ActiT est accuracy for the assessment of histological activity grades in patients with chronic hepatitis C, an overview using Obuchowski measure

T. Poynard\textsuperscript{a,}\textsuperscript{*}, M. Munteanu\textsuperscript{b}, Y. Ngo\textsuperscript{b}, L. Castera\textsuperscript{c}, P. Halfon\textsuperscript{d}, V. Ratziu\textsuperscript{a}, F. Imbert-Bismut\textsuperscript{a}, D. Thabut\textsuperscript{a}, M. Bourliere\textsuperscript{e}, P. Cacoub\textsuperscript{f}, D. Messous\textsuperscript{a}, V. de Ledinghen\textsuperscript{c}

\textsuperscript{a} UPMC Paris Liver Center, AP–HP, Paris, France  
\textsuperscript{b} Department of Biopredictive, Paris, France  
\textsuperscript{c} Service d’hépatogastroenterologie, hôpital Haut-Leveque, Pessac, France  
\textsuperscript{d} Laboratoire Alphabio, Marseille, France  
\textsuperscript{e} Service d’hépatogastroenterologie, hôpital Saint-Joseph, Marseille, France  
\textsuperscript{f} Service de médecine interne, groupe hospitalier Pitié-Salpêtrière, université Paris VI, Paris, France

Available online 2 July 2010

Summary

Background. — ActiT est (AT) is a biomarker of liver necro-inflammatory histological activity validated in patients with chronic hepatitis C (HCV).

Aim. — The aim was to assess the accuracy of AT in comparison with alanine aminotransferase (ALT) the standard of care.

Methods. — Methods used an integrated database of individual data and the new recommended Obuchowski measures. An updated “classical” meta-analysis of AT validation studies was also performed. The main end points were the area under the ROC curves (AUROCs) for the diagnosis of each histological activity grade defined using METAVIR scoring system. To avoid repeated tests and the spectrum effect of activity grades prevalence, the comparison of AT and ALT accuracies used the Obuchowski method.

Results. — For the individual analysis, a total of 1250 patients were included and for the meta-analysis six studies (2017 patients) were included. The overall accuracy of AT for the diagnosis of any activity grade (Obuchowski measure = 0.850) was significantly higher than the accuracy of ALT (Obuchowski measure = 0.837; \(P = 0.009\)). The updated standard meta-analysis confirmed the accuracy of AT (\(P < 0.0001\)) both in independent AUROC = 0.79 (95% CI, 0.73–0.85) and in non independent studies AUROC = 0.74 (95% CI, 0.67–0.81).

Conclusions. — The accuracy of AT for grading the necro-inflammatory activity of patients with HCV was significantly higher than ALT serum activity alone, the standard biomarker.

© 2010 Elsevier Masson SAS. All rights reserved.

\textsuperscript{*} Corresponding author.

E-mail address: tpoynard@teaser.fr (T. Poynard).

0399-8320/$ - see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.gcb.2010.05.001
Background

ActiTTest (AT) is a biomarker of liver necro-inflammatory histological activity initially validated in patients with chronic hepatitis C (HCV) [1, 2] and then in patients with chronic hepatitis B (HBV) [3]. AT is widely used associated with Fibrotest (FT) as a non invasive alternative to liver biopsy, which had several limitations [4—6]. AT combined alanine amino transferase serum activity (ALT) and the five components of FT (alpha2 macroglobulin, apolipoprotein A1, haptoglobin, gamma-glutamyl transpeptidase, and total bilirubin). Several studies and meta-analyses have been performed for FT validating its accuracy for the diagnosis of fibrosis [7—10], but only one meta-analysis for AT [7] and none concerning other activity biomarkers. Furthermore, the methodology of assessing the accuracy of liver injury biomarkers has been recently highly improved [8—14]. The Obuchowski method [15] better takes into account the spectrum effect [12,13], the ordinal scale of scoring system [13], and the multiple testing [13] than the standard recommended statistical methods [16,17]. The Obuchowski method has not been yet applied for activity biomarkers.

In patients with HCV, the non invasive assessment of necro-inflammatory grade, combined with fibrosis staging, is however still important for at least three reasons. The first one is for the indication of treatment, which could be discussed in patient without advanced fibrosis but with moderate or severe necro-inflammatory activity [18,19], and for maintenance therapy in patients with cirrhosis in order to reduce necrosis and inflammation [20,21]. The second is the risk of false positive induced by necro-inflammatory activity on fibrosis biomarkers including transaminases in their panels [22,23] and on liver stiffness measurements by fibroscan [14,24]. The third is the risk of over interpretation of fibrosis improvement after antiviral treatment if the impact on necro-inflammatory activity is not taken into account [23].

The aim of the present study was to assess the accuracy of AT compared to ALT, the standard of care, using an integrated database of individual data and the new recommended Obuchowski measures. An updated “classical” meta-analysis of AT validation studies was also performed.

Methods

This study was conducted according to the principles expressed in the declaration of Helsinki. Two meta-analyses were performed; one used an integrated database combining individual data provided by authors, and the other combined all identified published studies.

Design of individual data meta-analysis

Patients

Patients included in the database with individual data were already described in a previous overview of fibrotest in patients with chronic liver diseases [8]. For the present study, only patients with HCV were included. Patients’ data were included with the agreement of authors in the present study if they met the following optimal conditions of analyses; they must have been prospectively included in a validation study including AT and biopsy with METAVIR scoring system for activity score, measured on fresh serum, according to preanalytical and analytical recommendations, with a liver biopsy performed less than 7 days apart of the serum samples. All patients have a HCV with positive PCR. All biomarkers have been assessed blindly to clinical, biochemical or histological patients’ characteristics.

Biochemical analysis

Biochemical assays were performed with fresh serum decanted and stored for a maximum of 72 hours at +2—8° C, under no-light conditions. The dosage methods for each of the FT-AT parameters followed the standardized conditions, as listed in the technical recommendations (http://www.biopredictive.com). For ALT, the method used pyridoxal phosphate and the calibration approved by the IFCC [25,26].

Liver biopsies

Liver biopsies were processed using standard techniques. A pathologist who was unaware of the biomarkers evaluated the fibrosis stage and necrosis grade according to the METAVIR scoring system [27]. 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. Necro-inflammatory activity was graded on a scale of 0—3: A0: no activity; A1: minimal activity; A2: moderate; and A3: severe. 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.

Statistical analysis

The use of the AUROC raises two methodological issues. First, its use is based on the assumption that the gold standard is binary, whereas activity grading uses an ordinal scale. This difference implies that activity grades in the study sample have to be aggregated into two groups, a process that can lead to discordant conclusions, depending on how the groups are aggregated, as demonstrated for fibrosis stages [11—13]. Analysis based on the AUROC can also be biased by the way in which the proportion of each stage of fibrosis in the sample fits the distribution in the reference population to which the indices are applied. As a result, the comparison of different AUROCs based on samples with different stage distributions may be flawed (spectrum effect) [11—13]. The spectrum effect reflects the inherent variation in test performance among population subgroups. Subgroup variation is not a bias but is clinically relevant information to be identified and reported with appropriate analyses. If a spectrum effect is possible, heterogeneity should be assessed by subgroup analyses of test performance.

To overcome these methodological issues, two methods have been recently proposed and validated for fibrosis staging. We proposed a standardization of the AUROC for the distribution of fibrosis stages to deal with the spectrum effect [11,12]. Lambert et al. proposed in order to overcome both spectrum effect and ordinal scale, to use the Obuchowski measure [15]. This measure can be used in situa-
tions in which the gold standard is not binary. Furthermore, the Obuchowski measure allows to compare two biomarkers with a single test, avoiding appropriate correction for the type I error when comparing two biomarkers for different stages or grades [13,15].

**Obuchowski measure**

This measure is a multinomial version of the AUROC. With $N$ (= 4) categories of the gold standard outcome (histological activity grade) and AUROCst, the estimate of the AUROC of diagnostic tests for differentiating between categories s and t, the Obuchowski measure, is a weighted average of the $N(N - 1)/2$ (= 6) different AUROCst corresponding to all the pairwise comparisons between two of the $N$ categories.

Each pairwise comparison has been weighted to take into account the distance between activity grades (i.e., the number of units on the ordinal scale). A penalty function proportional to the difference in METAVIR units between grades was defined: the penalty function was 0.33 when the difference between stages was 1, 0.67 when the difference was 2, and 1 when the difference was 3. The Obuchowski measure can be interpreted as the probability that the non-invasive index will correctly rank two randomly chosen patient samples from different activity grades according to the weighting scheme, with a penalty for misclassifying patients [13,15]. Note that the overall Obuchowski measure is not equivalent to a usual area under the ROC curve as the measurements are weighted according to the distance between stages. The rationale was to reduce the risk of bias due to retrospective analysis favoring AT versus ALT in the weighting process. For grading activity, the AT cutoffs were those recommended for manufacturer since the first validation using biopsy: 0.29, 0.52 and 0.62 for A1, A2 and A3 respectively [1,7,11]. For ALT, the a priori simple cutoffs of ALT chosen were 50, 100 and 150 IU/L as we previously demonstrated than the expression of ALT activity using the upper limit of the normal was hazardous [28].

The Obuchowski measure can be interpreted as the probability that the non-invasive index will correctly rank two randomly chosen patient samples from different activity grades according to the weighting scheme, with a penalty for misclassifying patients [13,15].

**Main endpoint**

The main end point was the accuracy estimated with Obuchowski measure. The AT accuracy was compared with ALT, the standard marker of necro-inflammatory histological activity.

**Sensitivity analyses**

Sensitivity analyses were performed by comparing Obuchowski measures between AT and ALT (the standard diagnostic test) according to gender, age, biopsy length, fibrosis stage, and independency of authors (inventor of the AT or not). The studies performed in the Paris center included the FT-AT inventor group and studies performed in Marseille and Bordeaux were independent of the inventor group.

**Design of meta-analysis of published studies**

To select published studies, we used the standards for reporting of diagnostic accuracy (STARD) criteria and the Cochrane database of systematic reviews (CDSR) methods. Key STARD criteria include factors such as whether:

- the study population was relevant to the clinical question being addressed;

---

**Figure 1** Flowsheet of patients included in the ActiT test and transaminases ALT overview.
• there was a careful description of the population from which the patients were drawn, as well as actual inclusions and exclusions;
• recruitment and the mode of sampling were carefully described;
• researchers interpreting the non-invasive test were blinded to the reference test result;
• sufficient data were provided to complete a 2 × 2 table of true and false positive and negative diagnoses.

Studies published only with an abstract provided insufficient data and were excluded.

Search strategy
We searched MEDLINE with the key word “ActiT est”. We hand-searched key journals (Gastroenterology, Hepatology, Journal of Hepatology, Gut, Journal of Viral Hepatitis and American Journal of Gastroenterology) from February 2001 to December 2009 to validate the search, as well as the abstract books of the American Association and European Association for the Study of Liver Disease annual meetings.

Inclusion and exclusion criteria
Two reviewers (a hepatologist and a hepatologist-statistician) independently assessed the papers with predetermined STARD criteria. Disagreements were resolved through discussion with a third reviewer. We included studies in patients with HCV, with AT and liver biopsy, which provided data for true positives and negatives, false positives and negatives and AUROCs for advanced activity, which stated that the AT had been assessed blind to the biopsy and which stated the method used for defining the degree of activity. We were careful to avoid including data from duplicate publications. Inclusion of study was never dependent on the result of the non-invasive test under investigation.

Statistical methods used in AUROCs meta-analysis
The AUROC was estimated by the empirical (non-parametric) method of Delong et al., equivalent to the Mann–Whitney statistic and compared using the paired method of Zhou et al. [29,30]. The meta-analysis was stratified according to the independency of authors. The analysis used a random effect model and the heterogeneity between effects according to centers has been tested using Cochran’s Q heterogeneity test (Q). Analyses were performed on NCSS software (Kaysville, Utah, USA) [31] and on R software.

Results
For the individual analysis, a total of 1250 patients were included ( years 2002–2007) and for the meta-analysis six studies (2017 patients) were included (2002–2009) (Fig. 1). A total of 827 patients of the individual database were also included in the meta-analysis of studies, that is 41% (827/2017) of this population.

Table 1 Characteristics of patients included in the individual data base.

<table>
<thead>
<tr>
<th>Characteristic (number of patients)</th>
<th>1,250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>729 (58; 56—61)</td>
</tr>
<tr>
<td>Female</td>
<td>521 (42; 39—44)</td>
</tr>
<tr>
<td>Age year (median)</td>
<td>46 (45—47)</td>
</tr>
<tr>
<td>Center</td>
<td></td>
</tr>
<tr>
<td>Paris</td>
<td>490 (39; 36—42)</td>
</tr>
<tr>
<td>Marseille</td>
<td>600 (48; 46—51)</td>
</tr>
<tr>
<td>Bordeaux</td>
<td>160 (13; 11—15)</td>
</tr>
<tr>
<td>Independency</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>760 (61; 58—64)</td>
</tr>
<tr>
<td>No</td>
<td>490 (39; 36—42)</td>
</tr>
<tr>
<td>Activity grade</td>
<td></td>
</tr>
<tr>
<td>A0</td>
<td>136 (11; 9—13)</td>
</tr>
<tr>
<td>A1</td>
<td>610 (49; 46—52)</td>
</tr>
<tr>
<td>A2</td>
<td>446 (36; 33—38)</td>
</tr>
<tr>
<td>A3</td>
<td>58 (5; 4—6)</td>
</tr>
<tr>
<td>Fibrosis stage</td>
<td></td>
</tr>
<tr>
<td>F0</td>
<td>138 (11; 9—13)</td>
</tr>
<tr>
<td>F1</td>
<td>493 (39; 37—42)</td>
</tr>
<tr>
<td>F2</td>
<td>279 (22; 20—24)</td>
</tr>
<tr>
<td>F3</td>
<td>193 (15; 13—18)</td>
</tr>
<tr>
<td>F4</td>
<td>147 (12; 10—14)</td>
</tr>
<tr>
<td>Length biopsy mm (median)</td>
<td>16 (15—18)</td>
</tr>
<tr>
<td>ALT IU/L (median)</td>
<td>65 (62—68)</td>
</tr>
<tr>
<td>ActiT est (median)</td>
<td>0.43 (0.40—0.46)</td>
</tr>
<tr>
<td>Fibrotest (median)</td>
<td>0.37 (0.35—0.40)</td>
</tr>
</tbody>
</table>

Individual data analysis
Patients included
A total of 1250 patients were included for assessing AT accuracy (Table 1). The main characteristics of these patients were similar to the larger database previously published for FT meta-analysis (data not showed) [10,11]. Median age was 46 years, 58% of male, 85% of biopsy with minimal or moderate activity, and 12% of cirrhosis.

Accuracy of ActiT est and alanine aminotransferase
The AUROCs of AT and ALT are detailed for each pair wise grades comparison in Table 2. The overall mean (SE) accuracy of AT (Obuchowski measure) was 0.848 (0.005) greater than that of ALT, 0.834 (0.006), P = 0.008. The greater difference was observed for the accuracy between moderate and severe grade, 0.651 (0.033) and 0.564 (0.045) respectively.

Sensitivity analyses
Sensitivity analyses according to patients’ characteristics are detailed in Table 3. The same trend was observed in all subgroups, with accuracy of AT always greater than ALT, except for the subgroup of patients with advanced fibrosis.
Table 2  Accuracy (area under the ROC curves) of ActiT est and alanine aminotransferase for pair wise grades diagnosis in 1250 patients of the core group.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>A1 (n = 610)</th>
<th>A2 (n = 446)</th>
<th>A3 (n = 53)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0 (n = 136)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ActiT est</td>
<td>0.589 (0.537—0.641)</td>
<td>0.744 (0.698—0.746)</td>
<td>0.889 (0.833—0.941)</td>
</tr>
<tr>
<td>ALT</td>
<td>0.580 (0.530—0.593)</td>
<td>0.718 (0.670—0.766)</td>
<td>0.849 (0.793—0.905)</td>
</tr>
<tr>
<td>A1 (n = 610)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ActiT est</td>
<td>0.524 (0.494—0.554)</td>
<td>0.770 (0.702—0.838)</td>
<td></td>
</tr>
<tr>
<td>ALT</td>
<td>0.501 (0.467—0.535)</td>
<td>0.676 (0.596—0.756)</td>
<td></td>
</tr>
<tr>
<td>A2 (n = 446)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ActiT est</td>
<td></td>
<td>0.651 (0.585—0.717)</td>
<td></td>
</tr>
<tr>
<td>ALT</td>
<td></td>
<td>0.564 (0.474—0.654)</td>
<td></td>
</tr>
</tbody>
</table>

The overall mean with standard error (SE) accuracy of ActiT est (Obuchowski measure) was 0.848 (0.838—0.858) greater than that of ALT, 0.834 (0.822—0.846),  P = 0.008. Note that the overall Obuchowski measure is not equivalent to a usual area under the ROC curve as weighted according to the distance between stages.

Table 3  Sensitivity analyses of ActiT est accuracy according to patient characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Obuchowski measure mean a standard error (SE) ActiT est</th>
<th>Alanine aminotransferase mean a standard error (SE)</th>
<th>Significance P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core group (1,250)</td>
<td>0.848 (0.838—0.858)</td>
<td>0.834 (0.006)</td>
<td>0.008</td>
</tr>
<tr>
<td>Gender</td>
<td>0.842 (0.828—0.856)</td>
<td>0.832 (0.818—0.846)</td>
<td>0.19</td>
</tr>
<tr>
<td>Male (729)</td>
<td>0.850 (0.836—0.864)</td>
<td>0.833 (0.817—0.849)</td>
<td>0.02</td>
</tr>
<tr>
<td>Female (521)</td>
<td>0.855 (0.839—0.871)</td>
<td>0.839 (0.810—0.857)</td>
<td>0.08</td>
</tr>
<tr>
<td>Age</td>
<td>0.844 (0.830—0.858)</td>
<td>0.831 (0.817—0.845)</td>
<td>0.06</td>
</tr>
<tr>
<td>&gt; 50 year (482)</td>
<td>0.850 (0.838—0.862)</td>
<td>0.832 (0.818—0.846)</td>
<td>0.01</td>
</tr>
<tr>
<td>≤ 50 year (768)</td>
<td>0.842 (0.826—0.874)</td>
<td>0.847 (0.829—0.865)</td>
<td>0.20</td>
</tr>
<tr>
<td>Independency</td>
<td>0.840 (0.007)</td>
<td>0.827 (0.007)</td>
<td>0.02</td>
</tr>
<tr>
<td>Yes (760)</td>
<td>0.858 (0.842—0.874)</td>
<td>0.834 (0.009)</td>
<td>0.65</td>
</tr>
<tr>
<td>No (490)</td>
<td>0.829 (0.008)</td>
<td>0.827 (0.007)</td>
<td>0.04</td>
</tr>
<tr>
<td>Fibrosis stage</td>
<td>0.829 (0.008)</td>
<td>0.827 (0.007)</td>
<td>0.04</td>
</tr>
<tr>
<td>F0F1 (763)</td>
<td>0.840 (0.007)</td>
<td>0.829 (0.008)</td>
<td>0.04</td>
</tr>
<tr>
<td>F2F3F4 (487)</td>
<td>0.827 (0.007)</td>
<td>0.827 (0.007)</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Note that the overall Obuchowski measure is not equivalent to a usual area under the ROC curve.

Meta-analysis of published studies

A total of eight populations (2017 patients) analyzed in six articles were included in the meta-analysis (Table 4) [1,2,31—35]. Four out of the six articles were published by independent authors [32—35]. The mean of the observed AUROCs in published studies was 0.77 (95% CI, 0.72—0.81), significantly greater than the 0.50 non accuracy value. There was no significant difference between the AUROCs observed in non independent studies (0.79 [95% CI, 0.73—0.85]) compared to independent studies (0.74 [95% CI, 0.67—0.81]) (Fig. 2). A significant heterogeneity was observed between all studies (Q = 23.8; P = 0.01) and also among non-independent (Q = 9.9; P = 0.03) and among independent studies (Q = 9.6; P = 0.02).

Discussion

In the last 7 years, numerous biomarkers of liver fibrosis have been published, but few biomarkers of necro-inflammatory histological activity have been published [7,10,11,36,37]. To our knowledge, only ALT and AT has been validated as an alternative to biopsy for estimating the grade of activity in patients with HCV [7,36]. The present meta-
<table>
<thead>
<tr>
<th>First author [ref]</th>
<th>Number of patients</th>
<th>Methodology</th>
<th>Age</th>
<th>Area under ROC (95% CI)</th>
<th>A0</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>Biopsy length mm</th>
<th>Independent</th>
<th>Guidelines and fresh</th>
<th>Biopsy fibrotest median days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imbert-Bismut, 2001 [1]</td>
<td>205</td>
<td>Prospective Single center Training cohort</td>
<td>47</td>
<td>0.79 (0.73—0.85)</td>
<td>52</td>
<td>85</td>
<td>63</td>
<td>5</td>
<td>16</td>
<td>No</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>Imbert-Bismut, 2001 [1]</td>
<td>134</td>
<td>Prospective Single center Validation cohort</td>
<td>48</td>
<td>0.75 (0.69—0.81)</td>
<td>17</td>
<td>80</td>
<td>33</td>
<td>4</td>
<td>16</td>
<td>No</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>Poynard et al., 2003 [2]</td>
<td>352</td>
<td>Retrospective Randomized trial Multicenter Before treatment</td>
<td>45</td>
<td>0.75 (0.69—0.81)</td>
<td>5</td>
<td>55</td>
<td>117</td>
<td>17</td>
<td>17</td>
<td>No</td>
<td>No</td>
<td>137</td>
</tr>
<tr>
<td>Poynard et al., 2003 [2]</td>
<td>352</td>
<td>Retrospective Randomized trial Multicenter After treatment</td>
<td>47</td>
<td>0.86 (0.80—0.92)</td>
<td>83</td>
<td>136</td>
<td>82</td>
<td>51</td>
<td>17</td>
<td>No</td>
<td>No</td>
<td>12</td>
</tr>
<tr>
<td>Halfon et al., 2006 [32]</td>
<td>504</td>
<td>Prospective Multicenter</td>
<td>45</td>
<td>0.73 (0.69—0.77)</td>
<td>41</td>
<td>265</td>
<td>183</td>
<td>15</td>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>Morali et al., 2007 [33]</td>
<td>81</td>
<td>Prospective Multicenter</td>
<td>40</td>
<td>0.79 (0.73—0.85)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>Yes</td>
<td>&lt; 30</td>
<td></td>
</tr>
<tr>
<td>Fontanges et al., 2008 [34]</td>
<td>96</td>
<td>Prospective Multicenter</td>
<td>48</td>
<td>0.64 (0.58—0.70)</td>
<td>NA</td>
<td>66</td>
<td>27</td>
<td>3</td>
<td>23</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
</tr>
<tr>
<td>Anastasiou et al., 2009a [35]</td>
<td>65</td>
<td>Prospective Single center</td>
<td>50</td>
<td>0.83 (0.75—0.91)</td>
<td>28</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
</tbody>
</table>

NA not available.
a Twenty-seven patients had an hepatitis C.
b No details between A0 and A1.
of bias. All biopsy analyses were performed blindly to all

inventor’s group with a possible conflict of interest. How-

length.

AL T in the different subgroups stratified according to biopsy

25 mm. However, the accuracy of AT was always higher than

meta-analysis, no study has a median length greater than

length was 16 mm, with only 25% greater than 25 mm and in

the three populations including common patients (data not

showed).

analyses confirmed, by two methods, the accuracy of AT

for the diagnosis of necro-inflammatory histological activity

in patients with HCV. In comparison with the previous

overview [7], the present study included independent

studies, and demonstrated a higher accuracy of AT versus

ALT, the classical reference biomarker of activity.

This study has several limitations

First, 41% of patients analyzed in the meta-analysis were

also included in the individual database. Therefore, the
two analyses were not strictly independent. However, the
meta-analysis results were the same after exclusion of
the three populations including common patients (data not
showed).

Second, the standard reference, the liver biopsy has lim-

ited quality. The recommended length for biopsy length is

at least 25 mm [4]. In the individual database, the median

length was 16 mm, with only 25% greater than 25 mm and in

the meta-analysis, no study has a median length greater than

25 mm. However, the accuracy of AT was always higher than

ALT in the different subgroups stratified according to biopsy

length.

Third, the present overview was performed by the AT

inventor’s group with a possible conflict of interest. How-
ever, the methodology used prevented many possible risks

of bias. All biopsy analyses were performed blindly to all

patients’ characteristics including biomarkers results. Fur-

thermore, when independent data were analyzed separately

using both methods, there was still a significant accuracy

for AT, significantly greater than that of ALT (Table 3).
A significant effect was observed by the classical meta-

analysis (Fig. 2). However, this heterogeneity was not clearly

attributable to a center effect as this heterogeneity per-

sisted among the Parisian center. A spectrum effect is

possible but not established in the present analysis.

Fourth, it is not usual to compare a variable (ALT) with
a score (AT) including the same variable. By construction,
the diagnostic accuracies of AT and ALT would differ only by
the influence of other components of the AT score. There-
fore, the comparisons between ALT and AT would logically be
favorable to AT or at worst, not significantly different. One
main hypothesis to explain the better performance of AT,
as observed for acute alcoholic hepatitis, is the predictive
value of low apolipoprotein A1 for the diagnosis of moderate
or severe necrosis [37].

This study has several advantages

First, the individual database permitted to verify that the
AT accuracy was similar in almost all subgroup of patients,
which is reassuring for the clinical applicability. Interest-
ingly, there was a significant AT accuracy, better than ALT
alone in the subgroup of patients without advanced fibrosis
(Table 3), which could be an indication of treatment in case
of moderate or severe activity.

Second, the individual database permitted to assess the
accuracy of AT and ALT for all the activity grades diagno-
sis and not only for the diagnosis of advanced activity. As
already demonstrated for FT [8,11,12], AT AUROCs were around 0.60 for one stage/grade difference,
0.75 for two stages/grades difference and 0.90 for three
stages/grades difference (Table 2).

Thirdly, the use of Obuchowski method for the analysis
of individual data permitted to avoid the spectrum effect
and the risk of multiple testing when comparing the accu-

ricies of AT and ALT for the various activity grades [13].
This is a very important improvement in comparison to the
"classical" meta-analysis with aggregated classes, which
is highly impacted by spectrum effect [11–13] The present
standard meta-analysis illustrated the risk of spectrum bias
with very a heterogeneous spectrum of grades. The preva-
ience of grade 0 varied from 1 to 43% and grade 3 from 0

The mean difference between AUROCs and the

curves (AUROC) assessed in published studies of ActiTest diag-

stasis and not only for the diagnosis of advanced activity.
As

already demonstrated for FT the so-called "gray zone" or
"inaccurate zone" (lower AUROCs between intermediate
stages or grades) is mostly a mathematical consequence of
the spectrum effect for liver injury biomarkers or biopsy

5,11,32]. However, many reviews or guidelines have not

underlined this point [16–19,36]. As for FT [8,11,12], AT

AUROCs were around 0.60 for one stage/grade difference,
0.75 for two stages/grades difference and 0.90 for three
stages/grades difference (Table 2).

For a non statistician reader, the inter-

pretation of Obuchowski measure is not simple. The overall

measure is not equivalent to the usual area under the ROC
curve, as the Obuchowski measure is an accuracy weighted

measure is not equivalent to the usual area under the ROC
curve, as the Obuchowski measure is an accuracy weighted

obtained if more sensitive
cutoffs were used for ALT: 20, 40 and 60IU/L (data not showed).

Fourth, AT (including ALT) measurements were performed by validated laboratories, which agreed to respect the recommendations of preanalytical and analytical processes which reduced the ALT variability [25,26].

Finally for clinicians, the three main advantages of AT versus ALT are the following: an overall significantly higher accuracy, a complementary performance when assessed together with Fibrotest assessing activity grade together with fibrosis stage, and the use of validated analyzers and calibrated kits for this specific indication. Furthermore, using transaminases results expressed in upper limits of the normal is hazardous [25,26,28].

Conclusions
The individual database analysis and the updated meta-analysis confirmed the accuracy of AT for the grading of necro-inflammatory histological activity in patients with HCV. The accuracy estimated by Obuchowski measure was significantly higher than the ALT accuracy, the standard of care.

Competing interests
TP is a consultant and has a capital interest in Biopredictive, the company marketing FT, and MM and YN are full time employees of Biopredictive.

Biopredictive had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Authors’ contributions
TP conceived the study, performed the statistical analysis, and wrote the manuscript. PH, LC, VR, FIB, DT, DM, MB, and VdL participated in the coordination of the study, data monitoring and drafted the manuscript. MM, YN were involved in the data collection, checking the applicability of AT according to manufacturer rules. All authors read and approved the final manuscript.

References


