Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests

P. Calès, J. Boursier, V. de Lédinghen, P. Halfon, Y. Bacq, V. Leroy, N. Dib, F. Oberti, A. Sawadogo, M.-C. Rousselet, F. Lunel

Summary

Objective. — To evaluate the rates of reliable diagnosis of cirrhosis by two usual blood tests.

Methods. — Reliable diagnosis was mainly evaluated by comparing rates of positive (PPV) and negative (NPV) predictive values with FibroTest and FibroMeters, as either standard test or specifically designed for cirrhosis, in 1056 patients with chronic hepatitis C.

Results. — Using the diagnostic limits provided by fibrosis stage scales, the PPV for cirrhosis was: standard FibroMeters: 68.5% versus FibroTest: 37.1%. Using 95% PPV, the cirrhosis detection rate was: specific FibroMeter: 26.1% versus FibroTest: 2.0% ($P < 10^{-3}$). The cirrhosis detection rate increased from 26 to 65% by performing liver biopsy in 8% of patients with indeterminate results on specific FibroMeter between 95% NPV and PPV. On the other hand, specific FibroMeter provided three intervals of 95% reliable diagnosis with no biopsy: less than or equal to 95% NPV: no cirrhosis (threshold: diagnosis); significant fibrosis; and greater than or equal to 95% PPV: cirrhosis.
Conclusion. — The detection rate and PPV for cirrhosis using fibrosis scales were fair for standard FibroMeter and poor for FibroTest. Around one-fourth of cases of cirrhosis are detected by the 95% PPV of specific FibroMeter, and around two-thirds by performing an additional liver biopsy in only 8% of patients. Finally, specific FibroMeter can avoid liver biopsy by classifying patients into three categories: no cirrhosis; significant fibrosis; and cirrhosis.

Résumé
But. — Nous avons évalué le diagnostic fiable du Fibrotest et du FibroMètre pour le diagnostic de cirrhose.
Méthodes. — Le diagnostic fiable a été principalement évalué par les valeurs prédictives positive (VPP) et négative (VPN) chez 1056 patients atteints d’hépatite chronique C.
Résultats. — La VPP de cirrhose des compteurs de fibrose était : FibroMètre standard : 68,5 %, Fibrotest : 37,1 %. Le taux de détection de la cirrhose par la VPP 95 % était : FibroMètre spécifique de cirrhose 26,1 %, Fibrotest : 2,0 % (p < 10⁻³). Ce taux augmentait à 65 % en réalisant une biopsie chez 8 % des patients ayant un FibroMètre spécifique entre VPN et VPP 95 %. Alternativement, le FibroMètre spécifique permettait d’éviter toute biopsie en distinguant trois intervalles de diagnostic fiable à 95 % (seuil : diagnostic, respectivement) : VPN inférieure ou égale à 95 % : non cirrhose; fibrose significative; VPN supérieure ou égale à 95 % : cirrhose.
Conclusions. — La VPP fournie par les compteurs est assez bonne pour le FibroMètre standard et médicole pour le Fibrotest. Le FibroMètre spécifique détecte environ un quart des cirrhoses avec la VPP 95 % et environ deux tiers en ajoutant une biopsie chez 8 % des patients. Alternativement, le FibroMètre spécifique peut éviter toute biopsie en classant les patients en trois catégories : non cirrhose, fibrose significative, cirrhose.

© 2008 Elsevier Masson SAS. All rights reserved.

Abbreviations
AUROC area under the receiver operating characteristic
NPV negative predictive value
PPV positive predictive value
PACA Provence—Alpes—Côte d’Azur (French administrative region)
Se sensitivity
Spe specificity

Introduction
Blood tests for liver fibrosis are usually designed to diagnose significant fibrosis, including all stages with bridging fibrosis [1]. However, cirrhosis is also a clinically relevant target that warrants screening for liver complications, including esophageal varices and hepatocellular carcinoma. Unlike ultrasonographic elastometry [2,3], common blood tests have not been thoroughly evaluated for the diagnostic target of cirrhosis, especially in terms of diagnostic reliability.

Nevertheless, the diagnostic performance of several blood tests is considered excellent, especially in cirrhosis [4]. The use of FibroTest and ultrasonographic elastography was recently recommended by the French National Health Authority (Haute Autorité de santé, HAS) for the diagnosis of cirrhosis in untreated chronic hepatitis C patients without co-morbidity [5]. However, diagnostic accuracy depends on the blood-test value, with maximum accuracy at the extreme values and minimum accuracy with median values [4]. To apply these blood tests in clinical practice, the value intervals at which a reliable diagnosis can be made must be known. A reliable diagnosis corresponds to the intervals of blood-test values where the diagnostic accuracy is considered sufficiently reliable for clinical practice and, in patients with a reliable diagnosis, liver biopsy is considered avoidable. Previously, the intervals of reliable diagnosis were defined by the thresholds provided by 90% predictive values [6,7]. Yet, how those intervals were determined has never been examined in cirrhosis nor commented on in detail.

Recently, FibroMeter has shown higher accuracy for cirrhosis and more reliable diagnostic intervals for significant fibrosis than FibroTest [4]. However, these blood tests were specifically designed to diagnose significant fibrosis. For this reason, we have recently implemented a specific FibroMeter targeted at cirrhosis [8]. This test provides an improved positive predictive value (PPV) compared with the standard FibroMeter for the diagnosis of significant fibrosis in the pivotal study [9].

The aim of the present study was to comprehensively describe and compare the intervals of reliable diagnosis with Fibrotest [6] and FibroMeter [9], and to examine the possibility of improving reliable-diagnosis rates for cirrhosis to reduce the need for liver biopsy.

Methods
Data source
Using the Medline database and a manual search, we systematically reviewed the literature from 1997 to June 2007 for studies comparing FibroMeters and FibroTest in patients with chronic viral hepatitis C for whom liver biopsy data were available. Three independent publications were retrieved. The first study, involving one center—Angers—
was the original publication that described FibroMeters [9]. The second study included two independent centers—Provence—Alpes—Côte d’Azur (PACA) and Tours [10]—while the third was from Grenoble [7]. The PACA center incorporated three secondary—and two tertiary-care settings. Two additional, as yet unpublished, study populations were provided by Angers and Bordeaux [11]. Thus, we were able to perform a meta-analysis of individual patients’ data from five centers, that were independent for study design, patient recruitment, blood-marker determination or interpretation of liver histology.

**Patients**

Inclusion and exclusion criteria were similar across all five centers. Patients with chronic viral hepatitis C were prospectively included from 1994 to 2007 if they had anti-HCV antibodies, HCV RNA in serum, and available liver biopsy and blood markers. Fasting blood samples were collected immediately before, or no more than three months after, the liver biopsy was performed. Patients at the Tours and PACA centers were excluded if their liver specimens were less than 15 mm. Other exclusion criteria were additional causes of liver disease, particularly HIV or HBV co-infection, complicated cirrhosis, antifibrotic treatment in the previous six months, and alcohol consumption greater than 30 g/day in the five years prior to inclusion. Overall, the five centers provided 1535 patients, but 479 were excluded because of failure to meet inclusion criteria (such as HBV co-infection) or missing data (usually a blood marker), leaving a core study population of 1056 patients. The general characteristics of excluded patients did not significantly differ from those of included patients. The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and was approved by local ethics committees.

**Blood measurements**

Blood samples were independently processed at each center. The determined variables were: platelet count; urea; bilirubin; gamma-glutamyl transpeptidase; aspartate and alanine aminotransferases; prothrombin time; apolipoprotein A1; haptoglobin; hyaluronic acid; and alpha-2-macroglobulin. Direct markers were measured in either fresh blood or frozen serum stored at less than or equal to −20 °C. Indirect markers were usually measured in fresh blood. Automated and assay techniques varied among the centers except for apolipoprotein, alpha-2-macroglobulin (Dade Behring, Marburg, Germany) and hyaluronic acid (Corgenix, Broomfield, CO, USA and Biogenic SA, Pérols, France). Blood tests were calculated according to published patents or formulas [9,11]. The FibroMeter score was slightly improved by taking patients’ gender into account as described in a previous study [12]. Several studies performed in numerous laboratories have shown excellent interlaboratory reproducibility [13]. The area of liver fibrosis [14], expressed as a percentage of the whole liver, was estimated using the FibroMeter test specifically designed for that measurement [9].

**Liver biopsy**

Liver biopsies were performed using Menghini’s technique and a 1.4- to 1.6-mm needle diameter. Biopsy specimens were fixed in a formalin–alcohol–acetic acid solution and stained with hematoxylin–eosin–safron. Liver fibrosis was staged from F0—F4, according to the METAVIR staging system [15]. Three diagnostic targets were defined: significant fibrosis (main target), F2 + F3 + F4; severe fibrosis, F3 + F4; and cirrhosis, F4. Readings were carried out by independent blinded senior pathologists specializing in hepatology. Histological assessments were made twice by the same pathologist in Grenoble, once in Bordeaux and once by two pathologists each in Angers, Tours and PACA, with a consensual final reading in cases of disagreement.

**Statistics**

Data were analyzed according to STARD statements [16] and, thus, are reported on an intention-to-diagnose basis. Two kinds of cut-offs are used for blood-test values. The diagnostic cut-off can distinguish patients according to the diagnostic target and determine overall accuracy. However, this diagnostic cut-off was not available because the present diagnostic target (cirrhosis) is different from that (significant fibrosis) of the pivotal study. We thus calculated it a posteriori according to the maximum Youden index: sensitivity (Se) + specificity (Spe)—1. The other cut-off points used—called ‘thresholds’ here—were those of the predictive values.

The size of the population necessary to detect a significant difference between FibroMeter and Fibrotest was

| Table 1 | Diagnostic indices of blood tests for cirrhosis, with diagnostic cut-offs determined a posteriori according to the maximum Youden index.
| Inches diagnostiques des tests sanguins pour la cirrhose et seuils diagnostiques déterminés a posteriori selon l’index de Youden maximum. |
|----------------|----------------|----------------|----------------|
|                | Youden index | Sensitivity | Specificity |
| Standard FibroMeter | 0.628       | 95.7         | 71.1         |
| Specific FibroMeter  | 0.089       | 87.0         | 81.0         |
| Fibrotest            | 0.660       | 82.4         | 77.2         |

<table>
<thead>
<tr>
<th>Predictive values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>99.2</td>
</tr>
<tr>
<td>97.9</td>
</tr>
<tr>
<td>97.5</td>
</tr>
</tbody>
</table>
Table 2 Negative (NPV) and positive (PPV) predictive values for categories of cirrhosis and ‘borderline’ cirrhosis (F3/4), according to the limits of blood-test meters and corresponding rates (%) of cirrhosis detected.

<table>
<thead>
<tr>
<th>Fibrosis stage by blood-test meter</th>
<th>Cirrhosis by liver biopsy</th>
<th>Standard FibroMeter</th>
<th>Fibrotest</th>
<th>P^a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Meter limit</td>
<td>0.982</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NPV (%)</td>
<td>91.9</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV (%)</td>
<td>68.5</td>
<td>37.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cirrhosis detected (%)</td>
<td>32.2</td>
<td>61.8</td>
<td>&lt;10^-3</td>
</tr>
<tr>
<td>F4</td>
<td>Meter limit</td>
<td>0.834</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>&amp; F3/4</td>
<td>NPV (%)</td>
<td>96.4</td>
<td>95.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PPV (%)</td>
<td>39.8</td>
<td>35.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cirrhosis detected (%)</td>
<td>74.8</td>
<td>64.7</td>
<td>0.078</td>
</tr>
</tbody>
</table>

^a By McNemar test.
^b Borderline stage.

determined according to the findings of a previous study [12] of the diagnosis of cirrhosis. With an α risk of 0.05, β risk of 0.2, cirrhosis prevalence of 0.11, AUROC correlation of 0.75 and bilateral testing, the required sample size was 910 patients for the following AUROC values: FibroMeter: 0.93, Fibrotest: 0.89 [12]. Predictive values were indirectly obtained using mainly predictive values and, incidentally, Se and Sp [4]. For example, with the usual definition based on 90% predictive values, a reliable diagnosis for a patient means a greater than or equal to 90% chance of being F0 – F3 at the lower interval of blood-test values by NPV and greater than or equal to 90% chance of being F4 at the highest interval by PPV. The indeterminate interval is the zone outside of these reliable intervals. The statistical software programs used were SPSS 11.5.1 for Windows (SPSS Inc., Chicago, IL, USA) and SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

Results

Characteristics of patients

The principal characteristics of the 1056 patients were: gender: 59.5% male; mean age: 45.6 ± 12.5 years; METAVIR fibrosis stage: F0: 4.4%, F1: 43.5%, F2: 27.0%, F3: 14.0%, F4: 11.2%. Mean liver specimen length was 21 ± 8 mm, and 58.2% were greater than or equal to 20 mm.

Predictive values

Table 1 shows the overall predictive values for cirrhosis with Fibrotest and FibroMeter (for all values falling into two diagnostic categories according to diagnostic cut-off). The NPV was high (≥ 97.5%), but PPV was low (≤ 29.7%).

In addition, a 100% NPV was available for a large proportion of patients: 47.5% with standard FibroMeter versus 31.9% for Fibrotest (P < 10^-3). The PPV peak was 88% for standard FibroMeter and 100% for Fibrotest, although these thresholds were not clinically applicable as they included only a small proportion of patients (standard FibroMeter: 1.7%; Fibrotest: 0.2%; P < 10^-3).

In addition to the binary diagnosis of significant fibrosis, these two blood tests provide a scale that translates blood-test values into METAVIR fibrosis stages [12,17]. Thus, the available diagnostic limits of blood-test scale by which patients are classified into the cirrhosis stage enabled calculation of the corresponding PPV for cirrhosis: 68.5% for standard FibroMeter and 37.1% for Fibrotest (Table 2). However, the highest PPV with the standard FibroMeter was obtained at the expense of a lower rate of cirrhosis detection by the test limit: 32% for standard FibroMeter versus 62% for Fibrotest (P < 10^-3). The NPV for cirrhosis provided by these diagnostic limits were 91.9% for standard FibroMeter and 95.3% for Fibrotest.

Specific blood test

We have devised a specific FibroMeter test for the diagnosis of cirrhosis using the same variables as the standard FibroMeter, but constructed for significant fibrosis resulting in different coefficients of regression score [8]. Their values displayed different curves (Fig. 1, upper) and means: 0.50 ± 0.30 (mean ± S.D.) for standard FibroMeter, and 0.11 ± 0.21 for specific FibroMeter for cirrhosis (P < 10^-3).

The specific FibroMeter was able to provide a 100% PPV for cirrhosis in 1.3% of patients (11.3% of patients with cirrhosis) versus 0.2% with Fibrotest (P = 0.004). In addition, the PPV peak (100%) of the specific FibroMeter was well-shaped, following a progressive curve (Fig. 1, lower). In contrast, the PPV peak (88%) of the standard FibroMeter was followed by a marked decrease that precluded reliable clinical use (Fig. 1, middle). The 100% NPV threshold of the specific FibroMeter included only 15.6% of patients versus 47.5% (P < 10^-3) for the standard FibroMeter. In contrast, the proportion of patients with PPV greater than or equal to 75% was higher with the specific FibroMeter than with standard FibroMeter: 8.4% versus 3.3%, respectively; P < 10^-3 (Fig. 2).
Table 3 Comparison of reliable-diagnosis intervals for cirrhosis as a function of different levels of negative (NPV) and positive (PPV) predictive values for cirrhosis and FibroMeters type (standard or specific for cirrhosis): note that certain results do not precisely fit the expected fixed values due to available raw thresholds and subsequent changes in observed calculated results.

Comparaison des intervalles de diagnostic fiable pour la cirrhose en fonction de différents niveaux de valeurs prédicitives négative (NPV) et positive (PPV) de cirrhose et du type de FibroMètre (standard ou spécifique de cirrhose) : à noter que certains résultats ne correspondent pas exactement à la valeur fixée attendue du fait des seuils bruts disponibles et des changements induits dans les résultats calculés.

<table>
<thead>
<tr>
<th>FibroMeter</th>
<th>Threshold (%)</th>
<th>Interval(s) NPV</th>
<th>NPV</th>
<th>PPV</th>
<th>NPV + PPV</th>
<th>NPV + PPV</th>
<th>Entire population</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>100</td>
<td>47.5</td>
<td>0²</td>
<td>47.5</td>
<td>100</td>
<td>47.5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>85.6</td>
<td>0²</td>
<td>85.6</td>
<td>95.0</td>
<td>81.3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>98.0</td>
<td>0²</td>
<td>98.0</td>
<td>90.0</td>
<td>88.3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Specific</td>
<td>100</td>
<td>15.6</td>
<td>1.3</td>
<td>16.9</td>
<td>100</td>
<td>16.9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>88.8</td>
<td>3.1</td>
<td>91.9</td>
<td>95.1</td>
<td>87.4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>98.3</td>
<td>4.6</td>
<td>97.1</td>
<td>92.5</td>
<td>89.7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Standard/specific</td>
<td>100</td>
<td>47.5</td>
<td>1.3</td>
<td>48.8</td>
<td>99.4</td>
<td>48.5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>85.6</td>
<td>3.1</td>
<td>88.7</td>
<td>95.8</td>
<td>84.9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>98.0</td>
<td>4.6</td>
<td>97.4</td>
<td>92.5</td>
<td>90.0</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

a Proportion of patients included in the interval(s) defined by the predictive value(s) in the entire study population.
b Patients with a correct diagnosis (cirrhosis or no cirrhosis) by the predictive values within their intervals or in the entire population.
c Decreasing order of clinical interest based on threshold level and rate of patients correctly classified in the entire population.
d Not available.
e Grey zone of 3% of patients where NPV threshold was superior to PPV threshold.
f Standard for NPV/specific for PPV.
Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests

Figure 1  Comparison of standard and specific FibroMeters: (upper) blood-test values as a function of patients ranked by increasing blood-test values. The curves of FibroMeters (FM) designed for clinically significant fibrosis (CSF) and specifically for cirrhosis (F4) in the other graphs cannot be compared because of significant differences in blood-test values (see Fig. 2 for comparison with PPV); (middle and lower) negative (NPV) and positive (PPV) predictive values, and diagnostic accuracy (DA) for cirrhosis as a function of blood-test values with standard and specific FibroMeters.

Comparaison des FibroMètres standard et spécifique. Schéma supérieur : valeurs des tests sanguins en fonction des patients ordonnés par valeurs croissantes du test sanguin. Les courbes des FibroMètres (FM) construits pour la fibrose cliniquement significative (CSF) et spécifiquement pour la cirrhose (F4) ne peuvent être comparées, dans les schémas suivants, du fait d’une différence significative entre les valeurs des tests sanguins (voir Fig. 2 pour une comparaison des VPP). Schémas inférieurs : valeurs prédictives négative (NPV) et positive (PPV), et performance diagnostique (DA) pour la cirrhose en fonction des valeurs des tests sanguins du FibroMètre standard ou spécifique.

Reliable diagnosis

To optimize a reliable diagnosis, the calculations focused on the blood tests with the highest reliable diagnosis.

Determination of the best interval

Table 3 illustrates different combinations of reliable-diagnosis intervals as a function of thresholds of predictive values and FibroMeters type (standard or specific) for cirrhosis. The most clinically relevant predictive values were those provided by the 95% thresholds of specific FibroMeter, which avoided liver biopsy in 91.9% of patients. The rate of patients with cirrhosis (according to liver biopsy) in the corresponding intervals of specific FibroMeter values were: 0 to less than or equal to 95% NPV threshold: 39.1%; others: 34.8%; greater than or equal to 95% PPV threshold to 1: 26.1% (Table 4).

Table 4  Comparison of patients and cirrhosis rates (%) as a function of 95% negative (NPV) and positive (PPV) predictive value thresholds between specific FibroMeter and Fibrotest.

<table>
<thead>
<tr>
<th>Test</th>
<th>≤ 95% NPV</th>
<th>Others</th>
<th>&gt; 95% PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>Specific</td>
<td>88.8</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>FibroMeter</td>
<td>87.4</td>
<td>12.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.038</td>
<td>&lt;10⁻³</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td>Specific</td>
<td>39.1</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>FibroMeter</td>
<td>42.2</td>
<td>55.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.557</td>
<td>0.007</td>
</tr>
</tbody>
</table>

* McNemar test.
In terms of certainty (exclusion or affirmation of cirrhosis), the most clinically relevant predictive values were those provided by 100% thresholds of NPV by standard FibroMeters and of PPV by specific FibroMeter, which avoided liver biopsy in 49% of patients (Table 3).

Comparisons between blood tests
A 95% PPV was provided by Fibrotest in 0.2% of patients versus 3.1% \( (P < 10^{-3}) \) by specific FibroMeter (Table 4). A 95% NPV was provided by Fibrotest in 87.4% of patients, which was not significantly different from that with standard FibroMeter \( (85.6\%, P = 0.324) \), but significantly lower than with the specific FibroMeter \( (88.8\%, P = 0.038) \) (Table 4). The rate of patients with a correct diagnosis (cirrhosis or no cirrhosis) was significantly higher with specific FibroMeter \( (87.4\%) \) than with Fibrotest \( (83.3\%; P < 10^{-3}) \) using the 95% predictive values (Table 5). More important, the detection rate of cirrhosis (according to 95% PPV) was 26.1% with specific FibroMeter versus 2.0% with Fibrotest \( (P < 10^{-3}) \), whereas the rate of indeterminate results or cirrhosis in the indeterminate category was significantly higher with Fibrotest than with specific FibroMeter (Table 4).

Use of sensitivity (Se) and specificity (Spe)
We have compared Se and Spe—diagnostic indices useful for populations—with the most clinically useful predictive values—diagnostic indices useful for a single patient—of the specific FibroMeters (at the 95% thresholds). The reliable-diagnosis intervals defined by Spe and Se produced an indeterminate zone of 30.5% of patients, which was significantly higher than that by predictive values (8.1%) (based on Table 5). As a function of specific FibroMeter thresholds, the rate of patients with cirrhosis (according to liver biopsy) was: 0 to less than or equal to 95% Se; 5.2%; others: 33.0%; greater than or equal to 95% Spe to 1: 61.7%. So, as expected, the proportion of patients included in the intervals of 95% Se and Spe, and the rate of patients correctly classified in the whole population, were significantly lower than with the intervals of 95% predictive values (Table 4). On the other hand, the rate of cirrhosis detected was significantly higher with Spe than with Se and with predictive values: 62% versus 26% \( (P < 10^{-3}) \), whereas the subsequent PPV decreased to 62% (Table 5).

Optimization of patient classification
Fig. 3 shows the distribution of METAVIR stages as a function of 95% predictive value thresholds of specific FibroMeter. In the indeterminate interval (outside of the 95% predictive values), the prevalence of fibrosis stages was: cirrhosis \( (F4) \): 48.8%; severe fibrosis \( (F3 + F4) \): 80.5%; and significant fibrosis \( (F2 + F3 + F4) \): 95.1%. Thus, three reliable-diagnosis intervals could be distinguished according to 95% predictive value thresholds: less than or equal to 95% NPV: no cirrhosis (accuracy: 95.1%); others: significant fibrosis (accuracy: 96.8%), with an overall accuracy of 95.1% in the entire study population. In addition, the severity of cirrhosis increased as a function of these three intervals \( (<95\% \text{NPV}, \text{others and } \geq 95\% \text{PPV}, \text{respectively}) \), as indicated by the prothrombin index (%): 90.9 \( \pm 8.4 \); 78.7 \( \pm 11.4 \); and 72.1 \( \pm 13.2 \) \( (P < 10^{-3}) \) by ANOVA, and \( P < 0.04 \) within each

Table 5
<table>
<thead>
<tr>
<th>Blood test Interval(s)</th>
<th>Patients included (%)a</th>
<th>Patients correctly classified by diagnostic indices (%)b</th>
<th>NPV or Se PPV or Spe</th>
<th>NPV + PPV or Se + Spe</th>
<th>NPV + PPV or Se + Spe Entire population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific FibroMeter Se, Spe</td>
<td>&lt; 10^{-3}</td>
<td>92.9</td>
<td>92.9</td>
<td>89.1</td>
<td>91.9</td>
</tr>
<tr>
<td>Specific FibroMeter NPV, PPV</td>
<td>69.5</td>
<td>69.5</td>
<td>69.5</td>
<td>69.5</td>
<td>69.5</td>
</tr>
<tr>
<td>Specific FibroMeter P</td>
<td>1056 P. Calès et al.</td>
<td>1056 P. Calès et al.</td>
<td>1056 P. Calès et al.</td>
<td>1056 P. Calès et al.</td>
<td>1056 P. Calès et al.</td>
</tr>
</tbody>
</table>

- Patients included (%)a
- Patients correctly classified by diagnostic indices (%)b
- NPV or Se PPV or Spe
- NPV + PPV or Se + Spe
- NPV + PPV or Se + Spe Entire population

Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests

Figure 3  METAVIR fibrosis (F) stages according to liver biopsy as a function of reliable-diagnosis intervals determined by thresholds of 95% negative (NPV) and positive (PPV) predictive values for cirrhosis, and an indeterminate interval (Others), using specific FibroMeter. The percentages in parentheses indicate the proportions of patients in the interval.

Discussion

Predictive value curves

The diagnostic accuracy of blood tests for liver fibrosis is thought to be satisfactory for clinical applications [18]. However, the accuracy is not optimal for certain ranges of blood-test values [4]. This is probably attributable to the limitations of the reference, such as poor interobserver reproducibility for intermediate stages of liver fibrosis as determined by liver biopsy [19]. Nevertheless, it is essential to define the reliable-diagnosis intervals where the diagnostic error is considered to be acceptable. On a patient-by-patient basis, the clinician’s choice relies on predictive values. Thus, diagnostic performance is reflected by predictive values in a patient, but by Se and Spe in a population. For example, the PPV reflects the accuracy of a positive blood test for a cirrhosis diagnosis whereas the Se reflects the proportion of patients with detected cirrhosis. Usually, predictive values are provided for all values of blood tests according to a single diagnostic cut-off point chosen to optimize overall accuracy—in other words, an overall predictive value. However, the level of predictive values depends on blood-test values (Fig. 1).

Choice of predictive values

The diagnostic accuracy of blood tests is high for the diagnosis of cirrhosis. However, this is global diagnostic accuracy, and the potential pitfall is the low prevalence of this diagnostic target (usually 10—15%). In this case, the focus needs to be on diagnostic indices within the cirrhosis stage, especially the rate of cirrhosis detected by PPV.

These data are provided by the available scales (or meters) of the standard FibroMeter and Fibrotest that translate a binary diagnosis of significant fibrosis into a classification that corresponds to histological fibrosis stages. Using these meters, the NPVs for cirrhosis were satisfactory: 91.9% for standard FibroMeter; and 95.3% for Fibrotest. However, the PPVs were disappointing for Fibrotest (37%) and suboptimal for FibroMeter (69%). However, the Se for cirrhosis of the standard FibroMeter was significantly less than that of Fibrotest, using the limits provided by the meters, resulting in a lower rate of detected cirrhosis (the opposite was observed when the F3/4 stage was added; [Table 2]).
The specific FibroMeter test, specially constructed for the diagnosis of cirrhosis with the same variables as those included in the standard FibroMeter, increased the PPV peak from 88 to 100%. This 100% PPV was observed in 1.3% of patients, and was significantly higher than the 0.2% of patients with Fibrotest. Finally, the rate of detected patients with cirrhosis was much higher with the specific FibroMeter (26%) than with Fibrotest (2%) using the 95% PPV.

Reliable-diagnosis intervals

The objective of reliable-diagnosis intervals is to obtain a reliable diagnosis for the largest number of patients. Considering the usual accepted 5% risk of error in biomedicine, the best choice was to use the reliable-diagnosis intervals provided by 95% predictive values of the specific FibroMeter (Table 3). The resulting intervals included 92% of the population, leaving a grey zone of 8% where liver biopsy could be performed. However, a three-category classification—no cirrhosis; significant fibrosis; cirrhosis—would avoid liver biopsy in all patients (Fig. 3) with an acceptable risk for a missed diagnosis of cirrhosis (see below).

If we assume a 0% error, the best choice would be to use the standard FibroMeter for 100% NPV and the specific FibroMeter for 100% PPV simultaneously. This is how both FibroMeters, including the same markers, are used on the professional website to calculate whether or not a patient’s result lies in the intervals of 100% predictive values. One advantage of this level of diagnostic certainty is that the thresholds of NPV and PPV are the same as those of Se and Spe, respectively. In that case, NPV and PPV are not directly dependent on the prevalence of the diagnostic target in the evaluated population and are, thus, more easily adapted to clinical practice. In other circumstances, the reproducibility of predictive values should be checked by further studies. Another solution is to use thresholds of Se and Spe (when <100%) that are more adapted to individual screening than to individual diagnosis. Thus, by using Se and Spe, the rate of detected cirrhosis was higher, but at the expense of a lower PPV compared with diagnoses based on predictive value thresholds (Table 5).

Using the 95% PPV of the specific FibroMeter, only 26% of patients with cirrhosis were detected. By including the 8% of patients requiring a liver biopsy in the indeterminate interval, this rate would increase to 65% of cirrhosis cases detected. These are respectively, however, low and fair detection rates, despite being higher than those with Fibrotest. However, any missed diagnoses of cirrhosis in the first reliable-diagnosis interval (<95% NPV) and even in the indeterminate interval are attenuated as they correspond to a significantly earlier stage of cirrhosis, as reflected by a significantly higher prothrombin index and a significantly lower area of fibrosis (Fig. 4) compared with cirrhosis diagnosed by greater than or equal to 95% PPV. This results in three categories of cirrhosis: very early; early; and established cirrhosis. As clinical complications are closely related to the amount of fibrosis, especially esophageal varices and hepatocellular carcinoma [3,20], the missed diagnoses should have only a minor impact on patient care.

Finally, knowing that fibrosis usually progresses, the diagnosis of cirrhosis in the patients with early cirrhosis on initial testing may become more evident later on with repeated blood-test determinations, which is easily done in clinical practice. Alternatively, if the diagnosis of cirrhosis is made in the interval less than 95% PPV, especially in the indeterminate interval, other new techniques could be used, such as combinations of non invasive tools—for example, sequential algorithms [21] or even synchronous algorithms which combine blood tests and transient elastography [11]. The new algorithms now being developed will increase the detection rate of cirrhosis.

Evaluation and improvement of a reliable diagnosis of cirrhosis by blood tests

Limitations

Non invasive tests are usually thought to be underestimations [22] due to the limitations of liver biopsy [19,23]. However, in terms of cirrhosis diagnosis and the lack of false positives by liver biopsy, the implementation of reliable-diagnosis intervals is probably more robust. A sensitivity analysis has shown that the exclusion of the pivotal population of FibroMeters and liver specimens less than 20 mm had a weak effect on the diagnostic accuracy of blood tests [4]. The influence of clinical center has been evaluated in another study [12], and found only minor non significant effects of centers on diagnostic indices of standard Fibrometer and significant changes for Fibrotest.

Clinical practice

Considering the importance of cirrhosis diagnosis, the acceptable rate of error is less than or equal to 5%. In Fig. 5, we have presented the resulting clinical dilemma. A physician (or patient) accepting a 5% error will promptly discard the cirrhosis diagnosis in the majority of patients while proposing a liver biopsy in a minority of patients. A physician (or patient) demanding a (known) 0% error will discard the cirrhosis diagnosis in a minority of patients while proposing a liver biopsy in a small majority of patients. Such choices should be available to the physician and/or patient. We prefer the former choice, given the constraints of liver biopsy and the challenge that a 0% error represents.

Conclusion

Using the thresholds of standard blood tests, the diagnosis of cirrhosis by FibroMeter is more accurate, but less sensitive than by Fibrotest. The use of a specific FibroMeter test markedly increases the detection rate. Thus, with 95% predictive values, the detection rate of cirrhosis is around one in four, but can be increased to around two out of three by a liver biopsy performed in a minority of patients with indeterminate results. In addition, patients with missed diagnoses of cirrhosis are at a very early stage of disease. This means that, considering the possibility of liver complications, it may not be deleterious to delay the initial diagnosis of very early cirrhosis and postpone its eventual diagnosis until a more advanced stage by repeating the blood test on, say, an annual basis. However, this proposal needs to be validated by further studies. Another diagnostic choice might be to extend this concept of missed diagnoses with no significant clinical consequences to patients with indeterminate results, thereby avoiding liver biopsy in every patient and resulting in a three-category classification of no cirrhosis, significant fibrosis and cirrhosis.

Acknowledgments

The authors wish to thank other investigators from: Angers: Sophie Michalak, Isabelle Hubert-Fouchard, Anselme Konaté, Catherine Ternisien, Alain Chevailler, Gilles Hunault; PACA: Marc Bourlière, Denis Ouzan, Albert Tran, Danielle Botta, Christophe Renou, Christophe Sattonnet, Marie-Christine Saint-Paul, Thierry Benderitter, Stéphane Garcia, Henri Pierre Bonneau, Guillaume Penarandaa; Tours: Anne de Muret, Marie-Claude Bréchot; Grenoble: Nathalie Sturm, Marie-Noelle Hilleret, Jean-Pierre Zarski, Patrice Faure, Jean-Charles Renversez, François Morel, Candice Trocmé; Bordeaux: Juliette Foucher, Laurent Castéra, Patrice Couzigou, Pierre-Henri Bernard, Wassil Merrouche, Paulette Bioulac-Sage, Brigitte Le Bail; and Cluny: Claude Degott, Valérie Paradis; and Kevin L. Erwin (for English proofreading).

Grant: PHRC in 1994 and 2002 from the French Department of Health.

Authors’ disclosure: Paul Cales, Frédéric Oberti and Françoise Lunel have stock ownership in BioLiveScale Inc. that has a license for FibroMeters from Angers University.

References