Standardization of ROC Curve Areas for Diagnostic Evaluation of Liver Fibrosis Markers Based on Prevalences of Fibrosis Stages

Thierry Poynard,1* Philippe Halfon,1 Laurent Castera,1 Mona Munteanu,2 Françoise Imbert-Bismut,1 Vlad Ratziu, Yves Benhamou,1 Marc Bourlière,1 Victor de Ledinghen,1 and the FibroPaca Group

Background: The area under the ROC curve (AUC) is widely used as an estimate of the diagnostic value for fibrosis markers. Whether there is variability in the AUC related to the prevalence of fibrosis stages defining advanced and nonadvanced fibrosis is unknown. The aim of this study was to assess the relationships between the AUC and the prevalence of each fibrosis stage and to elaborate simple methods of standardization.

Methods: The AUCs of FibroTest (FT) for the diagnosis of advanced fibrosis were estimated in patients with chronic hepatitis C using an integrated database including 1312 patients with FT and biopsy, and in an overview of 18 diagnostic studies.

Results: In the integrated database considering stage prevalence, the FT AUC for advanced fibrosis varied (P < 0.001) from 0.67 (only stage F2 as advanced fibrosis and only F1 as nonadvanced fibrosis) to 0.98 (only F4 as advanced fibrosis and only F0 as nonadvanced fibrosis). The same results were observed in the overview where the FT AUC varied (P < 0.001) from 0.65 to 0.89 according to fibrosis stage prevalence. Two approaches for expressing standardized AUCs were developed: one approach assumed a uniform prevalence distribution of each fibrosis stage; the other approach used the prevalence distribution of fibrosis stages observed in the population.

Conclusions: The expressions of the AUCs of fibrosis markers should be standardized according to the prevalence of fibrosis stages defining advanced and nonadvanced fibrosis.

© 2007 American Association for Clinical Chemistry

Background

A major clinical challenge is finding the best means of evaluating and managing the increasing numbers of patients infected with the hepatitis C virus (HCV)3 (1). Liver biopsy, because of its limitations and risks, is no longer considered mandatory as the first line indicator of liver injury, and many markers have been developed as noninvasive alternatives (2-4).

These new markers quantitatively estimate liver fibrosis and their diagnostic accuracy is evaluated using the area under the ROC curve (AUC) (2-4). The AUC combines the sensitivity and specificity of a given quantitative marker for the diagnosis of a specific definition of fibrosis. Sensitivity is usually assessed in patients with advanced fibrosis (i.e., stages F2, F3, and F4 in the METAVIR scoring system) (4, 5) and specificity in nonadvanced fibrosis (i.e., stages F0, F1).

Since 1991, we have identified parameters associated with fibrosis and fibrosis progression and constructed several panels combining these parameters, including the PGA index (6), age/platelets index (7), and the FibroTest (FT) (8). We and others have observed that 2 factors are associated with the AUCs: biopsy length and biopsy fragmentation (4, 9-14).

1 Assistance Publique-Hôpitaux de Paris Groupe Hospitalier Pitié-Salpêtrière, Paris, France.
2 Biopredictive.
* Address correspondence to this author at: Assistance Publique-Hôpitaux Groupe Hospitalier Pitié-Salpêtrière, 47-83 Boulevard de l’Hôpital, 75651 Paris Cedex 13, France. Fax 33-142161427; e-mail tpoynard@teaser.fr.
Received January 13, 2007; accepted June 14, 2007.
Previously published online at DOI: 10.1373/clinchem.2007.085795

Nonstandard abbreviations: HCV, hepatitis C virus; AUC, area under the ROC curve; FT, FibroTest; DANA, difference between the mean fibrosis stage of advanced fibrosis minus the mean fibrosis stage of nonadvanced fibrosis; eDANA, estimated DANA; AduAUC, adjusted uniform AUC; ObAUC, observed AUC; AdnAUC, naturally-observed adjusted AUC.
No study has examined the importance of the spectrum of stages defining advanced or nonadvanced fibrosis. In the last 25 years, 35 reviews and 25 editorials of noninvasive fibrosis markers have been identified (see Appendix in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue9), with 19 of them published in 2005 and 2006. Among these 19 recent publications, all discussed the sensitivity, specificity, or AUC of noninvasive markers for the diagnosis of advanced fibrosis, but none of these studies took into account the prevalence of fibrosis stages defining advanced or nonadvanced fibrosis.

The aim of this study was to assess the relationships between the prevalence of each fibrosis stage defining nonadvanced and advanced fibrosis and the AUCs in a large number of patients to better standardize the expression of fibrosis markers. It was anticipated that such an evaluation might allow more accurate comparisons to be made between fibrosis markers. The relationships between AUCs and biopsy length and biopsy fragmentation have been analyzed in a separate study and found to be less highly associated with variability in AUCs than the prevalence of each fibrosis stage defining nonadvanced and advanced fibrosis (14).

Materials and Methods

Standardization of AUCs According to the Prevalence of Fibrosis Stages

AUC as used in the evaluation of markers for the diagnosis of advanced fibrosis is not dependent on the prevalence of this condition. However, it is not known whether there is variability in the AUC due to a change in prevalence of each fibrosis stage used to define advanced fibrosis (i.e., F2, F3, F4) and nonadvanced fibrosis (i.e., F0, F1). Our aim was to assess the relationship between the prevalence of each stage and the AUC of a fibrosis marker for advanced fibrosis and to construct an index for standardizing comparisons. The standardization was performed to transform any different prevalence profile to 2 standard distributions of fibrosis stages: a uniform distribution and a naturally-observed distribution.

The uniform distribution of fibrosis stages was defined by a prevalence of 0.20 for each of the 5 stages (uniform prevalence). In the case of a uniform prevalence, the mean fibrosis stage in METAVIR units is 3 for advanced fibrosis: [mean of (F2 + F3 + F4)/3 = (2 + 3 + 4)/3 = 3] vs 0.5 for nonadvanced fibrosis [mean of (F1 + F0)/2 = (1 + 0)/2 = 0.5]. In this uniform prevalence distribution, the difference between the mean fibrosis stage of advanced fibrosis minus the mean fibrosis stage of nonadvanced fibrosis (DANA) was 2.5.

The naturally-observed distribution of fibrosis stages was defined by the prevalence of stages observed among patients with chronic hepatitis C, including those in a previous large study on the natural history of fibrosis progression (15). With liver biopsy estimates (n = 2235), the stage prevalences were F0 10%, F1 36%, F2 21%, F3 15%, and F4 18%. In the case of naturally-observed stage prevalences, the mean fibrosis stage in METAVIR units was 2.94 for advanced fibrosis: = [(0.21 × 2 + 0.15 × 3 + 0.18 × 4)/0.54 = 2.94] vs 0.78 for nonadvanced fibrosis: (0.36/0.46) = 0.78. In this naturally-observed prevalence distribution, the DANA was 2.16 (2.94 − 0.78 = 2.16).

The DANA was used for the standardization of fibrosis marker AUCs. The 1st step was to plot the AUCs of a validated fibrosis marker, FT, for the diagnosis of advanced fibrosis in a large database vs different combinations of fibrosis stages, resulting in a wide spectrum of DANA. The DANA range may vary from 1 to 4. A DANA of 1 would be obtained if advanced fibrosis is composed of F2 only, and nonadvanced fibrosis of F1 only (the mean of F2 = 2 minus the mean of F1 = 1, giving a DANA of 1). A DANA of 4 would be obtained if advanced fibrosis is composed of F4 only, and nonadvanced fibrosis of F0 only (the mean of F4 = 4 minus the mean F0 = 0, giving a DANA of 4). The regression between the AUCs and the whole DANA spectrum enables the AUC to be estimated from the DANA. The regression formula for standardizing AUCs estimated from different stage prevalences was AUC = 0.582 + 0.105 × (DANA).

The 2nd step was to check that the relationship observed in the integrated database with individual data was similar to that observed in an overview of the published diagnostic studies of FT. In the published studies, the DANA was not given by the authors but was estimated (eDANA) by the following formula: eDANA = mean advanced fibrosis as estimated by [(prevalence of F2 × 2 + prevalence F3 × 3 + prevalence F4 × 4)/ (prevalence F2 + prevalence F3 + prevalence F4)] minus the mean nonadvanced fibrosis as estimated by [prevalence F1/(prevalence F0 + prevalence F1)]. The equivalence of eDANA (from 19 different prevalence populations) with the direct DANA estimate using individual data (n = 1312) was checked using the integrated database with an almost perfect correlation [R Pearson = 0.999991 and Spearman = 1.0 (n = 19)].

We defined the uniform DANA at 2.5 (each stage prevalence = 0.20, giving a mean of 3 METAVIR fibrosis in the advanced fibrosis group and a mean of 0.5 in the nonadvanced group). An adjusted uniform AUC (AdnAUC), relating the observed DANA value to that seen when fibrosis stages are uniformly distributed (DANA = 2.5), can be calculated from the regression formula linking the observed AUC (ObAUC) to DANA.

We defined the naturally-observed DANA at 2.16. The AUC, which was standardized at the DANA value of 2.16, is the value for a given test for the diagnosis of advanced fibrosis when fibrosis stages are distributed as observed in the population. A naturally-observed adjusted AUC (AdnAUC), relating the observed DANA to that found in the population (DANA = 2.16), can be calculated from the regression formula linking the ObAUC to DANA.
PATIENTS

Integrated database. The integrated database included published prospective studies of chronic hepatitis C patients with concomitant FT measurement and liver biopsy, who had liver biopsies scored using the METAVIR scoring system (5), had FT assessed on fresh serum using the recommended preanalytical and analytical procedures, and had individual data sent by the principal investigator.

Overview of published diagnostic studies. Literature was reviewed using PubMed with the following keywords: diagnostic, fibrosis, and FT. Studies were included if fibrosis had been evaluated using the METAVIR scoring system, the diagnostic value was expressed using the AUC and the prevalence of different fibrosis stages was given.

Liver biopsies. In the integrated database, liver biopsies were processed using standard techniques. A pathologist who was unaware of the biochemical markers evaluated the fibrosis stage and necrosis grade according to the METAVIR scoring system (5). Fibrosis was staged on a scale of 0–4: F0 = no fibrosis, F1 = portal fibrosis without septa, F2 = few septa, F3 = numerous septa without cirrhosis, and F4 = cirrhosis. Biopsies were performed with a 16-gauge Hepafix Luer Lock needle (Braun Melsungen) in the Paris center and the Bordeaux center, and with various needles in the multicenter study from Marseille.

Biochemical markers. The previously validated FT was used (8). FT (FT, Biopredictive; HCV-Fibrosure, Labcorp) has been validated for the assessment of liver fibrosis in patients with chronic hepatitis C and B, and in patients with alcoholic and nonalcoholic steatosis. FT is a noninvasive blood test that combines the quantitative results of 5 serum biochemical markers [alpha2 macroglobulin, haptoglobin, gamma glutamyl transpeptidase, total bilirubin, and apolipoprotein A1] with the patient’s age and gender in a patented artificial intelligence algorithm (U.S. Patent 6 631 330) to generate a measurement of fibrosis stage in the liver.

STATISTICAL ANALYSIS

The AUC was used as a measurement of discrimination and estimated using the empirical (nonparametric) method of DeLong et al. (16). The paired method of Zhou et al. (17) was used for statistical comparisons of AUCs. All analyses were performed with NCSS software (NCSS).

RESULTS

INTEGRATED DATABASE

A total of 1312 patients from 3 centers (Paris n = 537, Marseille n = 601, Bordeaux n = 174) were included in the integrated database. These were consecutive patients with chronic hepatitis C who were PCR positive before treatment and who had had contemporaneous FT and liver biopsy with METAVIR staging performed. The prevalences of stages were 11% for F0 (n = 141), 40% for F1 (n = 520), 22% for F2 (n = 295), 16% for F3 (n = 208), and 11% for F4 (n = 148); there was a 50% prevalence of advanced fibrosis (n = 651). The mean age was 48 years and 58% were male.

OVERVIEW OF PUBLISHED STUDIES

A total of 19 populations have been studied for the diagnostic value of FT in HCV, with publications in 13 articles (Table 1) (9, 10, 18–28). For 3 studies outlined in 2 publications (26, 28), the prevalence of fibrosis stages was not detailed in the publication but was provided by the principal investigators on request. For one study the AUC was unknown (24) and was not included, which therefore resulted in 18 included studies.

AUCS ACCORDING TO THE PREVALENCE OF FIBROSIS STAGES

In the integrated database combinations and according to stage prevalence, the FT AUCs varied (P <0.001) from 0.67 to 0.98 for DANA ranging from 1.00 (the lowest difference between only stage F2 as advanced fibrosis and only F1 as nonadvanced fibrosis) to 4.00 (the highest difference between only F4 as advanced fibrosis and F0 as nonadvanced fibrosis) (Table 2). There was a highly significant correlation between the FT AUC and DANA [Spearman R coefficient = 0.95 (P <0.0001)] (Fig. 1). The regression formula for standardizing AUCs estimated from different stage prevalences was AUC = 0.582 + 0.1056 × (DANA). The FT AUC, which was standardized at the uniform DANA value of 2.5, was 0.85. Therefore the AduAUC, which took into account the observed DANA vs a uniform DANA of 2.5, was calculated using the formula: AduAUC = ObAUC + (0.1056) (2.5 − ObDANA). The FT AUC, which was standardized at the naturally-observed DANA value of 2.16, was 0.81. Therefore the AdnAUC, which took into account the observed DANA vs a DANA of 2.16 found when fibrosis stages were distributed as in the population at large, was calculated using the formula: AdnAUC = ObAUC + (0.1056) (2.16 − ObDANA).

In the overview, the FT AUC varied (P <0.001) from 0.65 to 0.89 for DANA ranging from 1.47 to 3.05. There was also a close correlation between the FT AUC and DANA [Spearman R coefficient = 0.64, (P = 0.004)]. The regression in the overview was not different from that of the integrated database, with almost all points being included in the 95% CI of the regression line (Fig. 1).

DISCUSSION

This study highlights the major impact of a previously neglected factor: the prevalence of fibrosis stages defining advanced fibrosis and nonadvanced fibrosis, which is highly associated with the AUC estimate. It has already been established that the accuracy of a test can vary with the definition of the target condition, but the size of this
Table 1. AUCs of FT for the diagnosis of advanced fibrosis in HCV patients, observed in published studies, according to prevalence of fibrosis stages defining advanced and non-advanced fibrosis, and DANA scores.

<table>
<thead>
<tr>
<th>First author</th>
<th>Number of Patients</th>
<th>Methodology</th>
<th>F0 (%)</th>
<th>F1 (%)</th>
<th>F2 (%)</th>
<th>F3 (%)</th>
<th>F4 (%)</th>
<th>DANA score</th>
<th>AUC SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varaut1, 2005</td>
<td>50</td>
<td>Retrospective Single center Dialysis patients</td>
<td>1</td>
<td>28</td>
<td>11</td>
<td>8</td>
<td>2</td>
<td>1.61</td>
<td>0.53</td>
</tr>
<tr>
<td>Varaut2, 2005</td>
<td>60</td>
<td>Retrospective Single center Kidney recipients</td>
<td>10</td>
<td>21</td>
<td>17</td>
<td>10</td>
<td>2</td>
<td>1.84</td>
<td>0.71</td>
</tr>
<tr>
<td>Sebastiani1, 2006</td>
<td>65</td>
<td>Prospective PNALT</td>
<td>18</td>
<td>22</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>2.21</td>
<td>0.71</td>
</tr>
<tr>
<td>Poynard-2, 2003</td>
<td>352</td>
<td>Retrospective Randomized trial Multicenter Before treatment</td>
<td>6</td>
<td>206</td>
<td>63</td>
<td>32</td>
<td>32</td>
<td>1.71</td>
<td>0.73</td>
</tr>
<tr>
<td>Poynard-1, 2001</td>
<td>299</td>
<td>Retrospective Randomized trial Multicenter Before treatment</td>
<td>60</td>
<td>133</td>
<td>56</td>
<td>55</td>
<td>5</td>
<td>1.42</td>
<td>0.74</td>
</tr>
<tr>
<td>Wilson, 2006 *</td>
<td>115</td>
<td>Retrospective Multicenter 30% HIV</td>
<td>39</td>
<td>36</td>
<td>33</td>
<td>8</td>
<td>3</td>
<td>1.87</td>
<td>0.74</td>
</tr>
<tr>
<td>Rossi, 2003</td>
<td>125</td>
<td>Prospective Multicenter Non-validated analyzers</td>
<td>25</td>
<td>52</td>
<td>26</td>
<td>13</td>
<td>9</td>
<td>1.95</td>
<td>0.74</td>
</tr>
<tr>
<td>Poynard-3, 2003</td>
<td>352</td>
<td>Retrospective Randomized trial Multicenter After treatment</td>
<td>15</td>
<td>222</td>
<td>64</td>
<td>25</td>
<td>26</td>
<td>1.73</td>
<td>0.77</td>
</tr>
<tr>
<td>Halfon, 2006</td>
<td>504</td>
<td>Prospective Multicenter</td>
<td>58</td>
<td>216</td>
<td>110</td>
<td>91</td>
<td>29</td>
<td>1.87</td>
<td>0.79</td>
</tr>
<tr>
<td>Sebastianii2, 2006</td>
<td>125</td>
<td>Prospective EALT</td>
<td>6</td>
<td>30</td>
<td>53</td>
<td>15</td>
<td>21</td>
<td>1.83</td>
<td>0.81</td>
</tr>
<tr>
<td>Cales1, 2005</td>
<td>478</td>
<td>Prospective Single center mixed causes Training cohort</td>
<td>28</td>
<td>170</td>
<td>120</td>
<td>57</td>
<td>102</td>
<td>2.08</td>
<td>0.82</td>
</tr>
<tr>
<td>Sene, 2006</td>
<td>138</td>
<td>Prospective Single center Cryoglobulinemia vasculitis</td>
<td>37</td>
<td>36</td>
<td>30</td>
<td>16</td>
<td>19</td>
<td>3.05</td>
<td>0.83</td>
</tr>
<tr>
<td>Castera, 2005</td>
<td>183</td>
<td>Prospective Single center</td>
<td>0</td>
<td>47</td>
<td>53</td>
<td>37</td>
<td>46</td>
<td>1.95</td>
<td>0.84</td>
</tr>
<tr>
<td>Imbert-1, 2001</td>
<td>189</td>
<td>Prospective Single center Training cohort</td>
<td>36</td>
<td>91</td>
<td>40</td>
<td>18</td>
<td>20</td>
<td>2.03</td>
<td>0.84</td>
</tr>
<tr>
<td>Cales2, 2005</td>
<td>120</td>
<td>Prospective Single center Validation cohort</td>
<td>20</td>
<td>43</td>
<td>28</td>
<td>14</td>
<td>15</td>
<td>2.09</td>
<td>0.86</td>
</tr>
<tr>
<td>Myers, 2003</td>
<td>130</td>
<td>Retrospective Single center HCV-HIV coinfection</td>
<td>17</td>
<td>55</td>
<td>22</td>
<td>19</td>
<td>17</td>
<td>2.15</td>
<td>0.86</td>
</tr>
<tr>
<td>Imbert-2, 2001</td>
<td>134</td>
<td>Prospective Single center Validation cohort</td>
<td>20</td>
<td>54</td>
<td>28</td>
<td>10</td>
<td>22</td>
<td>2.16</td>
<td>0.87</td>
</tr>
<tr>
<td>Callewaert 2004</td>
<td>106</td>
<td>Prospective HCV and alcohol</td>
<td>0</td>
<td>28</td>
<td>20</td>
<td>10</td>
<td>48</td>
<td>2.73</td>
<td>0.89</td>
</tr>
<tr>
<td>Coletta, 2005</td>
<td>40</td>
<td>Prospective Multicenter PNALT</td>
<td>3</td>
<td>23</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

* Details for stages given out of 119 and not out of 115: ** half of Knodell F3 has been counted as METAVIR F2. NA.: Not assessed; PNALT.: persistently normal ALT; EALT.: elevated ALT.
spectrum effect is usually unknown (29). The STARD statement has been developed for this purpose. Item 18 invites authors to report the “distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition” (30). In the present clinical setting the observed size of this spectrum effect was dramatic and we suggest that standardization methods be used. In practice, the simple calculation of the DANA enables the spectrum effect to be estimated.

A total of 18 published populations have been included (triangle) and 19 combinations of fibrosis stages (circle). The uniform standardized point is the intersection of the regression line (AUCs vs DANA) and the horizontal line, assuming a DANA of 2.5 (dashed line), which is the DANA of a population of equal fibrosis stage prevalence (prevalence F0 = F1 = F2 = F3 = F4 = 20%). The AduAUC for FT is 0.85. The naturally-observed standardized point is at the intersection of the regression line and the horizontal DANA line, assuming a DANA of 2.16 (dotted line), which is the DANA of a population of naturally-observed fibrosis prevalences (F0 = 10%, F1 = 36%, F2 = 21%, F3 = 15%, and F4 = 18%). The AdnAUC for FT is 0.81.

**Table 2. AUCs of FT for the diagnosis of advanced fibrosis observed in the integrated database, according to prevalence of fibrosis stages defining advanced and non-advanced fibrosis, means of fibrosis and DANA scores.**

<table>
<thead>
<tr>
<th>Fibrosis stage combinations defining non-advanced and advanced fibrosis</th>
<th>Number of patients</th>
<th>Mean fibrosis stage</th>
<th>DANA score*</th>
<th>AUC (SE) for the diagnosis of advanced fibrosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-advanced</td>
<td>Advanced</td>
<td>Non-advanced fibrosis</td>
<td>Advanced fibrosis</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
<td>815</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F0F1</td>
<td>F2 F3F4</td>
<td>956</td>
<td>0.79</td>
<td>2</td>
</tr>
<tr>
<td>F0</td>
<td>F2F3</td>
<td>1164</td>
<td>0.79</td>
<td>2.41</td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
<td>436</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>F0F1</td>
<td>F2F3F4</td>
<td>1171</td>
<td>0.79</td>
<td>2.77</td>
</tr>
<tr>
<td>F0F1</td>
<td>F2F4</td>
<td>1104</td>
<td>0.79</td>
<td>2.67</td>
</tr>
<tr>
<td>F0 F1</td>
<td>F2 F3 F4</td>
<td>1312</td>
<td>0.79</td>
<td>2.77</td>
</tr>
<tr>
<td>F0F1</td>
<td>F3</td>
<td>869</td>
<td>0.79</td>
<td>3</td>
</tr>
<tr>
<td>F0</td>
<td>F2 F3 F4</td>
<td>792</td>
<td>0</td>
<td>2.77</td>
</tr>
<tr>
<td>F0F1</td>
<td>F3F4</td>
<td>1017</td>
<td>0.79</td>
<td>3.42</td>
</tr>
<tr>
<td>F0</td>
<td>F4</td>
<td>668</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>F0F1</td>
<td>F4</td>
<td>809</td>
<td>0.79</td>
<td>4</td>
</tr>
<tr>
<td>F0</td>
<td>F4</td>
<td>289</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

* DANA score = Difference in Advanced minus Non-Advanced fibrosis means

**AUCS ACCORDING TO THE PREVALENCE OF FIBROSIS STAGES**

First, this study demonstrated that the prevalence of liver fibrosis stages defining advanced fibrosis is a major factor of variability in assessing the diagnostic value of a fibrosis marker. In the present example, FT AUCs varied from 0.67 to 0.98 according to this prevalence. FT was chosen as an example in this study, but this conclusion may be applied to any quantitative marker if the diagnosis of a disease includes several stages. For a disease defined as a combination of stages vs a combination of other stages, the AUCs for the diagnosis of this disease must be expressed in a standardized fashion according to the prevalence of a given stage. We suggest using either a uniform distribution with the same prevalence for each stage (corresponding to a difference in DANA of 2.5 fibrosis METAVIR units between advanced and non-advanced fibrosis), or a naturally-observed distribution of stages (corresponding to a DANA of 2.16 fibrosis METAVIR units). In the present integrated database, the ObAUC was 0.80, the standardized AduAUC was 0.85, and the AdnAUC was 0.82.

This finding is clinically significant, because it is easier to discuss the apparently discordant results of a given biomarker observed in the literature. For example, the last 2 published studies of FT in chronic hepatitis C found an ObAUC of 0.74 for Wilson et al. (27) and an AUC of 0.83 for Sene et al. (28). This difference can be fully explained by the difference in fibrosis stages with a DANA of 1.87 for the former and 3.05 for the latter, for a net DANA difference of 1.18. According to the correspondence between the AUC and DANA for FT (Fig. 1), a difference of 1 point in DANA is equivalent to a difference of 0.1056 in AUCs. Therefore the DANA difference of 1.18 is 1.18
of fibrosis. However, even if the exact model is unknown, it is important to note that the METAVIR scoring system is one of the best-validated scoring systems and is recommended in several guidelines.

Another limitation is that such prevalences are not seen in practice. Indeed, the prevalence of stage F0 in observational studies is low, at 10%, being half of the uniform prevalence of 20%. However, it must be recognized that the prevalences seen in observational studies are not perfect, because it is ethically impossible to perform liver biopsy in a large population of individuals without symptoms and with normal liver function tests. Most of the observational studies have been performed using biopsies in reference center populations. Therefore there is still a possibility that by using noninvasive biomarkers, a more realistic prevalence of stage F0 could be closer to 20% (the neutral hypothesis with uniform fibrosis stage prevalence) than 10% (the naturally-observed prevalence, which could underestimate the F0 stage prevalence). Indeed, by using biomarkers in 33,731 patients with presumed chronic hepatitis C, the prevalence of stage F0 was 29% (9881 of 33,731) (35).

We used a more realistic standardization that mirrors the distribution across stages found in studies with population-based sampling and liver biopsy. One can question why standardization is needed at all. The ideal design for an accuracy study is the inclusion of a consecutive series of patients. In the present study, the 3 populations included were consecutive patients in each center, thus reflecting the naturally-observed prevalence of fibrosis stages. This finding is in accordance with the similar distribution of fibrosis stages observed in the present study and in our previous study of fibrosis progression (14). In previously published overviews and editorials comparing the accuracy of fibrosis biomarkers, the authors compared and discussed the AUC without any reference to the spectrum effect. The proposed standardizations can help the decision-maker estimate the real AUC of the biomarker corresponding to their own setting. Diagnostic studies comparing different markers need direct comparisons in the same patients to avoid bias and variability related to indirect comparisons, including spectrum bias. The number of new fibrosis markers is increasing rapidly and it will be increasingly difficult to compare all of them in the same patients (4). AUC standardization cannot replace direct comparisons between markers, but enables misinterpretation in indirect comparisons to be reduced.

LIMITATIONS OF THE STUDY

This study has several limitations:

Our statistical analysis used a resampling from the same database, which could reinforce unknown bias.

We used a simple standardization for the uniform DANA, giving the same weight to each fibrosis stage. First there is a controversy concerning the linear association between the METAVIR scoring system and the quantity of fibrosis, and also concerning the linear progression of fibrosis. However, even if the exact model is unknown,
prevalence of fibrosis stages defining advanced or nonadvanced fibrosis.

Apart from the diagnosis of liver fibrosis, these findings suggest that the evaluation of biomarker utility could be better standardized. “Grouping” of disease stages to control for differences in staging between populations should be used. The standardization of prevalences could be made according to the prevalence of stages in observational studies if well established, or according to a more neutral hypothesis if these prevalences are not available.

Grant/funding support: Thierry Poynard has grants from the Association pour la Recherche sur les Maladies Virales and from the Association de Recherche sur les Maladies Hépatiques.

Financial disclosures: Thierry Poynard has a potential conflict of interest as the inventor of FibroTest with a capital interest in Biopredictive, the company marketing the FibroTest. Mona Munteanu is a full-time employee of Biopredictive. The company marketing the FibroTest.

Acknowledgments: Special thanks to Pierre Bedossa, Patrice Cacoub, Alfredo Alberti, and Giada Sebastiani for furnishing details of their previous publications.

References


