Ultrasonographic assessment of liver fibrosis with computer-assisted analysis of liver surface irregularities

N. Huet a,*, I. Denis b, A. Martino a, B. Gallix c, N. Sturm d, V. Leroy e, I. Bricault a, f

a Service de radiologie et imagerie médicale, CHU de Grenoble, Cs 10217, boulevard de la Chantourne, 38043 Grenoble cedex 9, France
b Service de radiologie générale, Hôpital de Chambéry, 7, square Massalaz, 73000 Chambéry, France
c Service d’imagerie médicale, Hôpital Saint-Éloi, 80, avenue Augustin-Fliche, 34295 Montpellier cedex 5, France
d Service d’anatomo-pathologie, CHU de Grenoble, Cs 10217, 38043 Grenoble cedex 9, France
e Service d’hépato-gastro-entérologie, CHU de Grenoble, Cs 10217, 38043 Grenoble cedex 9, France
f Université Grenoble Alpes, TIMC-IMAG, 38000 Grenoble, France

Abstract

Purpose: The goal of this study was to evaluate the diagnostic accuracy of a software program that automatically analyzes the liver surface to diagnose significant fibrosis, by comparing it to the subjective analysis of a radiologist and to transient elastography (Fibroscan®).

Patients and methods: One hundred fourteen patients with chronic liver disease were included in the study. They underwent liver biopsy, FibroScan® and ultrasonographic examination of the liver surface. The liver surface was analyzed by a software program that gave a score of surface irregularities. This evaluation was compared to subjective analysis by a radiologist expert in liver imaging and by two general radiologists.

Results: Fifty percent of the patients had significant fibrosis according to the METAVIR score. The AUROC for the diagnosis of significant fibrosis by the software program was 0.80 (95%CI: 0.71–0.87), which was equivalent (P = 0.86) to that of FibroScan® (0.81; 95%CI: 0.71–0.89).

Keywords: Ultrasound; Liver fibrosis; Computer-assisted image analysis

Abbreviations: AUROC, Area under the receiver operating characteristic (ROC) curve; NASH, Non alcoholic steatohepatitis; ROC, Receiver operating characteristic; HBV, Hepatitis B virus; HCV, Hepatitis C virus; HIV, Human immunodeficiency virus; NPV, Negative predictive value; PPV, Positive predictive value.

* Corresponding author.
E-mail address: nhuet@chu-grenoble.fr (N. Huet).

http://dx.doi.org/10.1016/j.diii.2015.02.010
2211-5684/© 2015 Éditions françaises de radiologie. Published by Elsevier Masson SAS. All rights reserved.
Introduction

The development of fibrosis during follow-up of chronic liver disease, whatever the etiology, is a major prognostic factor. Precise analysis of the stage of fibrosis is essential to determine future management. Although liver biopsy is still considered to be the gold standard, it still has numerous limitations, in particular due to sampling errors. The rate of false negative findings for the diagnosis of cirrhosis can reach up to 20 to 30% [1,2]. Moreover, biopsy is invasive, expensive [3,4], and conveys a risk of severe complications (0.4%) and even death (0.03%) [5–7].

As a result, multiple non-invasive diagnostic tests have been developed to evaluate the liver as precisely as possible, to identify patients with significant lesions that could respond to treatment and to monitor disease progression. The most common complementary tools are blood tests, which consist of a combination of biological markers [8–11] and transient elastography (FibroScan®) to measure liver stiffness [12–18]. The morphological study of the liver by conventional ultrasonography (US) is also frequently used in daily clinical practice to confirm the diagnosis of cirrhosis [4,19,20]. Among the different criteria used, liver surface analysis using high resolution US helps diagnose liver fibrosis and differentiate between the early stages of fibrosis [21–23].

The goal of this study was to evaluate the diagnostic accuracy of a software program that automatically analyzes the liver surface to diagnose significant fibrosis, by comparing it to the subjective analysis of a radiologist and to transient elastography.

Materials and methods

Patients

One hundred fourteen patients (56 women, 58 men; mean age 51 ± 10.6 years old, range 18–82 years), presenting with unexplained persistent elevation of transaminases or chronic liver disease, were included in this prospective study. One hundred and five patients were hospitalized in the hepatogastroenterology unit of the Grenoble University Hospital to undergo liver biopsy and the remaining nine had clinically confirmed cirrhosis. US-guided percutaneous liver biopsy was performed by two hepatologists. Histological analysis of the liver biopsies was performed by an experienced pathologist. The degree of liver fibrosis was assessed using the 5 stages of the META VIR scoring system, from F0 to F4 (absent, minimal, moderate, severe and confirmed cirrhosis). The study was consistent with the Declaration of Helsinki; data were collected as part of a biological collection authorized by the Institutional Review board and registered under the reference DC-2008-727.

Imaging protocol

One hundred four patients had liver stiffness measurement by two trained operators using the FibroScan® (Echosens, Paris, France) US analysis of the liver surface was performed by a radiologist using an Aplio® (Toshiba Medical Systems, Tokyo, Japan) US unit and a 7.2 MHz probe within between one and four hours after performing the liver biopsy or during a hospital stay in patients with cirrhosis. US cine loops were recorded for each patient using a right lateral intercostal approach. It should be noted that the choice was made to only analyze the right liver surface. Indeed, the correlation with the results of liver biopsy (also performed in the right liver) appeared more pertinent. Moreover, data from the expert radiologist, in addition to preliminary tests performed with the software, showed that the quality of the US recordings of the left liver surface were often insufficient for reliable analysis of the irregularity because of the deeper location and higher slant of the left liver capsule.

Image analysis

An extract of each video recording was selected by the operator by positioning a region of interest outside of areas of poor echogenicity, areas with artifacts due to shadowing posterior to the ribs or areas where the capsule did not appear perpendicular to the US beam. The video extracts were read by a radiologist expert specialized in liver imaging and two general radiologists. Interpretations were performed blinded to clinical data and histopathological results. A score of 0 to 4 was used: 0 for a smooth, thin and regular hyperechogenic surface, 1 for a nearly normal surface with some irregularities, 2 intermediate, 3 for an irregular spotty surface, and 4 for a totally irregular and deformed liver surface.

These video extracts were then analyzed using software developed at the Grenoble University Hospital allowing automated analysis of the contours of the liver surface by US. This software quantitatively evaluated the irregularity of the liver surface using a score between 0 and 100% (Fig. 1).

Statistical analysis

All data were recorded and analyzed using NCSS and MedCalc software. The reference test was the histopathological
analysis of liver tissue samples. The results of the software analysis, of FibroScan® and of the radiologist interpretations for the diagnosis of significant fibrosis (F0-1 vs. F2-4), severe fibrosis (F0-2 vs. F3-4) and cirrhosis (F0-3 vs. F4) were evaluated using area under the ROC curve (AUROC) analysis. A P value less than 0.05 was considered statistically significant. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated after determining optimal threshold values for each method.

Correlations between liver surface irregularity scores (using software) and results of histopathological analysis were evaluated by a non-parametric Spearman’s correlation coefficient (\( \rho \)). The kappa test was used to evaluate interobserver agreement of radiologists for the diagnosis of significant fibrosis.

**Results**

The distribution of patients in relation to the stage of fibrosis and etiology is shown in Table 1. Fifty percent of the patients had significant fibrosis \( F \geq 2 \). The quality of 110/114 US videos (96.5%) was considered to be good enough for automatic software analysis. 4 other ultrasound videos were automatically rejected as failures by the software, because the slant of the capsule was detected as too high in relation to the axis of the ultrasound beam.

The scores of surface irregularity according to the software were between 11% and 89% (Fig. 2). They were significantly correlated to the degree of fibrosis (\( \rho = 0.508; P < 0.001 \)). The results of the software for the diagnosis of significant fibrosis evaluated by the AUROC were 0.80 (95%CI: 0.71–0.87). A threshold of 25% resulted in a sensitivity of 89% and a specificity of 61%. With a score below 25%, observed in 37% of the population, significant fibrosis was excluded with a NPV of 85.4%.

The AUROC for the diagnosis of significant fibrosis by the expert radiologist was 0.66 (95%CI: 0.56–0.75). Diagnosis by the expert radiologist appeared to be significantly poorer (\( P = 0.02 \)) than by the software (Fig. 3). If a subjective
radiological score of ≥ 2 was chosen to indicate significant fibrosis, the diagnosis would be made with a sensitivity of 66% and a specificity of 54%. Moreover, the interobserver agreement among the radiologists was poor, with kappa coefficients ranging between 0.25 and 0.37.

The AUROC for the diagnosis of significant fibrosis by FibroScan® was 0.81 (95%CI: 0.71–0.89), with no significant difference with that obtained using the software (P = 0.86).

Comparison of AUROCs obtained using software, FibroScan® and subjective analysis of the expert radiologist for the different stages of fibrosis is presented in Table 2.

Discussion

The results of our study show that automatic liver surface analysis is better than subjective analysis by radiologists. Our results show that the diagnosis of significant fibrosis by subjective radiological interpretation is not satisfactory. Moreover, as might be expected with the subjective evaluation, agreement among radiologists is poor. The results obtained with the software are similar to those of FibroScan® for all stages of fibrosis and are better than those reached by the expert radiologist (Table 2). The diagnostic accuracy of the software was significantly better than those of the expert for F ≥ 2 fibrosis (P = 0.02). Indeed, although there were numerous areas of overlap between the scores of liver irregularity by the software and the METAVIR score (Fig. 2), the discriminatory value of the software appears to be especially effective between the two groups corresponding to significant fibrosis (F2-F3-F4) and minimal/absent fibrosis (F0-F1).

Thus, the overall results of the software for the evaluation of significant fibrosis are good with an AUROC of 0.80. These results are basically similar to those of other conventional tests in the literature with an AUROC of 0.84 for FibroScan® [13] and between 0.78 and 0.86 for the main biological tests [9,10]. FibroScan® and the biological tests are alternative tests that are increasingly used and some of them have been validated in the diagnosis of HCV-related fibrosis or cirrhosis [16,19].

The software would be extremely useful if it were integrated into an US equipment to be used in routine clinical practice. The analysis could be performed during morphological liver US examination, which is part of the initial work up for all patients with chronic liver disease, or even more widely for screening during US for any indication. However, at a threshold of 25%, the software would not have detected significant fibrosis present in two patients. According to the three readers, the capsule in these recordings appeared thin and regular with no breaks or marked deformity. These features therefore result in a false negative finding of liver surface analysis rather than of the software. The study by Colli et al. reported false negatives in 26% of F3-F4 patients without any explanation for this [22]. Nevertheless, the parenchyma in our two false negatives was multinodular, a feature that is not studied by the software but that could allow the radiologist to reclassify these patients.

The potential causes of fibrosis stage overestimation were searched for in twenty F0-F1 patients in our study. Significant fragmentation of the biopsy samples could have influenced the level of disagreement because they were found in 50% of the overestimated F0-F1 patients compared to 13.4% of the patients for whom there was agreement. The fragmentation of the biopsy could result in a false negative biopsy, rather than a false positive US analysis. On the other hand, the presence of biopsy specimens that were too small with too few portal spaces did not seem to be a factor of disagreement. Finally, steatosis could play a role because the degree of fibrosis was overestimated in 3/4 of patients with NASH in our study [23,24]. The degree of fibrosis was also overestimated in these patients by most of our radiologists and in all patients by the expert reader. Perisinusoidal fibrosis without portal or periportal involvement, which is therefore not taken into account in the METAVIR score, is often associated with steatohepatitis [24] and could play a role in the development of capsular irregularities, especially since it was more present in the F0-F1 overestimated patients than in those for whom there was agreement.

The number of false positive and false negative findings with this diagnostic method are therefore similar to those obtained with other available tests. However, it has the advantage of having a different performance profile. Indeed, there is usually a peak rate of disagreement in F2 patients, which is due to the difficulty of classifying this group by usual non-invasive methods. In our study on the other hand, 82.6% of the F2 patients were correctly classified.

Four failures were automatically identified by the software because of the high slant between the liver capsule and the US beam axis. The 110 other US recordings were accepted and analyzed by the software. The liver surface is usually located superficially and is therefore easy to evaluate with a 7.2 MHz probe. Unlike FibroScan®, we can assume that ascites does not influence the success of this technique because it does not limit the effectiveness of the US beam and does not prevent satisfactory visualization of the capsule. On the other hand, when the liver is located deeper

Table 1 Distribution of patients in terms of chronic liver disease and the stage of fibrosis according to the METAVIR score.

<table>
<thead>
<tr>
<th>METAVIR score</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV</td>
<td>0</td>
<td>17</td>
<td>19</td>
<td>8</td>
<td>8</td>
<td>52</td>
</tr>
<tr>
<td>HBV</td>
<td>6</td>
<td>15</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Autoimmune hepatitis</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>NASH</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Hemochromatosis</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Sclerosing cholangitis</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Alcohol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>42</td>
<td>25</td>
<td>14</td>
<td>18</td>
<td>114</td>
</tr>
</tbody>
</table>

Miscellaneous contains abnormal liver test results, scleroderma, portal thrombosis.
Table 2: Areas under the ROC curves of FibroScan®, the software and the expert radiologist for the diagnosis of fibrosis in relation to the METAVIR stage.

<table>
<thead>
<tr>
<th></th>
<th>Significant fibrosis (F0-1 vs. F2-3-4)</th>
<th>Severe fibrosis (F0-1-2 vs. F3-4)</th>
<th>Cirrhosis (F0-1-2-3 vs F4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC 95% CI</td>
<td>AUROC 95% CI</td>
<td>AUROC 95% CI</td>
</tr>
<tr>
<td>FibroScan®</td>
<td>0.81 0.71–0.89</td>
<td>0.80 0.70–0.88</td>
<td>0.84 0.75–0.91</td>
</tr>
<tr>
<td>Automatic software analysis</td>
<td>0.80 0.71–0.87</td>
<td>0.76 0.67–0.85</td>
<td>0.81 0.72–0.88</td>
</tr>
<tr>
<td>Expert reader</td>
<td>0.66 0.56–0.75</td>
<td>0.70 0.61–0.78</td>
<td>0.70 0.61–0.79</td>
</tr>
</tbody>
</table>

The differences in AUROC observed between FibroScan® and the automatic software analysis were not significant. The AUROC of the software was higher than those of the expert; the threshold of significance was reached for the diagnosis of significant fibrosis (P = 0.02).

below the skin (i.e., more than 2 cm), ultrasound visualization of the liver surface may be less precise and it is possible that the evaluation of surface regularity will be less accurate. This hypothesis should be confirmed by a subgroup evaluation in a further study with a sufficiently large proportion of over weighted patients.

Our study has several limitations. The evaluation of this technique, like all non-invasive diagnostic techniques of liver fibrosis, was limited by the absence of a strong standard of reference. It is acknowledged that a biopsy fragment of 25 mm in length, which corresponds to the reference size, is associated with a rate of disagreement of 25% for the degree of fibrosis because of possible heterogeneous distribution of fibrosis in the liver [25].

The interobserver variability of this technique was not studied because it was not possible for each patient to undergo two US examinations by two different radiologists on the day of the biopsy. To evaluate intra-operator variability the different patients would have needed to return to a consultation after a short period, which was not possible for organizational reasons. There is however a factor of potential variability that seems important to analyze: the extent of the liver surface that was evaluated varied from patient to patient depending on the US scanning performed, and the irregularities of the liver capsule could be different depending on the parts of the liver that were explored. Thus, these preliminary results must be confirmed by further studies in which the influence of different US parameters, the analyzed area and the operator can be identified. If the software was included directly in the US machine, testing could be performed more simply by different radiologists on the same patient during an US examination or several days before or after the biopsy.

Moreover, it is impossible to draw firm conclusions from this preliminary study on the value of this liver surface analysis technique compared to existing tests, and on the role it could play. Additional studies are needed to further define the value of this approach and to compare it with other tests to define the best diagnostic thresholds and possible test combinations.

Conclusion

This computer-assisted analysis of the liver surface showed a better accuracy than subjective analysis by radiologists, with similar performance to FibroScan® for the diagnosis of significant fibrosis. The reproducibility of this technique as well as its value in association with other non-invasive diagnostic tests should be evaluated in larger studies.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

The authors would like to thank Doctors L. Felix and M. Michoud for their participation in this study.

References


