ for significant fibrosis (F ⩾2).**

| | n = 436* | | | | | n = 382‡ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AUROC | 95% CI | *p* Sidak | Obuch-owski | *p* | AUROC | 95% CI | *p* Sidak | Obuch-owski | *p* |
| FIBROMETER® | 0.82 | [0.78;0.86] | | 0.85 | | 0.83 | [0.80;0.87] | | 0.86 | |
| FIBROTEST® | 0.80 | [0.75;0.84] | 0.421 | 0.83 | 0.040 | 0.81 | [0.77;0.85] | 0.711 | 0.84 | 0.056 |
| FORNS' score | 0.75 | [0.71;0.80] | 0.004 | 0.79 | <0.001 | 0.77 | [0.72;0.82] | 0.011 | 0.81 | <0.001 |
| APRI | 0.76 | [0.72;0.81] | 0.005 | 0.79 | <0.001 | 0.78 | [0.73;0.82] | 0.010 | 0.80 | <0.001 |
| MP3 | 0.76 | [0.71;0.80] | 0.049 | 0.79 | <0.001 | 0.76 | [0.71;0.81] | 0.021 | 0.79 | <0.001 |
| ELFG | 0.78 | [0.74;0.83] | 0.266 | 0.82 | <0.001 | 0.78 | [0.74;0.83] | 0.069 | 0.82 | 0.004 |
| HEPASCORE® | 0.82 | [0.78;0.85] | 1.000 | 0.84 | 0.288 | 0.82 | [0.78;0.86] | 0.951 | 0.84 | 0.068 |
| FIB4 | 0.76 | [0.71;0.80] | 0.003 | 0.79 | <0.001 | 0.78 | [0.73;0.82] | 0.010 | 0.80 | <0.001 |
| HYALURONATE | 0.75 | [0.70;0.80] | 0.001 | 0.79 | <0.001 | 0.74 | [0.69;0.79] | <0.001 | 0.79 | <0.001 |
| FIBROSCAN™ (interpretable results) | - | - | - | - | - | 0.82 | [0.78;0.86] | 0.997 | 0.84 | 0.202 |

*CHC patients having all the blood tests; ‡CHC patients with all the tests and interpretable Fibroscan™.
CI = confidence interval.





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

**Table 3. Performance of blood tests and Fibroscan™ for the diagnosis of cirrhosis (F4).**

| | n = 436* | | | n = 382‡ | | |
|---|---|---|---|---|---|---|
| | AUROC | 95% CI | *p* Sidak | AUROC | 95% CI | *p* Sidak |
| FIBROMETER® | 0.89 | [0.86;0.93] | | 0.90 | [0.86;0.93] | |
| FIBROTEST® | 0.86 | [0.83;0.90] | 0.325 | 0.87 | [0.82;0.91] | 0.321 |
| APRI | 0.86 | [0.81;0.91] | 0.141 | 0.87 | [0.82;0.91] | 0.410 |
| ELFG | 0.88 | [0.83;0.92] | 0.883 | 0.87 | [0.83;0.92] | 0.860 |
| HEPASCORE® | 0.89 | [0.86;0.93] | 1.000 | 0.89 | [0.85;0.92] | 0.998 |
| FIB4 | 0.83 | [0.76;0.89] | 0.018 | 0.84 | [0.77;0.90] | 0.069 |
| FIBROSCAN™ (interpretable results) | - | - | - | 0.93 | [0.89;0.96] | 0.559 |

*CHC patients having all the blood tests; ‡CHC patients with all the tests and interpretable Fibroscan™.

best tests to 80–82% with the best combinations of tests. The proportion of "theoretically avoided liver biopsies" varied between 54% and 66% for the best combination (Fibrometer® and Hepascore®). For the diagnosis of cirrhosis no combination was superior to the best blood tests or Fibroscan™ alone in the "per-protocol" analysis (382 patients). However, when we considered the population of 436 patients ("intention to diagnose population") the combination of Fibroscan™ plus a blood test markedly improved the percentage of well classified patients for both significant fibrosis and cirrhosis.

*Other analyses*

We also calculated the number of "theoretically avoided liver biopsies" for the diagnosis of significant fibrosis using negative and positive predictive values of 90% (Supplementary Table S7). No difference was found between Fibrometer® (36.6%), Fibrotest® (35.6%), Hepascore® (30.5%), and interpretable Fibroscan™ (45.8%).

## Discussion

Blood tests and Fibroscan™ have been recently developed as alternatives to liver biopsy [24]. Retrospective studies [14,25,26] have compared several of these markers to liver biopsy but to our knowledge this is one of the first independent prospective validation of all relevant blood tests, and Fibroscan™ compared to liver biopsy in untreated patients with CHC. The true indicator of liver disease status would be the histological analysis of the entire liver, but impossible to obtain in routine practice and thus liver biopsy is considered at best as an "imperfect gold standard" [27]. Reduced sensitivity for the detection of significant fibrosis has been demonstrated with biopsies of less than 30 mm, fragmented specimens and steatosis. Concerning errors consecutive to the biopsy itself, Metha *et al.* [28] have demonstrated that the AUROC for a perfect marker would not exceed 0.90 or 0.83 according to 40% or 50% prevalence of significant disease in estimations where liver biopsy accuracy is highest (sensitivity and specificity of 90%). However, our study especially takes into consideration the methodological aspects so as to optimize the interpretation of the stage of fibrosis. Firstly, the liver specimens had to answer to quality criteria [29] to prevent a high risk of discordance for fibrosis staging [3,4,30]. Until now no study has included patients with such a high mean length of biopsy without fragmentation, cirrhosis excepted. By using the METAVIR scoring system, 65% of liver biopsies with a length of 15 mm are usually classified. This percentage increases to 75% for a length of 25 mm [3]. Also, a 25 mm biopsy is considered the optimal length for accurate liver evaluation. Considering this, in our study a sampling error for liver biopsy remains since only 50% of patients had a liver biopsy length greater than 25 mm. In addition, two senior liver pathologists independently reviewed biopsies [4] which were re-examined to reach a consensus in cases of disagreement. The agreement between the two expert pathologists was better than those previously published [4]. In order to exclude inter-laboratory variability the biochemical analyses were centralized with standardized methods and enzymatic calibration [31]. All serum samples were stored at −80 °C since the stability of different parameters could be affected by storage [32] such as marked transaminase activity loss at −20 °C [33].

The AUROCs of each test were comparable to those reported in the original publications [6–15,18,20] when expressed using observed-AUROCs according to the prevalence of stages defining advanced and non-advanced fibrosis. We observed similar AUROCs to those reported in meta-analyses [34–36] for the most validated biomarkers, Fibrotest®, Fibrometer®, and Apri and without major differences with interpretable Fibroscan™, Hepascore®, and ELFG. In diagnosing cirrhosis, the "Fibrostic" study [37] showed a significantly better performance of Fibroscan™ compared to serum markers while in contrast, our study shows that all the tests performed equivalently. This difference between these two recent multicentre studies might be due to the differences in design. Indeed in the "Fibrostic" study, Fibroscan™ was used in first intent and analysed apart from blood tests, while in our study we tried to compare in first intent all tests in "intention to diagnose". The methodology used for Fibroscan™ was equivalent in the two studies but the blood tests were performed in each centre in the Fibrostic study, using assay methods that might possibly have not always been homogeneous, while they were centralized in the Fibrostar study, except when impossible, and rigorously standardized analytical conditions were respected.

For differentiating between adjacent stages, F1 *vs.* F2, only Hyaluronate was inferior to Fibrometer®. For this adjacent comparison, AUROCs could appear low, but the performances were





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

# Research Article

**Table 4. Percentage of well classified patients in terms of the published cut-offs for the 382 patients with all tests and interpretable Fibroscan™.**

| Significant Fibrosis (F ≥2) | Published cut-off* | % well classified | Sensitivity (%) | Specificity (%) | NPV (%) | PPV (%) |
|---|---|---|---|---|---|---|
| FIBROMETER® | 0.411 | 70.9 | 87.6 | 56.4 | 83.9 | 63.7 |
| FIBROTEST® | 0.48 | 70.7 | 75.8 | 66.2 | 75.8 | 66.2 |
| APRI | 0.5 | 67.0 | 33.1 | 96.6 | 62.3 | 89.4 |
| HEPASCORE® | 0.5 | 73.6 | 74.7 | 72.5 | 76.7 | 70.4 |
| FIBROSCAN™ (interpretable results) | 5.2 | 63.6 | 96.6 | 34.8 | 92.2 | 56.4 |

| Cirrhosis (F4) | Published cut-off | % well classified | Sensitivity (%) | Specificity (%) | NPV (%) | PPV (%) |
|---|---|---|---|---|---|---|
| FIBROMETER® | 0.88(1) | 85.9 | 69.6 | 88.7 | 94.4 | 51.3 |
| FIBROTEST® | 0.74 | 79.6 | 71.4 | 81.0 | 94.3 | 39.2 |
| APRI | 2.0 | 86.1 | 7.1 | 99.7 | 86.2 | 80.0 |
| HEPASCORE | 0.84 | 80.6 | 76.8 | 81.3 | 95.3 | 41.3 |
| FIBROSCAN™ (interpretable results) | 12.9 | 87.7 | 76.8 | 89.6 | 95.7 | 55.8 |

*Degos *et al.* [37].
% well classified = comparison between the dichotomised score with the published cut-off and the stage of fibrosis; (1)cut-off with optimal PPV; CI, confidence interval.

**Table 5. Percentage of well classified patients and of theoretically avoided liver biopsies (in italics) according to one, or a combination of two tests (95% CI).**

| Significant Fibrosis (F≥2) | APRI | | FIBROMETER® | | HEPASCORE® | | FIBROTEST® | | FIBROSCAN™ | |
|---|---|---|---|---|---|---|---|---|---|---|
| APRI | 72% | [68-76] | | | | | | | | |
| FIBROMETER® | 78%<br>*62%* | [73-82]<br>*[57-66]* | 72% | [68-76] | | | | | | |
| HEPASCORE® | 80%<br>*60%* | [76-85]<br>*[55-64]* | 76%<br>66% | [72-81]<br>*[61-70]* | 73% | [69-77] | | | | |
| FIBROTEST® | 80%<br>*57%* | [75-84]<br>*[52-62]* | 76%<br>*63%* | [71-80]<br>*[58-68]* | 76%<br>*64%* | [71-80]<br>*[59-68]* | 70% | [66-75] | | |
| FIBROSCAN™ | 78%<br>*61%* | [73-83]<br>*[55-66]* | 81%<br>*59%* | [76-86]<br>*[54-64]* | 82%<br>*59%* | [77-86]<br>*[54-64]* | 82%<br>*57%* | [76-86]<br>*[52-62]* | 72% | [67-76] |
| FIBROSCAN™ In Intention to Diagnose | 78%<br>*60%* | [73-82]<br>*[55-64]* | 80%<br>*56%* | [75-85]<br>*[54-64]* | 81%<br>*57%* | [76-85]<br>*[52-62]* | 80%<br>*54%* | [75-84]<br>*[50-59]* | 63% | [58-68] |
| **Cirrhosis (F4)** | **APRI** | | **FIBROMETER®** | | **HEPASCORE®** | | **FIBROTEST®** | | **FIBROSCAN™** | |
| APRI | 86% | [83-90] | | | | | | | | |
| FIBROMETER® | 89%<br>*84%* | [85-92]<br>*[80-87]* | 87% | [84-90] | | | | | | |
| HEPASCORE® | 91%<br>*83%* | [87-93]<br>*[79-87]* | 90%<br>*85%* | [86-92]<br>*[81-88]* | 88% | [85-91] | | | | |
| FIBROTEST® | 90%<br>*83%* | [86-92]<br>*[79-86]* | 91%<br>*83%* | [87-93]<br>*[77-86]* | 91%<br>*84%* | [87-93]<br>*[80-87]* | 87% | [83-90] | | |
| FIBROSCAN™ | 93%<br>*84%* | [90-95]<br>*[79-87]* | 93%<br>*85%* | [90-96]<br>*[81-88]* | 93%<br>*86%* | [90-95]<br>*[82-89]* | 93%<br>*85%* | [90-96]<br>*[81-88]* | 92% | [88-94] |
| FIBROSCAN™ In Intention to Diagnose | 93%<br>*84%* | [90-95]<br>*[81-88]* | 93%<br>*85%* | [90-95]<br>*[81-88]* | 93%<br>*86%* | [90-95]<br>*[82-89]* | 93%<br>*85%* | [90-95]<br>*[81-88]* | 80% | [76-84] |





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

similar relative to liver biopsy. Indeed comparison between a biopsy of 25 mm (mean length in our study) and the true gold standard consisting of a large surgical sample showed 25% of false negative/positives and an AUROC evaluated at 85% for F2 *vs.* F1 [3].

Failed Fibroscan™ or non reliable results occurred in 22% of patients. This proportion of non-interpretable Fibroscan™ is not so different from that recently reported in a large mono-center series [21] the principal reasons were age, obesity, and BMI. Indeed in our study, contrary to blood tests, Fibroscan™ was not centralized but performed in each center by several operators having different levels of experience. However "in intention to diagnose" the Fibroscan™ performance was markedly reduced due to 22% of non interpretability but as recently published by Poynard *et al.* [38] applying manufacturers' recommendations increased the strength of concordance between Fibroscan™ and blood tests.

As reported, AUROCs may also vary according to the prevalence of each stage of fibrosis within the studied population (spectrum bias) especially when extreme stages (F0 and F4) are over-represented. In order to prevent this spectrum bias we used the Obuchowski measure. The Obuchowski measure [22,23] summarizes all pair-wise comparisons. Here it eliminated the bias related to the distribution of fibrosis stages and corrected the inflated type I error. By this measure and only in patients having all tests, Fibrometer® Fibrotest®, Hepascore®, and interpretable Fibroscan™ were the most accurate tests compared to liver biopsy. The choice of a linear penalty function to quantify the difference between observed and predicted fibrosis is open to discussion. However as previously reported [23], a linear function could have been used instead and would have permitted a comparison of the discriminative ability of these tests.

We evaluated combinations of tests in order to improve the diagnostic performance for significant fibrosis and cirrhosis. As previously published [16,39–42] we found that a synchronous algorithm combining Fibrotest®, Fibrometer® or Hepascore® and Fibroscan™ improved the accuracy for significant fibrosis and markedly decreased the requirement for biopsy. When Fibroscan™ was not interpretable; Apri in combination with one of the three best blood tests could be used. For the diagnosis of cirrhosis, contrary to recent studies [43,44] the diagnostic performance of Fibroscan™ and the three best blood tests were similar. Indeed, a combination seems to be unnecessary.

We also tested the applicability of the tests for the diagnosis of significant fibrosis. The values outside the cut-offs are zones where the diagnostic accuracy of the test is considered sufficiently reliable for use in clinical practice, and biopsy could be theoretically avoided. Using the conventional definition based on 90% NPV and 90% PPV, interpretable Fibroscan™, Fibrometer®, Fibrotest®, and Hepascore® performed equivalently and were better at discriminating than all other tests, confirming by another statistical method their higher accuracy.

Finally we calculated the diagnostic performance of the tests using previously published cut-offs [37]. No substantial difference was observed in the classification of tests when we compared published cut-offs and our cut-offs.

In conclusion this multicentre prospective and independent study definitely confirms the importance of non invasive markers to assess liver fibrosis in CHC. Contrarily to blood tests, performance of Fibroscan™ was reduced due to 22% of results not being interpretable. Fibrometer®, Hepascore®, and Fibrotest® performed better than all other blood tests and similarly to interpretable Fibroscan™. The combination of one of the three best blood tests with Fibroscan™, or Apri, improves the diagnostic performance for significant fibrosis. For the diagnosis of cirrhosis one of the best blood tests or Fibroscan™, when interpretable, can be used alone.

## Conflict of interest

Dr. Hubert-Fouchard holds stocks/stock options in 'Biolivescale' who market 'Fibrometer®'. There are no other conflicts of interest.

## Financial support

This study was funded by the French Agency for Research on AIDS and Viral Hepatitis (A.N.R.S) which gave grants to the participating hospitals and agreed to submission of the manuscript for publication.

## Acknowledgments

This study was supported by the French Clinical Biology Society (Société Francaise de Biologie Clinique) and the Association for Liver Research (l'Association pour l'Etude du Foie). We thank Jérôme Lambert and Fabrice Carrat for explaining and providing the program for calculating the Obuchowski measure; Sophie Thoret, Julie France and Jean-Louis Quesada for statistical analysis and Alison Foote (Inserm CIC3) for critically editing the manuscript.

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhep.2011.05.024.

## References

[1] Cadranel JF, Rufat P, Degos F. Practices of liver biopsy in France. Results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). Hepatology 2000;32:477–481.
[2] Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol 2002;97:2614–2618.
[3] Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;38:1449–1457.
[4] Rousselet MC, Michalak S, Dupre F, Croué A, Bedossa P, Saint-André J-P, et al. Sources of variability in histological scoring of chronic viral hepatitis. Hepatology 2005;41:257–264.
[5] Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. Hepatology 1996;24:289–293.
[6] Halfon P, Bourliere M, Penaranda G, Deydier R, Renou C, Botta-Fridlund D, et al. Accuracy of hyaluronic acid level for predicting liver fibrosis stages in patients with hepatitis C virus. Comp Hepatol 2005;4:6.
[7] 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.





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024

# Research Article

[8] Wai CT, Greenson J, 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.

[9] Cales P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet M-C, Konaté A, et al. A novel panel of blood markers to assess the degree of liver fibrosis. Hepatology 2005;42:1373–1381.

[10] Adams L, 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.

[11] Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. Hepatology 2007;46:32–36.

[12] Forns X, Ampurdanes S, Llovet J, Aponte J, Quintó L, Martínez-Bauer E, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. Hepatology 2002;36:986–992.

[13] 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.

[14] Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P, et al. Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. J Hepatol 2007;46:775–782.

[15] Rosenberg W, Voelker M, Thiel R, Burt A, Schuppan D, Hubscher S, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004;127:1704–1713.

[16] 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.

[17] Friedrich-Rust M, Ong MF, Martens SJ, 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.

[18] Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. Hepatology 2005;41:48–54.

[19] 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.

[20] Becker L, Salameh W, Sferruzza A, Zhang K, Chen R, Malik R, et al. Validation of Hepascore, compared to simple indices of fibrosis, in US patients with chronic hepatitis C virus infection. Clin Gastroenterol Hepatol 2009;7:696–701.

[21] Castera 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.

[22] Obuchowski N. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Stat Med 2006;25:481–493.

[23] Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. Clin Chem 2008;54:1372–1378.

[24] Sebastiani G. Non-invasive assessment of liver fibrosis in chronic liver diseases: implementation in clinical practice and decisional algorithms. World J Gastroenterol 2009;15:2190–2203.

[25] Bourliere M, Penaranda G, Renou C, Botta-Fridlund D, Tran A, Portal I, et al. Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies. J Viral Hepat 2006;13:659–670.

[26] Halfon P, Bacq Y, De Muret A, Penaranda G, Bourliere M, Ouzan D, et al. Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C. J Hepatol 2007;46:395–402.

[27] Bedossa P, Carrat F. Liver biopsy: the best, not the gold standard. J Hepatol 2009;50:1–3.

[28] Mehta S, Lau B, Afdhal N, Thomas D. Exceeding the limits of liver histology markers. J Hepatol 2009;50:36–41.

[29] 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.

[30] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. Hepatology 1994;20:15–20.

[31] Myara A, Guechot J, Imbert-Bismut F, Lasnier E, Piton A, Voitot H, et al. Harmonization of liver enzyme results: calibration for aminotransferases and gamma glutamyltransferase. J Hepatol 2004;41:501–502.

[32] Imbert-Bismut F, Messous D, Thibault V, Myers R, Piton A, Thabut D, et al. Intra-laboratory analytical variability of biochemical markers of fibrosis (Fibrotest) and activity (Actitest) and reference ranges in healthy blood donors. Clin Chem Lab Med 2004;42:323–333.

[33] Williams K, Williams A, Kline L, Dodd R. Stability of serum alanine aminotransferase activity. Transfusion 1987;27:431–433.

[34] Poynard T, Morra R, Halfon P, Castera L, Ratziu V, Imbert-Bismut F, et al. BMC Gastroenterol 2007;7:40.

[35] Leroy V, Halfon P, Bacq Y, Boursier J, Rousselet MC, Bourlière M, et al. Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. Clin Biochem 2008;41:1368–1376.

[36] Lin ZH, Xin YN, Dong QJ, Wang Q, Jiang XJ, Zhan SH, et al. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. Hepatology 2011;53:726–736.

[37] Degos F, Perez P, Roche B, Mahmoudi A, Asselineau J, Voitot H, et al. for the FIBROSTIC study group. Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study) J Hepatol 2010;53:1013–1021.

[38] Poynard T, Ingiliz P, Elkrief L, Munteanu M, Lebray P, Morra R, et al. Concordance in a world without a gold standard: a new non-invasive methodology for improving accuracy of fibrosis markers. PLoS One 2008;8:1–8.

[39] Boursier J, Vergniol J, Sawadogo A, Dakka T, Michalak S, Gallois Y, et al. The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. Liver Int 2009;29:1507–1515.

[40] Castéra L, Sebastiani G, Le Bail B, de Lédinghen V, Couzigou P, Alberti A. Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. J Hepatol 2010;52:191–198.

[41] 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.

[42] 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.

[43] Castéra L, Le Bail B, Roudot-Thoraval F, Bernard PH, Foucher J, Merrouche W, et al. Early detection in routine clinical practice of cirrhosis and oesophageal varices in chronic hepatitis C: comparison of transient elastography (FibroScan) with standard laboratory tests and non-invasive scores. J Hepatol 2009;50:59–68.

[44] Cross TJ, Calvaruso V, Maimone S, Carey I, Chang TP, Pleguezuelo M, et al. Prospective comparison of Fibroscan, King's score and liver biopsy for the assessment of cirrhosis in chronic hepatitis C infection. J Viral Hepat 2010;17:546-554.





Please cite this article in press as: Zarski J-P et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: The ANRS HCEP-23 study. J Hepatol (2011), doi:10.1016/j.jhep.2011.05.024