EASL-ALEH Clinical Practice Guidelines:
Non-invasive tests for evaluation of liver disease severity and prognosis
Clinical Practice Guidelines

EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis

European Association for the Study of the Liver*, Asociación Latinoamericana para el Estudio del Hígado

Introduction

Liver fibrosis is part of the structural and functional alterations in most chronic liver diseases. It is one of the main prognostic factors for the amount of hepatic fibrosis in the whole liver due to viral and non-viral chronic liver diseases [1,2]. Liver biopsy has traditionally been considered the reference method for evaluation of tissue damage such as hepatic fibrosis in patients with chronic liver disease. Pathologists have proposed robust scoring system for staging liver fibrosis such as the semi-quantitative META VIR score [3,4]. In addition computer-aided morphometric measurement of collagen proportional area, a partly automated technique, provides an accurate and linear evaluation of the amount of fibrosis [5]. Liver biopsy gives a snapshot and not an insight into the dynamic changes during the process of fibrogenesis (progression, static or regression). However, immunohistochemical evaluation of cellular markers such as smooth muscle actin expression for hepatic stellate cell activation, cytokeratin 7 for labeling ductular proliferation or CD34 for visualization of sinusoidal endothelial capillarization or the use of two-photon and second harmonic generation fluorescence microscopy techniques for spatial assessment of fibrillar collagen, can provide additional “functional” information [6,7]. All these approaches are valid provided that the biopsy is of sufficient size to represent the whole liver [4,8]. Indeed, liver biopsy provides only a very small part of the whole organ and there is a risk that this part might not be representative for the amount of hepatic fibrosis in the whole liver due to heterogeneity in its distribution [9]. Extensive literature has shown that increasing the length of liver biopsy decreases the risk of sampling error. Except for cirrhosis, for which micro-fragments may be sufficient, a 25 mm long biopsy is considered an optimal specimen for accurate evaluation, though 15 mm is considered sufficient in most studies [10]. Not only the length but also the caliber of the biopsy needle is important in order to obtain a piece of liver of adequate size for histological evaluation, with a 16 gauge needle being considered as the most appropriate [11] to use for percutaneous liver biopsy. Interobserver variation is another potential limitation of liver biopsy which is related to the discordance between pathologists in biopsy interpretation, although it seems to be less pronounced when biopsy assessment is done by specialized liver pathologists [12]. Beside technical problems, liver biopsy remains a costly and invasive procedure that requires physicians and pathologists to be sufficiently trained in order to obtain adequate and representative results – this again limits the use of liver biopsy for mass screening. Last but not least, liver biopsy is an invasive procedure, carrying a risk of rare but potentially life-threatening complications [13,14]. These limitations have led to the development of non-invasive methods for assessment of liver fibrosis. Although some of these methods are now commonly used in patients for first line assessment, biopsy remains within the armamentarium of hepatologists when assessing the etiology of complex diseases or when there are discordances between clinical symptoms and the extent of fibrosis assessed by non-invasive approaches.

Methodological considerations when using non-invasive tests

The performance of a non-invasive diagnostic method is evaluated by calculation of the area under the receiver operator characteristic curve (AUROC), taking liver biopsy as the reference standard. However, biopsy analysis is an imperfect reference standard: taking into account a range of accuracies of the biopsy, even in the best possible scenario, an AUROC >0.90 cannot be achieved for a perfect marker of liver disease [15]. The AUROC can vary based on the prevalence of each stage of fibrosis, described as spectrum bias [16]. Spectrum bias has important implications for the study of non-invasive methods, particularly in comparison of methods across different study populations. If extreme stages of fibrosis (F0 and F4) are over-represented in a population, the sensitivity and specificity of a diagnostic method will be higher than in a population of patients that has predominantly middle stages of fibrosis (F1 and F2). Several ways of preventing the “spectrum bias” have been proposed including the adjustment of AUROC using the DANA method (standardization according to the prevalence of fibrosis stages that define advanced (F2–F4) and non-advanced (F0–F1) fibrosis) [17,18] or the Obuchowski measure (designed for ordinal gold standards) [19]. What really matters in clinical practice is the number of patients correctly classified by non-invasive methods for a defined endpoint according to the reference standard (i.e. true positive and true negative).
Clinical Practice Guidelines

General statements

- Even though liver biopsy has been used as the reference method for the design, evaluation and validation of non-invasive tests, it is an imperfect gold standard. In order to optimize the value of liver biopsy for fibrosis evaluation, it is important to adhere to the following recommendations: (i) sample length >15 mm by a 16G needle; (ii) use of appropriate scoring systems according to liver disease etiology; and (iii) reading by an experienced (and if possible specialized) pathologist.
- Non-invasive tests reduce but do not abolish the need for liver biopsy; they should be used as an integrated system with liver biopsy according to the context.

Methodology

These Clinical Practice Guidelines (CPGs) have been developed by a panel of experts chosen by the EASL and ALEH Governing Boards. The recommendations were peer-reviewed by external expert reviewers and approved by EASL and ALEH Governing Boards. The CPGs were established using data collected from PubMed and Cochrane database searches. The CPGs have been based, as far as possible, on evidence from existing publications, and, if evidence was unavailable, the experts’ provide personal experiences and opinion. When possible, the level of evidence and recommendation are cited. The evidence and recommendations in these guidelines have been graded according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system. The strength of recommendations thus reflects the quality of underlying evidence. The principles of the GRADE system have been enunciated [20]. The quality of the evidence in the CPG has been classified into one of three levels: high (A), moderate (B) or low (C). The GRADE system offers two grades of recommendation: strong (1) or weak (2) (Table 1). The CPGs thus consider the quality of evidence: the higher the quality of evidence, the more likely a stronger recommendation is warranted; the greater the variability in values and preferences, or the greater the uncertainty, the more likely a weaker recommendation is warranted.

The non-invasive tests CPG Panel has considered the following questions:

What are the currently available non-invasive tests?

How do serum biomarkers perform for staging liver fibrosis?

What is the added value of combining TE and serum biomarkers?

What are the indications for non-invasive tests for staging liver disease in viral hepatitis?

What are the indications for non-invasive tests for staging liver disease in non-alcoholic fatty liver disease (NAFLD)?

What is the added value of combining TE and serum biomarkers?

What are the indications for non-invasive tests for staging liver disease in chronic liver diseases?

What is the prognostic value of non-invasive tests in chronic liver disease?

Table 1. Evidence grading used for the EASL-ALEH guidelines (adapted from the GRADE system).

<table>
<thead>
<tr>
<th>Evidence quality</th>
<th>Notes</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Further research is very unlikely to change our confidence in the estimate of effect</td>
<td>A</td>
</tr>
<tr>
<td>Moderate</td>
<td>Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate</td>
<td>B</td>
</tr>
<tr>
<td>Low</td>
<td>Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Any change of estimate is uncertain</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Notes</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Factors influencing the strength of the recommendation included the quality of the evidence, presumed patient-important outcomes, and cost</td>
<td>1</td>
</tr>
<tr>
<td>Weak</td>
<td>Variability in preferences and values, or more uncertainty. Recommendation is made with less certainty, higher cost or resource consumption</td>
<td>2</td>
</tr>
</tbody>
</table>

Guidelines

Currently available non-invasive methods

Non-invasive methods rely on two different approaches: a “biological” approach based on the quantification of biomarkers in serum samples or a “physical” approach based on the measurement of liver stiffness (LS). Although these approaches are complementary, they are based on different rationales. Serum biomarkers indicate several, not strictly liver specific clinical and serum parameters that have been associated with fibrosis stage, as assessed by liver biopsy, whereas LS corresponds to a genuine and intrinsic physical property of liver parenchyma.

Serum biomarkers of liver fibrosis

Many serum biomarkers have been proposed for staging liver fibrosis, mainly in patients with chronic hepatitis C. They are...
Table 2. Currently available serum biomarkers for non-invasive evaluation of liver fibrosis in chronic liver disease.

| HCV | FibroScan® (Biopredictive, Paris, France) patented formula combining α-2-macroglobulin, γGTP, apolipoprotein A1, haptoglobin, total bilirubin, age and gender |
| Summary: The practical advantages of analyzing serum biomarkers to evaluate liver fibrosis are that they are widely available (non-patented) (Table 3). However, none are liver specific and their results may be influenced by changes in clearance and excretion of each individual parameters. For instance, increased levels of hyaluronate occur in the post-prandial state [45] or in aged patients with chronic inflammatory processes such as rheumatoid arthritis [46]. Also, the reproducibility of measurement of some parameters included in "indirect" serum markers, such as aspartate aminotransferase (AST) levels or platelet count, is questionable [47]. In addition, the interpretation of each test requires a critical analysis in order to avoid false positive or false negative results. For instance, when using FibroTest®, the existence of hemolysis or Gilbert syndrome that can lead to false positive results in bilirubin, respectively should be taken into account [48]. Similarly, acute hepatitis can produce false positive results in tests, since all include serum levels of amino transferases in their formulas.

Liver stiffness measurement

Transient elastography

Liver fibrosis can be staged using 1-dimensional ultrasound TE (FibroScan(R), Echosens, Paris, France) [49], which measures the...
Clinical Practice Guidelines

Table 3. Respective advantages and disadvantages of currently available non-invasive methods in patients with chronic liver disease.

<table>
<thead>
<tr>
<th>Serum biomarkers</th>
<th>Transient elastography</th>
<th>ARFI (pSWE)</th>
<th>Measurement of liver stiffness</th>
<th>MR elastography</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Good reproducibility</td>
<td>• Most widely used and validated technique: standard to be beaten</td>
<td>• Can be implemented on a regular US machine</td>
<td>• Can be implemented on a regular US machine</td>
<td>• Can be implemented on a regular MRI machine</td>
</tr>
<tr>
<td>• High applicability (95%)</td>
<td>• User-friendly (performed at bedside; rapid, easy to learn)</td>
<td>• ROI smaller than TE but location chosen by the operator</td>
<td>• ROI can be adjusted in size and location and chosen by the operator</td>
<td>• Examination of the whole liver</td>
</tr>
<tr>
<td>• No cost and wide availability (non-patented)</td>
<td>• High range of values (2-75 kPa)</td>
<td>• Higher applicability than TE (ascites and obesity)</td>
<td>• Measures liver stiffness in real-time</td>
<td>• Higher applicability than TE (ascites and obesity)</td>
</tr>
<tr>
<td>• Well validated</td>
<td>• Quality criteria well defined</td>
<td>• Performance equivalent to that of TE for significant fibrosis and cirrhosis</td>
<td>• High range of values (2-150 kPa)</td>
<td>• High performance for cirrhosis</td>
</tr>
<tr>
<td>• Can be performed in the outpatient clinic</td>
<td>• Good reproducibility</td>
<td>• Prognostic value in cirrhosis</td>
<td>• Good applicability</td>
<td>• Prognostic value in cirrhosis</td>
</tr>
</tbody>
</table>

| **Disadvantages** |                        |             |                               |                 |
| • Non-specific of the liver | • Requires a dedicated device | • Unable to discriminate between intermediate stages of fibrosis | • Further validation warranted especially in comparison with TE | • Further validation warranted especially in comparison with TE |
| • Unable to discriminate between intermediate stages of fibrosis | • ROI cannot be chosen | • Performance not as good as TE for cirrhosis | • Unable to discriminate between intermediate stages of fibrosis | • Not applicable in case of iron overload |
| • Performance not as good as TE for cirrhosis | • Unable to discriminate between intermediate stages of fibrosis | • Cost and limited availability (proprietary) | • Unavailable for comparison with TE | • Requires a MRI facility |
| • Cost and limited availability (proprietary) | • Applicability (80%) lower than serum biomarker: (obesity, ascites, operator experience) | • Limitations (hemolysis, Gilbert syndrome, inflammation…) | • Learning curve? | • Time-consuming |
| • Limitations (hemolysis, Gilbert syndrome, inflammation…) | • False positive in case of acute hepatitis, extra-hepatic cholestasis, liver congestion, food intake and excessive alcohol intake | • Quality criteria not well defined | • Influence of inflammation? | • Costly |
| • False positive in case of acute hepatitis, extra-hepatic cholestasis, liver congestion, food intake and excessive alcohol intake | • Unable to discriminate between intermediate stages of fibrosis | • Quality criteria not well defined | • Prognostic value in cirrhosis? | • Prognostic value in cirrhosis |

ROI, region of interest.

velocity of a low-frequency (50 Hz) elastic shear wave propagating through the liver. This velocity is directly related to tissue stiffness, called the elastic modulus (expressed as $E = 3 \rho v^2$, where $v$ is the shear velocity and $\rho$ is the density of tissue, assumed to be constant). The stiffer the tissue, the faster the shear wave propagates.

TE is performed on a patient lying supine, with the right arm elevated to facilitate access to the right liver lobe. The tip of the probe is contacted to the intercostal skin with coupling gel in the 9th to 11th intercostal space at the level where a liver biopsy would be performed. The operator, assisted by a time-motion image, locates a liver portion at least 6 cm deep and free of large vascular structures. The operator then presses the probe button to start the measurements (“shots”). TE measures LS in a volume that approximates a cylinder 1 cm wide and 4 cm long, between 25 mm and 65 mm below the skin surface. The software determines whether each measurement is successful or not. When a shot is unsuccessful, the machine does not return a value. The entire procedure is considered to have failed when no value is obtained after ten shots. The final result of a TE session can be regarded as valid if the following criteria are fulfilled: 1) a number of valid shots of at least 10; 2) a success rate (the ratio of valid shots to the total number of shots) above 60%; and 3) an interquartile range (IQR, reflecting the variability of measurements) less than 30% of the median LS measurements (M) value ($\text{IQR/M} \leq 0.30\%$) [50].

The results are expressed in kilopascals (kPa), and range from 1.5 to 75 kPa with normal values around 5 kPa, higher in men and in patients with low or high body mass index (BMI) (U-shaped distribution) [51–54].

Advantages of TE include a short procedure time (<5 min), immediate results, and the ability to perform the test at the bedside or in an outpatient clinic (Table 3). Finally, it is not a difficult procedure to learn which can be performed by a nurse or a technician after minimal training (about 100 examinations) [55]. Nevertheless, the clinical interpretation of TE results should always be in the hands of an expert clinician and should be made with full knowledge of patient demographics, disease etiology and essential laboratory parameters.

Although TE analysis has excellent inter- and intra-observer agreement [56,57] (with an intra-class correlation coefficient (ICC) of 0.98), its applicability is not as good as that of serum biomarkers. In the largest TE series reported to date ($n = 13,369$ examinations), failure to obtain any measurement has been reported in 3.1% of cases and unreliable results (not meeting manufacturer’s recommendations) in 15.8% [58], mostly due to patient obesity or limited operator experience. Similar results have been reported in a large series of Asian patients ($n = 3205$).
with failure and unreliable results rates of 2.7% and 11.6%, respectively [59].

An important question in clinical practice is whether unreliable results translate into decreased accuracy. It has been suggested that among the recommendations, the IQR/M <30% is the most important parameter for good diagnostic accuracy [60,61]. In a recent study [62] in 1165 patients with chronic liver diseases (798 with chronic hepatitis C) taking liver biopsy as reference, TE reliability was related to two variables in multivariate analysis: the IQR/M and LS measure. Indeed, the presence of an IQR/M >30% and LS measure median >7.1 kPa resulted in a lower accuracy (as determined by AUROC) than that of the whole study population and these cases were therefore considered “poorly reliable”. Conversely, the highest accuracy was observed in the group with an IQR/M <10% regardless of the LS measure. Also a recent study reported a significant discrepancy in up to 20% of cases in patients without cirrhosis between different FibroScan devices (402 vs. 502) [63]. These results require further validation before any recommendation can be made.

In order to minimize the number of patients with unreliable results due to obesity, a new probe (XL, 2.5 MHz transducer), allowing measurement of LS between 35 to 75 mm depth, has been developed [64–68]. Myers et al. [66] showed that in 276 patients with chronic liver disease (42% viral hepatitis, 46% NAFLD) and a BMI >28 kg/m², measurement failures were significantly less frequent with the XL probe than with the M probe (1.1% vs. 16%; p <0.00005). However, unreliable results were still observed with the XL probe in 25% of cases instead of 50% with the M probe (p <0.00005). Also it is important to note that stiffness values obtained with XL probe are lower than that obtained with the M probe (by a median of 1.4 kPa).

Apart from obese patients, TE results can also be difficult to obtain from patients with narrow intercostal space and are nearly impossible to obtain from patients with ascites [49]. As the liver is an organ with a distensible but non-elastic envelope (Glisson’s capsule), additional space-occupying tissue abnormalities, such as edema, inflammation, extra-hepatic cholestasis, or congestion, can interfere with measurements of LS, independently of fibrosis. Indeed, the risk of overestimating LS values has been reported with other confounding factors including alanine aminotransferase (ALT) flares [69–71], extra-hepatic cholestasis [72], congestive heart failure [73], excessive alcohol intake [74–76], and food intake [77–80], suggesting that TE should be performed in fasting patients (for at least 2 h) and results always interpreted being aware of these potential confounding [81]. The influence of steatosis is still a matter of debate with conflicting results: some studies suggest that steatosis is associated to an increase in LS [82–84] whereas others do not [85,86].

Other liver elasticity-based imaging techniques
Several other liver elasticity-based imaging techniques are being developed, including ultrasound-based techniques and 3-D magnetic resonance (MR) elastography [87]. Ultrasound elastography can be currently performed by different techniques, which are based on two physical principles: strain displacement/imaging and shear wave imaging and quantification [88]. The latter allows a better estimation of liver tissue elasticity/stiffness, and includes point shear wave elastography (pSWE), also known as acoustic radiation force impulse imaging (ARFI) (Virtual touch tissue quantification™, Siemens; elastography point quantification, ElastPQ™, Philips) and 2D-shear wave elastography (2D-SWE) (Aixplorer™ Supersonic Imagine, France). pSWE/ARFI involves mechanical excitation of tissue using short-duration (~262 μsec) acoustic pulses that propagate shear waves and generate localized, μ-scale displacements in tissue [89]. The shear wave velocity (expressed in m/sec) is measured in a smaller region than in TE (10 mm long and 6 mm wide), but the exact location where measurements are obtained can be selected by the operator under B-mode visualization. A major advantage of pSWE/ARFI is that it can be easily implemented on modified commercial ultrasound machines (Acuson 2000/3000 Virtual Touch™ Tissue Quantification, Siemens Healthcare, Erlangen, Germany; ElastPQ™, IU22xMATRIX, Philips, Amsterdam, The Netherlands). Its failure rate is significantly lower than that of TE (2.9% vs. 6.4%, p <0.001), especially in patients with ascites or obesity [90]. Also its reproducibility is good, with ICC ranging from 0.84 to 0.87 [91–93]. However, like TE, pSWE/ARFI results are influenced by food intake [94] as well as necro-inflammatory activity and the serum levels of aminotransferases [95], both of which lead to an overestimation of liver fibrosis and have to be taken into account when interpreting the results. LS values obtained with pSWE/ARFI, in contrast to TE values, have a narrow range (0.5–4.4 m/sec). This limits the definitions of cut-off values for discriminating certain fibrosis stages and thus for making management decisions. Finally, quality criteria for correct interpretation of pSWE results remain to be defined.

2D-SWE is based on the combination of a radiation force induced in tissues by focused ultrasonic beams and a very high frame rate ultrasound imaging sequence capable of catching in real time the transient propagation of resulting shear waves [96]. The size of the region of interest can be chosen by the operator. 2D-SWE has also the advantage of being implemented on a commercially ultrasound machine (Aixplorer™. Supersonic Imagine, Aix en Provence, France) with results expressed either in m/sec or in kPa at a wide range of values (2–150 kPa). Its failure rate is significantly lower than that of TE [97–99], particularly in patients with ascites [98,99], but not in obese patients when the XL probe is used for TE (10.4% vs. 2.6%, respectively) [100]. Similar to pSWE/ARFI, quality criteria for 2D-SWE remain to be defined.

MR elastography uses a modified phase-contrast method to image the propagation characteristics of the shear wave in the liver [101]. Elasticity is quantified by MR elastography (expressed in kPa) using a formula that determines the shear modulus, which is equivalent to one-third the Young’s modulus used with TE [102]. The theoretical advantages of MR elastography include its ability to analyze almost the entire liver and its good applicability in patients with obesity or ascites. However, MR elastography remains currently too costly and time-consuming to be used in routine practice and cannot be performed in livers of patients with iron overload, because of signal-to-noise limitations.
Clinical Practice Guidelines

Recommendations

- Non-invasive tests should always be interpreted by specialists in liver disease, according to the clinical context, considering the results of other tests (biochemical, radiological and endoscopic) and taking into account the recommended quality criteria for each test and its possible pitfalls (A1)

- Serum biomarkers can be used in clinical practice due to their high applicability (>95%) and good inter-laboratory reproducibility. However, they should be preferably obtained in fasting patients (particularly those including hyaluronic acid) and following the manufacturer’s recommendations for the patented tests (A1)

- TE is a fast, simple, safe and easy to learn procedure that is widely available. Its main limitation is the impossibility of obtaining results in case of ascites or morbid obesity and its limited applicability in case of obesity and limited operator experience (A1)

- TE should be performed by an experienced operator (>100 examinations) following a standardized protocol with the patient, fasting for at least 2 hours, in the supine position, right arm in full abduction, on the mid-axillary line with the probe-tip placed in the 9th to 11th intercostal space with a minimum of 10 shots (A1)

- Correct interpretation of TE results in clinical practice must consider the following parameters:
  - IQR/ median value (<30%)
  - Serum aminotransferases levels (<5 x ULN)
  - BMI (use XL probe above 30 kg/m² or if skin-to-capsule distance is >25 mm)
  - Absence of extra-hepatic cholestasis
  - Absence of right heart failure, or other causes of congestive liver
  - Absence of ongoing excessive alcohol intake (A1)

- Although alternative techniques, such as pSWE/ARFI or 2D-SWE seem to overcome limitations of TE, their quality criteria for correct interpretation are not yet well defined (A1)

- At present correct interpretation of pSWE/ARFI results in clinical practice should systematically take into account the potentially confounding parameter:
  - fasting for at least 2 hours, transaminases levels (<5 x ULN), absence of extra-hepatic cholestasis and absence or right heart failure (B1)

- MR elastography is currently too costly and time-consuming for routine clinical practice use and seems more suited for research purposes (A1)

Endpoints for staging liver fibrosis

In patients with viral hepatitis and HIV-HCV coinfection, the clinically relevant endpoints are: (1) detection of significant fibrosis (META VIR, F ≥ 2 or Ishak, ≥ 3), which indicates that patients should receive antiviral treatment. However, with the availability of novel antiviral agents able to achieve sustained virological response (SVR) rates above 90% with limited side effects, it is likely that significant fibrosis will no longer represent an important decision making endpoint in HCV-infected patients. (2) Detection of cirrhosis (META VIR, F4 or Ishak, 5–6) indicates that patients should not only potentially be treated for longer duration/different regimens in HCV but also monitored for complications related to portal hypertension (PH) and regularly screened for hepatocellular carcinoma (HCC). In NAFLD, representing another major etiology of chronic liver disease, the presence of significant fibrosis does not represent a relevant endpoint in the absence of standardized treatment regimens. However, detection of septal (advanced) fibrosis-cirrhosis seems clinically more relevant in NAFLD patients. In alcoholic liver disease (ALD), cholestatic liver diseases, and other etiologies, cirrhosis represents the most relevant clinical endpoint.

Performance of serum biomarkers for staging liver fibrosis

The diagnostic performances of serum biomarkers of fibrosis are summarized in Table 4. Overall, biomarkers are less accurate in detecting intermediate stages of fibrosis than cirrhosis. The most widely used and validated are the APRI (a free non-patented index) and the FibroTest® (a patented test that is not widely available), mainly in viral hepatitis C. A recent systematic review including 172 studies conducted in hepatitis C [103] reported median AUROCs of 0.79 and 0.86 for FibroTest® and of 0.77 and 0.84 for APRI, for significant fibrosis and cirrhosis, respectively. A meta-analysis by the developer [104] that analyzed data from 6378 subjects (individual data from 3282 subjects) who received the FibroTest® and biopsies (3501 with HCV infection and 1457 with HBV) found that the mean standardized AUROC for diagnosis of significant fibrosis was 0.84, without significant differences between patients with HCV (0.85) and HBV (0.80). Another meta-analysis [105] analyzed results from 6259 HCV patients from 33
studies; the mean AUROC values of APRI in diagnosis of significant fibrosis and cirrhosis were 0.77 and 0.83, respectively. Another meta-analysis of APRI in 1798 HBV patients found mean AUROC values of 0.79 and 0.75 for significant fibrosis and cirrhosis, respectively [106]. In the largest comparative study to date (n = 510 patients mono-infected with hepatitis B or C matched on fibrosis stage), overall diagnostic performances of blood tests (FibroTest®), FibroMeter®, and HepaScore® were similar between hepatitis B and C with AUROC ranging from 0.75 to 0.84 for significant fibrosis, 0.82 to 0.85 for extensive fibrosis and 0.84 to 0.87 for cirrhosis, respectively [107].

In HIV-HCV coinfected patients, performance of non-patented tests (e.g., APRI, FIB-4, and the Forns index) for predicting fibrosis seems less accurate than in HCV mono-infected patients: they are accurate for the diagnosis of cirrhosis, but relatively inaccurate for the diagnosis of significant fibrosis [108–110]. As for patented tests, such as FibroTest®, FibroMeter®, and HepaScore®, they outperform the non-patented tests in HIV-HCV coinfected patients because of the risk of false positive results related to HIV-induced thrombocytopenia, antiretroviral treatment-induced hyperbilirubinemia or increased serum γ-glutamyl transferase levels [111,112]. Importantly, one should be aware of false positive results with APRI and FIB-4 (related to HIV-induced thrombocytopenia) as well as with FibroTest® and HepaScore® (related to hyperbilirubinemia induced by the use of antiretroviral treatment such as atanazavir) or FibroTest® and Forns Index (related to increase in γ-glutamyl transferase induced by nevirapine) [111].

In patients with NAFLD, the NAFLD fibrosis score [40] is currently the most studied [85,113–118] and validated biomarker [119]. The NAFLD fibrosis score seems to perform better in Caucasians than Asians, probably related to the ethical difference in fat distribution and its influence on the BMI [102].

### Recommendations

- Serum biomarkers of fibrosis are well validated in patients with chronic viral hepatitis (with more evidence for HCV than for HBV and HIV/HCV coinfected). They are less well validated in NAFLD and not validated in other chronic liver diseases (A1).
- Their performances are better for detecting cirrhosis than significant fibrosis (A1).
- Caution is needed in patients with HIV-HCV coinfected because of the risk of false positive results related to HIV-induced thrombocytopenia, antiretroviral treatment-induced hyperbilirubinemia or increased serum γ-glutamyl transferase levels (A2).
- FibroTest®, APRI and NAFLD fibrosis score are the most widely used and validated patented and non-patented tests (A2).

### Comparative performance of patented and non-patented serum biomarkers for staging liver fibrosis

When compared and validated externally in patients with hepatitis C [120–125], the different patented tests had similar levels of performance in diagnosis of significant fibrosis. In the largest independent study (1370 patients with viral hepatitis; 913 HCV...
Clinical Practice Guidelines

and 284 HBV patients), which prospectively compared the widely used patented tests (FibroTest®, FibroMeter®, and HepaScore®) with the non-patented test (APRI), the AUROC values for significant fibrosis ranged from 0.72 to 0.78 with no significant differences among scores [124]. In patients with cirrhosis, the AUROC values were higher for all tests, ranging from 0.77 to 0.86, with no significant differences among the tests. Although non-patented tests such as the Forns index, FIB-4, and APRI were not as accurate as patented tests [125], there are no additional costs, they are easy to calculate, and are widely available.

Recommendations

- When compared in HCV patients, the different patented tests have similar levels of performance in diagnosing significant fibrosis and cirrhosis (A1)
- Although non-patented tests might have lower diagnostic accuracy than patented tests, they are not associated with additional costs, are easy to calculate, and are widely available (A2)

Table 5. Diagnostic performance of TE for significant fibrosis (F ≥2) and cirrhosis (F4) in patients with viral hepatitis B and C.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Etiologies</th>
<th>Year</th>
<th>Patient (n)</th>
<th>F≥2 (%)</th>
<th>F4 (%)</th>
<th>Cut-offs (kPa)</th>
<th>AUROC</th>
<th>Se (%)</th>
<th>Sp (%)</th>
<th>CC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castéra et al. [126]</td>
<td>HCV</td>
<td>2005</td>
<td>183</td>
<td>74</td>
<td></td>
<td>7.1</td>
<td>0.83</td>
<td>67</td>
<td>89</td>
<td>73</td>
</tr>
<tr>
<td>Ziol et al. [127]</td>
<td>HCV</td>
<td>2005</td>
<td>251</td>
<td>65</td>
<td></td>
<td>8.6</td>
<td>0.79</td>
<td>56</td>
<td>91</td>
<td>68</td>
</tr>
<tr>
<td>Arena et al. [86]</td>
<td>HCV</td>
<td>2008</td>
<td>150</td>
<td>56</td>
<td></td>
<td>7.8</td>
<td>0.91</td>
<td>83</td>
<td>82</td>
<td>83</td>
</tr>
<tr>
<td>Lupšor et al. [128]</td>
<td>HCV</td>
<td>2008</td>
<td>324</td>
<td>65</td>
<td></td>
<td>7.4</td>
<td>0.86</td>
<td>76</td>
<td>84</td>
<td>79</td>
</tr>
<tr>
<td>Wang et al. [134]</td>
<td>HCV</td>
<td>2009</td>
<td>214</td>
<td>42</td>
<td></td>
<td>9.5</td>
<td>0.82</td>
<td>70</td>
<td>83</td>
<td>n.a.</td>
</tr>
<tr>
<td>Degos et al. [124]</td>
<td>HCV</td>
<td>2010</td>
<td>913</td>
<td>62</td>
<td></td>
<td>12.9</td>
<td>0.75</td>
<td>90</td>
<td>85</td>
<td>57</td>
</tr>
<tr>
<td>Zarski et al. [125]</td>
<td>HCV</td>
<td>2012</td>
<td>382</td>
<td>47</td>
<td></td>
<td>5.2</td>
<td>0.93</td>
<td>97</td>
<td>85</td>
<td>57</td>
</tr>
<tr>
<td>Coco et al. [69]</td>
<td>HBV (HCV)</td>
<td>2007</td>
<td>228</td>
<td>62</td>
<td></td>
<td>8.3</td>
<td>0.93</td>
<td>85</td>
<td>91</td>
<td>87</td>
</tr>
<tr>
<td>Oliveri et al. [130]</td>
<td>HBV</td>
<td>2008</td>
<td>188</td>
<td>26</td>
<td></td>
<td>7.5</td>
<td>0.97</td>
<td>94</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>Marcellin et al. [131]</td>
<td>HBV</td>
<td>2009</td>
<td>173</td>
<td>50</td>
<td></td>
<td>7.2</td>
<td>0.81</td>
<td>70</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>Chan et al. [132]</td>
<td>HBV</td>
<td>2009</td>
<td>161</td>
<td>71</td>
<td></td>
<td>12-13.4*</td>
<td>0.93</td>
<td>98</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>Kim et al. [133]</td>
<td>HBV</td>
<td>2009</td>
<td>91</td>
<td>43</td>
<td></td>
<td>9.7</td>
<td>0.80</td>
<td>82</td>
<td>59</td>
<td>62</td>
</tr>
<tr>
<td>Wang et al. [134]</td>
<td>HBV</td>
<td>2009</td>
<td>88</td>
<td>42</td>
<td></td>
<td>8.0</td>
<td>0.86</td>
<td>80</td>
<td>77</td>
<td>n.a.</td>
</tr>
<tr>
<td>Degos et al. [124]</td>
<td>HBV</td>
<td>2010</td>
<td>284</td>
<td>42</td>
<td></td>
<td>5.2</td>
<td>0.78</td>
<td>89</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>Sporea et al. [135]</td>
<td>HBV</td>
<td>2010</td>
<td>140</td>
<td>76</td>
<td></td>
<td>12.9</td>
<td>0.85</td>
<td>89</td>
<td>52</td>
<td>93</td>
</tr>
<tr>
<td>Cardoso et al. [136]</td>
<td>HBV</td>
<td>2012</td>
<td>202</td>
<td>42</td>
<td></td>
<td>7.2</td>
<td>0.87</td>
<td>74</td>
<td>88</td>
<td>82</td>
</tr>
<tr>
<td>Goyal et al. [137]</td>
<td>HBV</td>
<td>2013</td>
<td>357</td>
<td>25</td>
<td></td>
<td>6.0</td>
<td>0.84</td>
<td>82</td>
<td>67</td>
<td>n.a.</td>
</tr>
<tr>
<td>Afdhal et al. [129]</td>
<td>HCV/HBV</td>
<td>2015</td>
<td>560**</td>
<td>66.7</td>
<td></td>
<td>8.4</td>
<td>0.73</td>
<td>58</td>
<td>75</td>
<td>70</td>
</tr>
</tbody>
</table>

HCV, chronic hepatitis C; HBV, chronic hepatitis B; AUROC, area under ROC curve; Se, sensitivity; Sp, specificity; CC, correctly classified: true positive and negative; n.a, not available.

*More than half of patients with «clinical» cirrhosis; adapted to ALT levels.
**Validation cohort: HCV 92%; HBV 8%.
*Adapted to LT levels.
AUROC values of 0.94 and 0.84, respectively [147]. In a recent performance of TE for cirrhosis than for fibrosis, with mean meta-analysis of data from individual patient data.

AUROC values for diagnosing cirrhosis and significant fibrosis (AUROC values, 0.65–0.97; correct classification ranging from 80% to 98%) than TE more accurately detects cirrhosis (AUROC values, 0.80–0.99; correct classification ranging from 80% to 98%) than significant fibrosis (AUROC values, 0.65–0.97; correct classification from 57% to 90%) (Table 5 and Table 6). Several meta-analyses [145–149] have confirmed the better diagnostic performance of TE for cirrhosis than for fibrosis, with mean AUROC values of 0.94 and 0.84, respectively [147]. In a recent meta-analysis of 18 studies including 2772 HBV patients [150], mean AUROC values for diagnosing cirrhosis and significant fibrosis were 0.93 and 0.86, respectively. However, we are still lacking a meta-analysis of data from individual patient data.

Different cut-offs have been proposed for cirrhosis according to etiologies ranging from 9.7 kPa in HBV [133] to 22.7 kPa in ALD [151]. However, it must be kept in mind that these cut-off values have been defined in a single population using ROC curves in order to maximize sensitivity and specificity – and not applied to a validation cohort. Difference between cut-offs may be simply related to difference in cirrhosis prevalence in the studied populations (ranging from 8% to 54%; Tables 5 and 6), known as the spectrum bias [16,17]. Based on a meta-analysis, some authors have proposed an optimal cut-off of 13 kPa for the diagnosis of cirrhosis [147]. However, the cut-off choice must also consider the pre-test probability of cirrhosis in the target population (varying from less than 1% in the general population to 10% to 20% in tertiary referral centres). For instance, it has been shown that in a population with a pre-test probability of 13.8%, at a cut-off <7 kPa, cirrhosis probability ranged from 0% to 3% whereas at a cut-off >17 kPa cirrhosis probability was 72% [124]. When compared, the performances of TE have been shown to be similar between patients with HBV and HCV [135,136]. Serum levels of aminotransferases should always be taken into account when interpreting results from TE, especially in patients with hepatitis B (who might have flares) [152]. To avoid the risk of false positive results, some authors have proposed to adapt TE cut-offs based on levels of ALT [132], a strategy that might not apply to patients with fluctuating levels of ALT or hepatitis flares (Table 5). Conversely, in hepatitis e antigen (HBeAg)-negative patients with normal levels of ALT, non-invasive methods, particularly TE, could be used as adjunct tools to measure HBV DNA, to follow inactive carriers or better identify patients who require liver biopsy (those with ongoing disease activity or significant fibrosis, despite normal levels of ALT) [130,153–155].

### Performance of TE for staging liver fibrosis

Performances of TE for diagnosing significant fibrosis and cirrhosis are summarized in Table 5 (viral hepatitis) & Table 6 (non-viral hepatitis). The two index studies suggesting the interest of TE in the assessment of liver fibrosis have been conducted in patients with chronic hepatitis C [126,127]. LS values strongly correlated with METAVIR fibrosis stages. However, it should be emphasized that despite high AUROC values, a substantial overlap of LS values was observed between adjacent stages of hepatic fibrosis, particularly for lower fibrosis stages. Many other groups have since confirmed these results [86,124,125,128,129] also in patients with hepatitis B [69,124,129–137] as well as in patients with HIV-HCV coinfection [138–143].

TE is a reliable method for the diagnosis of cirrhosis in patients with chronic liver diseases, better at ruling out than ruling in cirrhosis (negative and positive predictive values 96% and 74%) [144]. TE more accurately detects cirrhosis (AUROC values, 0.80–0.99; correct classification ranging from 80% to 98%) than significant fibrosis (AUROC values, 0.65–0.97; correct classification from 57% to 90%) (Table 5 and Table 6). Several meta-analyses [145–149] have confirmed the better diagnostic performance of TE for cirrhosis than for fibrosis, with mean AUROC values of 0.94 and 0.84, respectively [147]. In a recent meta-analysis of 18 studies including 2772 HBV patients [150], mean AUROC values for diagnosing cirrhosis and significant fibrosis were 0.93 and 0.86, respectively. However, we are still lacking a meta-analysis of data from individual patient data.

### Table 6. Diagnostic performance of TE for F ≥ 2 and F4 in chronic liver diseases other than viral hepatitis.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Etiologies</th>
<th>Year</th>
<th>Patient (n)</th>
<th>F≥2 (%)</th>
<th>F4 (%)</th>
<th>Cut-offs (kPa)</th>
<th>AUROC</th>
<th>Se (%)</th>
<th>Sp (%)</th>
<th>CC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpechot et al. [163]</td>
<td>PBC-PSC</td>
<td>2006</td>
<td>95</td>
<td>60</td>
<td>16</td>
<td>7.3</td>
<td>0.92</td>
<td>84</td>
<td>87</td>
<td>75</td>
</tr>
<tr>
<td>Corpechot et al. [164]</td>
<td>PBC</td>
<td>2012</td>
<td>103</td>
<td>50</td>
<td>8.8</td>
<td>17.3</td>
<td>0.96</td>
<td>93</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Ganne-Carrie et al. [144]</td>
<td>Mixed</td>
<td>2006</td>
<td>1007</td>
<td>15</td>
<td>14.6</td>
<td>16.9</td>
<td>0.99</td>
<td>79</td>
<td>95</td>
<td>92</td>
</tr>
<tr>
<td>Foucher et al. [162]</td>
<td>Mixed</td>
<td>2007</td>
<td>354</td>
<td>13</td>
<td>17.6</td>
<td>16.9</td>
<td>0.99</td>
<td>77</td>
<td>97</td>
<td>n.a.</td>
</tr>
<tr>
<td>Fraquelli et al. [56]</td>
<td>Mixed</td>
<td>2007</td>
<td>200</td>
<td>12</td>
<td>7.9</td>
<td>11.9</td>
<td>0.90</td>
<td>91</td>
<td>89</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nguyen-Khac et al. [165]</td>
<td>ALD</td>
<td>2008</td>
<td>103</td>
<td>75</td>
<td>7.8</td>
<td>19.5</td>
<td>0.91</td>
<td>80</td>
<td>91</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nahon et al. [151]</td>
<td>ALD</td>
<td>2008</td>
<td>147</td>
<td>75</td>
<td>7.8</td>
<td>32</td>
<td>0.92</td>
<td>86</td>
<td>84</td>
<td>n.a.</td>
</tr>
<tr>
<td>Yoneda et al. [156]</td>
<td>NAFLD</td>
<td>2008</td>
<td>97</td>
<td>50</td>
<td>6.6</td>
<td>17.0</td>
<td>0.96</td>
<td>88</td>
<td>74</td>
<td>n.a.</td>
</tr>
<tr>
<td>Nobili et al. [157]</td>
<td>NAFLD</td>
<td>2008</td>
<td>50</td>
<td>24</td>
<td>4.7</td>
<td>17.0</td>
<td>0.99</td>
<td>100</td>
<td>97</td>
<td>n.a.</td>
</tr>
<tr>
<td>Luporso et al. [158]</td>
<td>NAFLD</td>
<td>2010</td>
<td>72</td>
<td>25</td>
<td>6.8</td>
<td>17.0</td>
<td>0.99</td>
<td>100</td>
<td>92</td>
<td>n.a.</td>
</tr>
<tr>
<td>Wong et al. [85]</td>
<td>NAFLD</td>
<td>2010</td>
<td>246</td>
<td>10</td>
<td>7.0</td>
<td>10.3</td>
<td>0.95</td>
<td>92</td>
<td>88</td>
<td>n.a.</td>
</tr>
<tr>
<td>Gaia et al. [82]</td>
<td>NAFLD</td>
<td>2011</td>
<td>72</td>
<td>46</td>
<td>7.0</td>
<td>12.5</td>
<td>0.94</td>
<td>78</td>
<td>96</td>
<td>80</td>
</tr>
<tr>
<td>Petta et al. [159]</td>
<td>NAFLD</td>
<td>2011</td>
<td>169</td>
<td>47</td>
<td>7.25</td>
<td>15.5</td>
<td>0.94</td>
<td>78</td>
<td>96</td>
<td>80</td>
</tr>
<tr>
<td>Myers et al. [66]</td>
<td>PBC-PSC</td>
<td>2012</td>
<td>75</td>
<td>n.a.</td>
<td>7.8</td>
<td>22.3</td>
<td>0.88</td>
<td>80</td>
<td>91</td>
<td>n.a.</td>
</tr>
<tr>
<td>Wong et al. [68]</td>
<td>NAFLD</td>
<td>2012</td>
<td>193</td>
<td>45</td>
<td>7.8</td>
<td>10.3</td>
<td>0.89</td>
<td>78</td>
<td>94</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; NAFLD, non-alcoholic fatty liver disease; ALD, alcoholic liver disease.

AUROC, area under ROC curve; Se, sensitivity; Sp, specificity; CC, correctly classified; true positive and negative; n.a., not available.

Please cite this article in press as: EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol (2015), http://dx.doi.org/10.1016/j.jhep.2015.04.006
Clinical Practice Guidelines

TE has also been investigated in NAFLD patients but in a smaller number of studies [66,68,82,85,156–159] (Table 6). Like in viral hepatitis, TE performances are better for cirrhosis than for significant fibrosis with AUROCs ranging from 0.94 to 0.99 and from 0.79 to 0.99, respectively. However, the performance of TE in NAFLD deserves several comments: Firstly, these studies have been conducted in heterogeneous and special populations such as Asian children or patients with low BMI (<28 kg/m²); secondly, most of them are underpowered with small sample size (<100 patients) and very few patients with cirrhosis; thirdly, the histological scoring systems such as those proposed by Brunt et al. [160] or Kleiner et al. [161] and endpoints (significant fibrosis or severe fibrosis) were heterogeneous in most studies evaluating fibrosis by TE in NAFLD. These differences in the study designs are likely the explanation for the observed differences among proposed cut-offs for a given endpoint (ranging for instance from 10.3 to 22.3 kPa for cirrhosis) (Table 6), known as the spectrum bias [16,17]. Finally, all these studies have been conducted in tertiary referral centres with a higher proportion of patients with severe fibrosis than in the general population, making it difficult to extrapolate the performance of TE in detecting cirrhosis in large populations. Nevertheless, TE could be of interest to exclude confidently severe fibrosis and cirrhosis with high negative predictive value (around 90%) in NAFLD patients [85].

TE has also been evaluated in a variety of chronic liver diseases [56,144,162], as well as in primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) [163,164], and ALD [151,165] (Table 6). However, in the latter it has been suggested by several groups that the presence of alcoholic hepatitis may influence LS results [74–76] and thus, TE should be ideally performed only after alcohol withdrawal in order to improve diagnostic accuracy.

Recommendations

- TE can be considered the non-invasive standard for the measurement of LS (A1)
- TE is well validated in viral hepatitis with performance equivalent in hepatitis B and C and in HIV-HCV coinfection (A1)
- TE is less well validated in NAFLD and in other chronic liver diseases (A1)
- TE performs better for detection of cirrhosis than for detection of significant fibrosis (A1)
- TE is a reliable method for the diagnosis of cirrhosis in patients with chronic liver diseases, that generally performs better at ruling out than ruling in cirrhosis (with negative predictive value higher than 90%) (A1)

Performance of other techniques for staging liver fibrosis

Point shear wave elastography using acoustic radiation force impulse quantification

Performances of pSWE/ARFI (Siemens) for diagnosing significant fibrosis and cirrhosis are summarized in Table 7. Most studies evaluated patients with mixed chronic liver disease with viral hepatitis being the predominant liver disease [166–177]. Similar to TE, pSWE/ARFI more accurately detects cirrhosis (AUROC values: 0.81–0.99) than significant fibrosis (AUROC values: 0.77–0.94). The largest study evaluating pSWE/ARFI for staging of chronic hepatitis C was a retrospective pooled analysis of 914 international patient data [178], part of which were published in smaller single centre studies previously [166,167,170,171,174,179]. It reported sensitivity and specificity of pSWE/ARFI for the diagnosis of significant fibrosis of 0.69 and 0.80 and for the diagnosis of liver cirrhosis of 0.84 and 0.76, respectively [178].

Meta-analyzes have confirmed the better diagnostic performance of pSWE/ARFI for cirrhosis than for fibrosis [180,181]. In a pooled meta-analysis including 518 individual patients with chronic liver disease (83% with viral hepatitis) mean AUROCs were 0.87 for the diagnosis of significant fibrosis, and 0.93 for the diagnosis of liver cirrhosis [180]. In a meta-analysis of 36 studies (21 full paper publications and 15 abstracts) comprising 3951 patients mean AUROCs were 0.84 (diagnostic odds ratio [DOR]: 11.54) for the diagnosis of significant fibrosis, and 0.91 (DOR: 45.35) for the diagnosis of liver cirrhosis [181]. Cut-off values suggested in the two meta-analyses were 1.34–1.35 m/sec for the diagnosis of significant fibrosis and 1.80–1.87 m/sec for the diagnosis of cirrhosis. Only few studies have evaluated pSWE/ARFI in chronic hepatitis B [182,183] and reported comparable results as for chronic hepatitis C and mixed chronic liver disease.

In a few studies pSWE/ARFI has also been investigated in NAFLD [184–187]. Such as in viral hepatitis, pSWE/ARFI performances are better for severe fibrosis and cirrhosis than for significant fibrosis with AUROCs ranging from 0.91 to 0.98 and from 0.66 to 0.86, respectively. Interestingly, 60% of patients with BMI between 30 and 40 kg/m² and 58% of patients with BMI >40 kg/m² could be successfully evaluated using pSWE/ARFI [186]. Finally, pSWE/ARFI has also been evaluated in a variety of chronic liver diseases (ALD, PBC, PSC, and autoimmune hepatitis (AIH)). However, since most studies included mixed chronic liver diseases with predominantly viral hepatitis, the value of pSWE/ARFI for less common etiologies of chronic liver disease needs further evaluation.

2D-shear wave elastography

Only few studies [96,97,188,189] have evaluated 2D-SWE for the staging of liver fibrosis, two of which used liver biopsy as reference method [97,189]. In a pilot study in 121 patients with chronic hepatitis C (METAIVIB fibrosis stage 41% F0/F1, 27% F2, 12% F3, and 20% F4), AUROCs of 2D-SWE for the diagnosis of significant fibrosis and cirrhosis were 0.92 and 0.98, respectively [189]. In another study in 226 patients with chronic hepatitis B (METAIVIB fibrosis stage 17% F0, 23% F1, 25% F2, 20% F3, and 15% F4), 2D-SWE had AUROCs of 0.88 and 0.98 for the diagnosis of significant fibrosis and cirrhosis, respectively [97]. Sensitivities and specificities were 85% and 92% for the diagnosis of significant fibrosis using a cut-off of 7.1 kPa, and 97% and 93% for the diagnosis of cirrhosis using a cut-off of 10.1 kPa.

Other elastography methods such as strain elastography (a quasi-static technique) are available, but data for the staging of liver fibrosis are insufficient and seem to suggest that strain elastography has a worse diagnostic performance as compared to shear wave elastography [190].
Transient elastography vs. other techniques

Studies comparing TE and pSWE using ARFI show varying results. While many studies reported comparable results for both methods [167,174,179,191,192], some studies report better results for ARFI [172] and others better results for TE [168,174], respectively. In a recent meta-analysis [90] including 13 studies (n = 1163 patients) comparing pSWE using ARFI with TE (11 full-length articles and two abstracts), no significant difference in DOR were found between ARFI and TE. Summary sensitivities and specificities for the diagnosis of significant fibrosis were 0.74 and 0.83 for ARFI and 0.78 and 0.84 for TE, respectively and 0.87 and 0.87 for ARFI and 0.89 and 0.87 for TE for the diagnosis of cirrhosis, respectively.

2D-SWE has been compared to TE in only three studies [97,100,189]. In chronic hepatitis C [189], AUROCs of SWE were significantly higher than with TE for the diagnosis of significant fibrosis (0.92 vs. 0.84, respectively; \( p = 0.002 \)) but not for cirrhosis (0.98 vs. 0.96, \( p = 0.48 \)). In chronic hepatitis B, AUROCs for SWE were significantly higher for both significant fibrosis (0.88 vs. 0.78) and cirrhosis (0.98 vs. 0.92) [97]. In 349 patients with chronic liver disease [100], SWE had a higher accuracy than TE for the diagnosis of severe fibrosis (\( \geq F3 \)) (\( p = 0.0016 \)), and a higher accuracy than pSWE using ARFI for the diagnosis of significant fibrosis (\( \geq F2 \)) (\( p = 0.0003 \)).

MR elastography has been compared to TE in patients with chronic liver diseases in three studies with conflicting results [193–195]. Two studies (a pilot Belgian study [193] and a Japanese retrospective study [195]) in 96 and 113 patients with chronic liver disease) suggested that MR elastography might be more accurate than TE in diagnosis of significant fibrosis whereas another study from the Netherlands [194] in 85 patients with viral hepatitis reported similar accuracy for significant fibrosis. Further data are required to evaluate if MR elastography has superior accuracy for detecting significant fibrosis and cirrhosis as compared to TE, pSWE/ARFI, or 2D-SWE.

**Recommendations**

- pSWE/ARFI performs better for detecting cirrhosis than significant fibrosis and is better validated in chronic hepatitis C than for hepatitis B, HIV-HCV coinfection, NAFLD and other liver diseases (A1)
- pSWE/ARFI shows equivalent performance to TE for detecting significant fibrosis and cirrhosis (A1)
- 2D-SWE is a promising technique that is currently under investigation. It seems to be at least equivalent to TE and pSWE/ARFI for non-invasive staging of liver fibrosis in viral hepatitis (B1)
- Comparison between MR elastography and TE has provided conflicting results. Further data are needed (A1)
Clinical Practice Guidelines

Comparison of performance of TE and serum biomarkers for staging liver fibrosis

Many studies have compared the performances of TE and serum biomarkers, mostly in viral hepatitis [124–126,143,196–203] but also in NAFLD and ALD [85,165]. TE and serum biomarkers have been shown to have equivalent performance for detecting significant fibrosis [124–126] but TE outperforms serum biomarkers for detecting cirrhosis [124,196,199]. However, given the lower applicability of TE (80% vs. 95% for serum biomarkers), performance could finally not differ for intention-to-diagnose analysis [125].

Recommendations

- TE and serum biomarkers have equivalent performance for detecting significant fibrosis in patients with viral hepatitis (A1)
- TE is the most accurate non-invasive method for detecting cirrhosis in patients with viral hepatitis (A1)

Algorithms combining different tests (LS and/or serum biomarkers)

Since the first proposal of a strategy combining TE and FibroTest® to increase diagnostic accuracy in patients with hepatitis C [126], many algorithms combining either TE and serum biomarkers [125,143,198–200,202,204,205] or several serum biomarkers [122,206–210] have been proposed, mainly in patients with viral hepatitis. Although these algorithms are more effective in detecting significant fibrosis than individual tests, they do not increase diagnostic accuracy for cirrhosis [125,196,199]. However, given the important clinical implications, in terms of prognosis, monitoring and treatment decisions that follow the diagnosis of cirrhosis, it seems justified to confirm a diagnosis of cirrhosis by two concordant but unrelated tests. Also ultrasound and other imaging methods hold a high specificity for the diagnosis of cirrhosis in this context, and may be useful as an unrelated method.

The advantage of combining two unrelated methods, such as TE and serum biomarkers, over the combination of two serum biomarkers is that TE provides more direct measurement of the liver structure than biomarkers, and that there is no relationship between the applicability of TE (success rate and interquartile range) and that of a biomarker [204,211]. Also, the combination of TE and serum biomarkers might be more effective than the combination of two serum biomarkers for detecting significant fibrosis (significantly greater number of saved liver biopsies) [200,212]. However, this strategy has only been validated in studies of patients with hepatitis C, is more costly, and could be hampered by the lower applicability of TE, compared with biomarkers. Most importantly, in case of unexplained discordance of non-invasive tests, a liver biopsy should still be performed.

Recommendations

- Among the different available strategies, algorithms combining TE and serum biomarkers appear to be the most attractive and validated one (A2)
- In patients with viral hepatitis C, when TE and serum biomarkers results are in accordance, the diagnostic accuracy is increased for detecting significant fibrosis but not for cirrhosis. In cases of unexplained discordance, a liver biopsy should be performed if the results would change the patient management. Such strategy remains to be validated in patients with hepatitis B and NAFLD (A1)

Indications for non-invasive tests for staging liver disease in viral hepatitis

HCV including HIV-HCV

In the clinical management of HCV patients including those coinfected with HIV, there are several specific indications where the clinician can use non-invasive tests to aid in disease management. Either alone or in combination these tests allow for rapid staging of liver disease without the need for liver biopsy. The current gold standard for utilization of non-invasive tests to stage liver disease is to combine a serum biomarker with TE. The key for accuracy is to have concordance between the tests, which increases the diagnostic accuracy (Fig. 1). Every patient with chronic HCV infection should have liver disease staging at least
once by non-invasive tests. Once a diagnosis of cirrhosis has been established, both AASLD and EASL guidelines recommend that those patients should be screened for PH and HCC [213,214]. Therefore all HCV patients need to be staged as part of routine HCV care to exclude cirrhosis. The diagnostic accuracy of TE for cirrhosis has been confirmed by multiple studies and meta-analyses and has proven superior to that reported by serum biomarkers.

Recommendations

- All HCV patients should be screened to exclude cirrhosis by TE if available. Serum biomarkers can be used in the absence of TE (A1)
- HCV patients who were diagnosed with cirrhosis based on non-invasive diagnosis should undergo screening for HCC and PH and do not need confirmatory liver biopsy (A1)

HBV

In chronic hepatitis B, TE generally has a higher AUROC as compared to serum biomarkers for advanced liver fibrosis [198,202]. Among inactive carriers with normal transaminases, TE also has less fluctuation over time as compared to FibroTest® or APRI score [155]. LS of <5–6 kPa often indicates absent or minimal liver fibrosis [132,153]. On the other hand, LS of >12–14 kPa often indicates liver cirrhosis (Table 5). Among patients with intermediate LS measurements, the accuracy of staging is lower. In doubtful cases, liver biopsy is recommended (Fig. 2). Among chronic hepatitis B patients who have elevated ALT levels or ALT flares, interpretation of LS measurement should be taken with caution. LS can be misleadingly high among patients who have severe acute exacerbation of chronic hepatitis B, even 3–6 months after ALT has been normalized [215].

For HBeAg-positive patients, particularly among those who are older than 35 years of age with high normal ALT levels, non-invasive assessment of liver fibrosis is useful to differentiate whether patients are in immune tolerance phase or already have significant liver fibrosis secondary to immune clearance [216]. In HBeAg-negative patients, the low replicative phase is indicated by normal ALT level and low HBV DNA (<2000 IU/ml). On the other hand, the reactivation phase is characterized by elevated HBV DNA levels with intermittent elevation of ALT levels. Patients who have repeated and prolonged reactivation have higher risks of developing liver cirrhosis [217]. Non-invasive assessment of liver fibrosis is preferred over liver biopsy among HBeAg-negative patients with low (<2000 IU/ml) or borderline (>2000 to 20,000 IU/ml) HBV DNA and normal ALT levels, as the
Clinical Practice Guidelines

risks of advanced fibrosis and cirrhosis in these patients are usually below 10% [218].

Recommendations

- TE has better prediction for advanced liver fibrosis and cirrhosis than serum biomarkers in chronic hepatitis B (B1)
- TE is best used to determine liver fibrosis in hepatitis B patients with active viraemia (HBV DNA >2000 IU/ml) but normal ALT (A1)
- TE can be used to exclude severe fibrosis and cirrhosis in inactive carriers (HBsAg-negative, low viral load (HBV DNA <2000 IU/ml) and normal ALT). Liver biopsy should only be considered in doubtful cases after TE (A1)
- LS measurement should be interpreted with caution among patients with elevated ALT, and should not be used in patients with very high ALT levels (>10 x ULN) (A1)

Use of non-invasive tests for staging liver disease in NAFLD

NAFLD is a very common condition with reported prevalence of approximately 20% in different parts of the world [219,220]. Simple steatosis does not increase mortality. Fibrosis is the most important prognostic factor in NAFLD and is correlated with liver-related outcomes and mortality [2,221]. Advanced fibrosis, as determined by non-invasive serum biomarker, has been shown to predict liver-related complications and mortality [222,223]. Not all NAFLD patients will develop advanced fibrosis. Biopsy series suggested a prevalence of advanced fibrosis in 50% of NAFLD patients [222], but a population-based study in Hong Kong revealed only 3.7% of the 264 NAFLD patients had advanced fibrosis [224]. NAFLD patients with metabolic syndrome and those with type 2 diabetes mellitus, had been shown to be at increased risk of having liver fibrosis in both Western and Asian cohorts [220,225]. Fibrosis progression is possible among patients with simple steatosis or non-alcoholic steatohepatitis; approximately 25% to 37% of patients will have fibrosis progression in 3–5 years [226–228] [229]. Histologic inflammation and maybe metabolic factors are associated with higher risk of fibrosis progression among patients with simple steatosis or steatohepatitis [230].

Among the different serum biomarkers studied in NAFLD, only NFS and FIB-4 have been externally validated more than once, in different NAFLD populations and with consistent results [119]. These tests perform best at excluding severe fibrosis/cirrhosis (with negative predictive values >90%) and could therefore be used as a first line triage to identify patients at low risk of severe fibrosis. TE has excellent diagnostic accuracy for cirrhosis with a higher rate of false positive results than of false negative results and higher negative than positive predictive values. Therefore its ability to rule in severe fibrosis-cirrhosis may be insufficient for clinical decision making and may require histological confirmation.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of non-invasive tests for staging liver disease in NAFLD</strong></td>
<td></td>
</tr>
<tr>
<td>NAFLD is a very common condition with reported prevalence of approximately 20% in different parts of the world [219,220]. Simple steatosis does not increase mortality. Fibrosis is the most important prognostic factor in NAFLD and is correlated with liver-related outcomes and mortality [2,221]. Advanced fibrosis, as determined by non-invasive serum biomarker, has been shown to predict liver-related complications and mortality [222,223]. Not all NAFLD patients will develop advanced fibrosis. Biopsy series suggested a prevalence of advanced fibrosis in 50% of NAFLD patients [222], but a population-based study in Hong Kong revealed only 3.7% of the 264 NAFLD patients had advanced fibrosis [224]. NAFLD patients with metabolic syndrome and those with type 2 diabetes mellitus, had been shown to be at increased risk of having liver fibrosis in both Western and Asian cohorts [220,225]. Fibrosis progression is possible among patients with simple steatosis or non-alcoholic steatohepatitis; approximately 25% to 37% of patients will have fibrosis progression in 3–5 years [226–228] [229]. Histologic inflammation and maybe metabolic factors are associated with higher risk of fibrosis progression among patients with simple steatosis or steatohepatitis [230].</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recommendations</strong></td>
<td></td>
</tr>
<tr>
<td>• Screening of liver fibrosis for NAFLD patients is recommended, particularly among patients with metabolic syndrome or type 2 diabetes mellitus who have higher risk of liver fibrosis (A1)</td>
<td></td>
</tr>
<tr>
<td>• Non-invasive assessment including serum biomarkers or TE can be used as first line procedure for the identification of patients at low risk of severe fibrosis/cirrhosis (A1)</td>
<td></td>
</tr>
<tr>
<td>• The identification of significant fibrosis is less accurate with non-invasive tests as compared to liver biopsy and may necessitate, according to the clinical context, histological confirmation (A1)</td>
<td></td>
</tr>
<tr>
<td>• Follow-up assessment by either serum biomarkers or TE for progression of liver fibrosis should be performed among NAFLD patients at a 3 year interval (B1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Use of non-invasive tests for staging liver disease in other liver diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Alcoholic liver disease</td>
<td></td>
</tr>
<tr>
<td>Although the use of non-invasive tests in ALD has been explored, the methodological quality of existing studies is considerably heterogeneous without evaluation in large cohorts of ALD patients. Existing information on the usefulness of serum biomarkers has been recently summarized in the EASL guidelines for ALD and in recent reviews [231–233]. While a good performance has been reported for the use of FibroTest® in detecting significant fibrosis and cirrhosis (AUROC = 0.84 for F2-F4, AUROC = 0.95 for the diagnosis of cirrhosis), APRI has been found of limited use in the setting of ALD. Of note, FibroMeter® and HepaScore® have shown similar diagnostic accuracies than FibroTest® [234] with AUROC around 0.80 for significant fibrosis and 0.90 for cirrhosis. In addition, ELF® has also been shown to be useful in assessing fibrosis in ALD [34]. Interestingly, available data suggest that serum biomarkers of fibrosis may also be able to predict clinical outcomes [234,235]. Information on elasticity-based techniques, mainly TE, in ALD is limited due to the scarcity of single-etiolo-gy studies. A recent systematic review from the Cochrane Collaboration, based on five retrospective and nine prospective cohort studies with a total of 834 patients, suggests that TE may be used as a diagnostic method to rule out severe fibrosis or cirrhosis in patients with ALD using cut-offs of 9.5 and 12.5 kPa, respectively [236]. However, the authors point out the risk of outcome reporting bias as well as caution on the use of currently recommended cut-offs as they are insufficiently validated and because there is the risk of overestimation of LS values in patients that are not abstinent from alcohol consumption.</td>
<td></td>
</tr>
<tr>
<td>Cholestatic liver disease</td>
<td></td>
</tr>
<tr>
<td>Available information regarding the use of non-invasive tests in cholestatic diseases is indeed more limited than that for viral hepatitis and NAFLD. This is due to the fact that patients with these diseases are usually part of cohorts of chronic liver disease and disease specific data on non-invasive tests performance is</td>
<td></td>
</tr>
</tbody>
</table>
In the case of PSC, no specific studies are available in this regard. To differentiate between early and advanced fibrosis in PBC [241].

Thus, PBC patients often undergo a liver biopsy and the use of non-invasive tests of liver fibrosis may be advantageous in this setting. Several reports have tested the usefulness of serum biomarkers of liver fibrosis including serum levels of hyaluronic acid, procollagen III amino-terminal propeptide, collagen IV, and FibroTest® in patients with PBC [238–240]. While earlier studies did not provide estimates of diagnostic accuracy to readily assess clinical performance more recent reports do provide AUC values of ROC curves which in most of the cases are below 0.8. Thus, current evidence allows the conclusion that no single serum measurement has the ability to differentiate between early and advanced fibrosis in PBC [241].

In the case of PSC, no specific evidence is available in this regard.

Reported data on the use of TE in PBC is encouraging. The report by Corpechot et al. of two cohorts of PBC patients evaluated with TE showed that this technique is currently one of the best surrogate markers of liver fibrosis in this disease [164]. This data is in agreement with findings from Floreani’s group in Italy [242] and with an earlier study by Gomez-Dominguez from Spain [243]. In addition, TE may be useful to monitor PBC progression. In fact, prospective data from Corpechot et al. showed that progression of LS over time is predictive of poor outcome [164]. As for PSC, a recent study from the same group [244] showed that TE efficiently differentiates severe from non-severe liver fibrosis stages in this disease, and that both baseline LS measurements and increase over the time are able to predict patients’ outcomes. Thus, TE seems to be a reliable non-invasive method for assessing biliary fibrosis in PSC patients [163,244]. However, untreated dominant stricture of the common bile duct or primary hepatic ducts should be ruled out in PSC patients since obstructive cholestasis influences LS assessment [72]. Finally, with the availability of smaller probes (S1, S2), the use of TE has recently been tested in children with biliary atresia, a disease where fibrosis monitoring may help predict outcomes before surgery [245].

However, more data on non-invasive fibrosis evaluation in patients with cholestatic liver diseases is needed to make firm recommendations on the use of TE in this disease.

Autoimmune hepatitis

AH may have insidious onset in a significant proportion of the cases, which result in a large number of cases (30% to 80%) being at the cirrhotic stage at the moment of diagnosis. Since a significant number of cases could be diagnosed without performing a liver biopsy [246], non-invasive tests for liver fibrosis may have a role in liver disease staging. Non-invasive tests could be also useful for monitoring response to immunosuppressive treatment in AH, since fibrosis and cirrhosis can be reversible in this setting [247,248]. However, specific data of either serum markers of fibrosis or imaging techniques is scarce to make recommendations. Of note, disproportionally high results of TE [249] have been reported in patients with AH, which is likely related to the inflammatory activity considering that values decreased rapidly upon induction of disease remission.

Recommendations

- TE may be used to rule out severe fibrosis or cirrhosis in patients with ALD (B2)
- Non-invasive assessment of fibrosis, using TE should be considered in patients with PBC or PSC (B2)
- Follow-up assessment of liver fibrosis with TE should be performed in PBC and PSC since worsening of LS predicts patients’ outcomes. More data is needed to define the optimal time-frame between repeated examination (B2)
- Untreated dominant stricture of the common bile duct or primary hepatic ducts must be ruled out in PSC patients since obstructive cholestasis influences LS assessment (A1)
- No recommendation can be made with current evidence on the use of non-invasive tests in AIH (A1)

Use of non-invasive methods when deciding for treatment in viral hepatitis

HCV including HIV-HCV

The current recommendations for treatment of HCV vary significantly between countries and healthcare systems according to the availability of therapy. However, the EASL and AASLD guidelines are clear as to the prioritization of treatment based on disease stage [213,214]. There is some controversy in how best to use non-invasive tests in HCV therapeutic decisions. In countries where antiviral treatment is only indicated in patients with severe fibrosis or cirrhosis, both TE and serum biomarkers are effective – either alone, or in combination – to assess liver fibrosis. However, the controversy is with significant fibrosis, where all staging parameters including non-invasive tests and liver biopsy have the greatest discordance and risk for inaccuracy. Since the diagnostic accuracy of non-invasive tests in differentiating between stage F1 and F2 has the most variability, this represents a challenge for clinicians [17,18]. Although cut-offs for both TE and serum tests have been suggested for significant fibrosis, they have not been well validated and in a large, prospective US biopsy controlled study in over 700 HCV patients, TE, APRI, and FIB-4 all performed less well for significant fibrosis [129]. Combination of serum biomarkers with TE may marginally improve differentiation of F0–F1 from F2–F4 but has never been validated in actually differentiating between the single stages of F1 from F2.

In HIV-HCV coinfected patients, a priority has been given to treat all patients since this special population shows a more rapid disease progression so disease staging is less important for therapeutic decisions. However, in some countries anti-HCV treatment in HIV/HCV coinfected patients follows the same guidelines as for HCV monoinfection. There may be a reduced diagnostic accuracy of serum biomarkers for fibrosis in HIV-HCV patients and TE should be preferred.

There have been suggestions that as therapy becomes simpler and more effective with the advent of new direct-acting antiviral (DAA) agents and an increased uptake of HCV...
Clinical Practice Guidelines

screening, that community based physicians, infectious disease physicians and internists may treat HCV [250]. In this situation, the role of non-invasive tests is also very important for determination of appropriate referral of patients with more advanced liver disease to specialists for appropriate disease monitoring [81].

Recommendations

- Non-invasive tests, using either TE or serum biomarkers, are adequate for diagnosis of severe fibrosis/cirrhosis in HCV and HIV-HCV coinfected patients and can be used to prioritize patients for HCV therapy based on disease stage (A1)
- For the diagnosis of significant fibrosis a combination of tests with concordance may provide the highest diagnostic accuracy (A2)
- Non-invasive tests should be utilized prior to therapy by treating non-specialists to make sure that patients with severe fibrosis/cirrhosis are referred for appropriate disease specific specialist evaluation (A1)

HBV

Liver cirrhosis is the most important risk factor for liver-related complications and HCC in chronic hepatitis B. According to all regional guidelines, patients with liver cirrhosis and significant viraemia (HBV DNA >2000 IU/ml) should receive antiviral treatment regardless of the ALT levels [251–253]. Hence, non-invasive assessment of liver fibrosis can be considered in all patients in whom liver cirrhosis is suspected. Among hepatitis B patients who have elevated ALT but not yet reached two times ULN, liver fibrosis assessment can assist the decision of antiviral therapy. Patients who have significant liver fibrosis and HBV DNA >2000 IU/ml should be considered for antiviral therapy even if their ALT levels are below two times ULN [251,252]. Among patients who have persistently elevated ALT >2 times ULN and HBV DNA >2000 IU/ml, all regional guidelines recommend commencement of antiviral therapy and liver fibrosis assessment may not be necessary.

Recommendations

- Non-invasive assessment of liver fibrosis, using either serum biomarkers or TE, should be considered for patients with significant viraemia (HBV DNA >2000 IU/ml) when liver cirrhosis is suspected (A1)
- Among patients who have HBV DNA >2000 IU/ml, antiviral therapy should be considered for patients who have advanced fibrosis or cirrhosis as determined by non-invasive assessment of liver fibrosis, either by serum biomarkers or TE, regardless of the ALT levels (A1)

Use of non-invasive methods for monitoring treatment response in viral hepatitis

HCV including HIV-HCV

A major advantage of non-invasive tests, compared with liver biopsy, is that they can be easily repeated over time in patients receiving antiviral therapy and that they could be used to monitor response to treatment and to evaluate fibrosis regression. Several studies reported a significant decrease in LS and biomarkers values, compared with baseline values, in patients with HCV who achieved SVR [254–263], consistent with significant histologic improvement documented in studies of paired liver biopsies from HCV patients who achieved SVR [264,265]. However, the changing levels of ALT and inflammation of successfully treated HCV patients can confound results of TE or biomarkers. Indeed, the major component of the significant decrease observed in LS and biomarkers values is not just reversal of fibrosis but also reduction in liver injury, edema and inflammation.

There are two important clinical questions about the use of non-invasive tests after antiviral treatment. First, what is the evidence of fibrosis and particularly cirrhosis reversal by non-invasive tests? The reversal of cirrhosis has important consequences in that it may alter long-term prognosis particularly for HCC occurrence in HCV patients and change the approach to screening and surveillance for HCC after SVR. This leads into the second question, what are the cut-off thresholds post SVR for determination of decreased risk of liver-related outcomes?

In HCV, there is only one study that has examined reversal of cirrhosis in 33 patients with cirrhosis with pre- and post-treatment liver biopsies and TE after SVR [266]. There was reversal of cirrhosis by biopsy in 19 patients with 11 of the 19 being METAVIR F3 and the remainder F1 or F2. Using a cut-off of 12 kPa, TE had a sensitivity of 61% and a specificity of 95%. The low sensitivity makes TE a poor tool to be utilized clinically as evidence of cirrhosis regression. Non-invasive tests, including TE and serum biomarkers, have been shown to predict liver-related outcomes in HCV patients [267,268]. In both these studies clinical cut-offs of LS between 9.5 and 10.5 kPa were able to stratify patients at increased risk of clinical liver-related outcome. The best timing for repeated assessment of LS after therapy has not been established yet.

Recommendations

- Routine use of non-invasive tests during treatment or after SVR in non-cirrhotic patients does not add to clinical disease management (A1)
- Routine use of non-invasive tests after SVR in patients with HCV cirrhosis has a high false negative rate and cannot be used to determine which patients no longer need HCC screening or for the diagnosis of cirrhosis reversal (A2)
- Routine use of non-invasive tests after SVR has not yet established thresholds that predict low risk of liver-related events (A1)
HBV

Prolonged treatment with antiviral therapy is associated with resolution of liver fibrosis and regression of liver cirrhosis [269–271]. Non-invasive tests are an attractive strategy to monitor changes in fibrosis. A significant decrease in LS and biomarkers values, compared with baseline values, in HBV-infected patients treated with analogs has been reported [272–281]. However, like for HCV patients, improvement of LS in HBV patients who start antiviral therapy when ALT is elevated may be related to normalization of ALT instead of fibrosis improvement [274]. In this case, a LS measurement a few months after commencing treatment and normalization of ALT is recommended as baseline to monitor the changes in liver fibrosis. The value of non-invasive tests to predict the occurrence of complications or survival in patients with undetectable HBV DNA and cirrhosis prior to antiviral treatment remains to be determined [282–284].

Table 8. Prognostic performance of TE for predicting development of HCC in patients with chronic liver disease.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Etiologies</th>
<th>Year</th>
<th>Total patients (n)</th>
<th>HCC patients (n)</th>
<th>Region</th>
<th>Design</th>
<th>Follow-up duration (months)</th>
<th>AUROC</th>
<th>Cut off value (kPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masuzaki et al.</td>
<td>HCV</td>
<td>2008</td>
<td>265</td>
<td>85</td>
<td>Asia</td>
<td>Cross-sectional</td>
<td>-</td>
<td>0.805</td>
<td>25</td>
</tr>
<tr>
<td>Kuo et al.</td>
<td>Mixed</td>
<td>2010</td>
<td>435</td>
<td>106</td>
<td>Asia</td>
<td>Cross-sectional</td>
<td>-</td>
<td>0.736</td>
<td>24</td>
</tr>
<tr>
<td>Feier et al.</td>
<td>HCV</td>
<td>2013</td>
<td>144</td>
<td>72</td>
<td>Europe</td>
<td>Cross-sectional</td>
<td>-</td>
<td>0.680</td>
<td>38.5</td>
</tr>
<tr>
<td>Masuzaki et al.</td>
<td>HCV</td>
<td>2009</td>
<td>866</td>
<td>77</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>36.0</td>
<td>n.a.</td>
<td>25</td>
</tr>
<tr>
<td>Akima et al.</td>
<td>HCV*</td>
<td>2011</td>
<td>157</td>
<td>41 (10)**</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>40.7</td>
<td>0.787</td>
<td>12.5</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>HCV</td>
<td>2013</td>
<td>198</td>
<td>10</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>47.8</td>
<td>n.a.</td>
<td>12</td>
</tr>
<tr>
<td>Narita et al.</td>
<td>HCV*</td>
<td>2013</td>
<td>151</td>
<td>9</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>24.1</td>
<td>n.a.</td>
<td>14</td>
</tr>
<tr>
<td>Jung et al.</td>
<td>HBV</td>
<td>2011</td>
<td>1130</td>
<td>57</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>30.7</td>
<td>n.a.</td>
<td>8</td>
</tr>
<tr>
<td>Chon et al.</td>
<td>HBV</td>
<td>2012</td>
<td>1126</td>
<td>63</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>30.7</td>
<td>0.789</td>
<td>n.a.</td>
</tr>
<tr>
<td>Fung et al.</td>
<td>HBV*</td>
<td>2011</td>
<td>528</td>
<td>7</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>35.0</td>
<td>n.a.</td>
<td>10</td>
</tr>
<tr>
<td>Kim et al.</td>
<td>HBV</td>
<td>2012</td>
<td>128</td>
<td>13</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>27.8</td>
<td>0.722</td>
<td>19</td>
</tr>
<tr>
<td>Kim Do et al.</td>
<td>HBV*</td>
<td>2013</td>
<td>162</td>
<td>12</td>
<td>Asia</td>
<td>Longitudinal</td>
<td>24.0</td>
<td>0.736</td>
<td>12</td>
</tr>
<tr>
<td>Robic et al.</td>
<td>Mixed</td>
<td>2011</td>
<td>100</td>
<td>4</td>
<td>Europe</td>
<td>Longitudinal</td>
<td>24.0</td>
<td>0.837</td>
<td>21.1</td>
</tr>
<tr>
<td>Klibansky et al.</td>
<td>Mixed</td>
<td>2012</td>
<td>667</td>
<td>16</td>
<td>USA</td>
<td>Longitudinal</td>
<td>28.7</td>
<td>0.870</td>
<td>10.5</td>
</tr>
<tr>
<td>Poynard et al.</td>
<td>HCV</td>
<td>2014</td>
<td>3927</td>
<td>84</td>
<td>Europe</td>
<td>Longitudinal</td>
<td>144</td>
<td>0.860</td>
<td>50</td>
</tr>
</tbody>
</table>

*Liver cirrhosis with Child-Pugh class A. ^Mostly HCV with 85.4%. **41 patients with HCC at the time of enrollment, 10 patients developed HCC during follow-up. ¥All patients treated with interferon; 1All patients were HBeAg-negative; ¥All patients completed 2-year entecavir treatment.

TE, transient elastography; HCC, hepatocellular carcinoma; AUROC, area under the receiver operating characteristic curve; kPa, kilopascal; HCV, hepatitis C virus; HBV, hepatitis B virus; n.a., not available.
Clinical Practice Guidelines

Use of non-invasive tests for monitoring disease progression

Portal hypertension

There is substantial evidence indicating that TE can be quite effective in detecting patients with a high risk of having (or not having) developed clinically significant elevations of hepatic venous pressure gradient (HVPG) or varices. Several studies have shown that there is a good correlation between LS values and HVPG in patients with advanced liver diseases in both pre- and post-transplant settings [285–288]. According to a recent meta-analysis, the diagnostic performance of TE for predicting clinically significant PH (CSPH, HVPG > 10 mmHg) in the setting of patients with compensated chronic liver disease/cirrhosis is excellent, with an AUROC of 0.93 [289]; a 90% sensitive cut-off for CSPH diagnosis is 13.6 kPa, and a 90% specific cut-off in this setting is 21 kPa. These cut-offs have been shown to allow a correct stratification of presence/absence of CSPH in patients with compensated cirrhosis and potentially resectable HCC, thus reducing the need for invasive hemodynamic assessment [290]. However, while the correlation is excellent for HVPG values between 5 and 10–12 mmHg (typical of cirrhosis without evident clinical manifestations related to PH), it hardly reaches statistical significance for values above 12 mmHg [286]. This is because, with the progression of cirrhosis, the mechanisms of PH become less and less dependent on the intrahepatic resistance to portal flow due to tissue fibrosis and progressively more dependent on extra-hepatic factors (i.e. hyperdynamic circulation, splanchnic vasodilatation) [291]. This observation sets a key limitation to the use of LS measurements as a non-invasive surrogate of HVPG beyond the prediction of clinically significant (HVPG > 10 mmHg) and severe (HVPG > 12 mmHg) PH, and, accordingly, TE of the liver is unlikely to be useful in monitoring the hemodynamic response to the administration of beta-blockers or disease progression in the decompensated phase. Conversely, repeated LS measurements could be useful during the first year after liver transplantation to identify patients with hepatitis C recurrence characterized by a rapid progression towards cirrhosis [292]; in addition, a LS > 8.7 kPa one year after orthotopic liver transplantation is associated to a worse prognosis in this setting [293].

More uncertain and controversial is the possibility of predicting the presence and the size of oesophageal varices (OV) based on LS values. A correlation between LS values and the presence of OV has been reported in several studies [196,286–288,294–296] with AUROCs ranging from 0.74 to 0.85 and cut-offs from 13.9 kPa to 21.5 kPa. Although the sensitivity for the prediction of the presence of OV was high (76–95%), specificity was in general not satisfactory (43–78%). Regardless, the general features of these studies, i.e. single-centre retrospective, heterogeneous etiology of cirrhosis and stages of disease progression, subjective assessment of OV size, do not allow any sound conclusion on the utility of LS assessment in predicting the presence of OV and to screen cirrhotic patients without endoscopy [297].

Recently, studies employing different technical approaches have highlighted the potential utility of spleen stiffness (SS) assessment for predicting the presence of OV and the degree of PH in cirrhotic patients [298–302]. In particular, the study by Colecchia and co-workers [300] measured SS and LS by TE in 100 consecutive patients with HCV–induced cirrhosis. All patients also underwent measurements of HVPG and upper GI endoscopy. The ability of both SS and LS to predict clinically significant PH and the presence of OV was compared to that of the previously proposed methods, i.e. the LS–spleen diameter to platelet ratio score (LSPS) and platelet count to spleen diameter [303–305]. SS and LS were more accurate than other non-invasive parameters in identifying patients with OV and different degrees of PH. Further validation is needed before the place of SS in clinical practice can be defined.

Several biological parameters have been proposed for the non-invasive detection of CSPH including prothrombin time [287], a score combining platelet count and total bilirubin [306], and FibroTest® [307]. In particular, a score combining platelet count with total bilirubin had an AUROC of 0.91 for predicting CSPH with 8% sensitivity and 86% specificity at a cut-off of 1.0.

Similarly, several non-invasive tools have been proposed for the detection of OV including routine biological parameters [308], FibroTest® [309], and combination of simple biological and ultrasound parameters [303]. In the largest study to date comparing retrospectively a panel of serum markers (platelet count, AST/ALT ratio, APRI, Forns index, Lok index, FIB-4, and Fibroindex) in more than 500 patients with chronic liver diseases, the combination of Lok index (cut-off = 1.5) and Forns index (cut-off = 8.8) had the best diagnostic performance (AUROC of 0.80 and negative predictive value of 90%) for predicting clinically relevant OV [308].

In conclusion, the evidence accumulated so far indicates that both HVPG and upper GI endoscopy cannot be replaced by non-invasive methods, although an initial non-invasive approach may be helpful in selecting patients in whom these procedures are indicated with a certain level of urgency.

Hepatocellular carcinoma

To date, several cross-sectional studies [310–313] identified that high LS value measured by TE is significantly associated with the risk of presence of HCC (Table 8). However, these cross-sectional studies only describe the ‘static’ phenomenon that patients with HCC have high LS values than those without HCC, not considering the ‘dynamic’ association between the progression or regression of liver fibrosis and the risk of future HCC development. To overcome this limitation, several longitudinal prospective studies [267,314–323] have recently been published (Table 8).

A large prospective cohort study with chronic hepatitis C (866 patients) was conducted in Japan [317]. Along with age, male gender, and clinical cirrhosis, stratified LS values were identified as an independent risk factor for HCC development. Compared with patients with LS values ≤10 kPa, patients with higher LS values were at significantly increased risk of developing HCC (LS values, 10.1–15 kPa, hazard ratio [HR], 16.7; LS values, 15.1–20 kPa, HR, 20.9; LS values, 20.1–25 kPa, HR, 25.6; and LS values, >25 kPa, HR, 45.5). In addition, the cumulative incidence rates of HCC showed a stepwise increase according to the stratified LS values (p <0.001 by the log-rank test). In addition, Jung et al. [316] further validated the usefulness of TE in prediction of HCC development in patients with chronic hepatitis B (n = 1130). Compared with patients with LS values ≤8 kPa, patients with higher LS values were at significantly increased risk of developing HCC (LS values, 8.1–13 kPa, HR, 3.07; LS values, 13.1–18 kPa, HR, 4.68; LS values, 18.1–23 kPa, HR, 5.55; and LS values, >23 kPa, HR, 6.60). Furthermore, changes in the risk of HCC development according to the pattern of changes in the measured LS was also shown in the study, suggesting a potential role for serial LS measurement.
as a dynamic monitoring tool for risk estimation of HCC development in patients with chronic hepatitis B. All these results implied that TE is useful in estimating the risk of HCC development in patients with chronic liver disease across the etiology despite different carcinogenic mechanisms of HCV and HBV.

Based on the close relationship between TE and the risk of HCC development, several studies have tried to develop and validate LS-based prediction models for HCC development [310,320,324]. Wong et al. [324] evaluated the accuracy of LS-HCC score, refined from their previous CU-HCC score [325] in 1555 Asian patients with chronic hepatitis B. The AUROC of LS-HCC score was higher than that of CU-HCC score in predicting HCC development (0.83 vs. 0.75 at 3 year, 0.89 vs. 0.81 at 5 year). More recently, Kim et al. [320] also introduced a predictive model based on a Cox proportional hazards model using age, male gender, LS value, and HBV DNA in patients with chronic hepatitis B. This model showed good discrimination capability, with an AUROC of 0.806 (95% CI 0.738–0.874) and AUROC remained largely unchanged between iterations, with an average value of 0.802 (95% CI 0.791–0.812). The predicted risk of HCC occurrence calibrated well with the observed risk, with a correlation coefficient of 0.905 (p < 0.001).

According to all the results regarding non-invasive markers, it can be concluded that non-invasive methods are not merely an alternative to biopsy for staging fibrosis, but also predictive of the incidence of liver-related complications of liver fibrosis, including HCC development. However, further studies focusing on diverse etiologies of chronic liver disease, such as ALD or NAFLD, are needed to expand the clinical prognostic usefulness of non-invasive methods. In addition, optimal cut-off values with respect to the different etiologies of chronic liver disease to assess the risk of HCC development should be set up in subsequent larger longitudinal prospective studies. In spite of several limitations, non-invasive methods to assess and monitor the risk of HCC development will help physicians to establish optimum treatment strategies. It should be further investigated whether the accuracy of the surveillance strategy can be enhanced, if these non-invasive methods are incorporated into the routine surveillance strategy.

Recommendations

- Non-invasive tests cannot replace HVPG for a detailed PH evaluation and upper GI endoscopy for detecting varices (A1)
- However, in settings where HVPG is not available, TE could be considered to stratify the risk of CSPH (A2)
- Although TE could be useful to identify patients at risk of developing HCC, more data are needed before it can be integrated into an HCC surveillance program (A1)

Determining prognosis

There is increasing evidence for the prognostic value of non-invasive tests in patients with chronic liver diseases. Several recent studies have shown that in patients with chronic liver disease, LS could also predict clinical decompensation as well as survival [244,268,282,321,322,326,327,328]. For instance, Robic et al. [322] found that TE was as effective as HVPG in predicting clinical decompensations in 100 patients with chronic liver disease with a 2 year follow-up. Both HVPG >10 mmHg and LS >21.1 kPa had 100% NPV for portal-hypertensive complications. Similarly, in a cohort of 128 Korean patients with active HBV cirrhosis, LS at a cut-off of 19 kPa, had a hazard ratio of 7 for development of clinical decompensation [321]. In a cohort of 1457 HCV patients, LS values and FibroTest had the highest 5 year predictive values for predicting survival and liver-related death, which did not change after adjustment for treatment response, patient age, and estimates of necro-inflammatory grade [268]. Interestingly, Corpechot et al. [244] have shown in 168 patients with PSC that, not only those with high baseline but also those with increase in LS values (>1.5 kPa/year) were at a very high risk (approximately 10 times the risk estimated in the other group) of death, liver transplantation, or hepatic complications within a 4 year period. In another study in 1025 patients with chronic hepatitis C, the prognosis of patients with LS between 7 and 14 kPa on inclusion was significantly impaired when an increase >1 kPa/year was observed [263].

Finally, it has been recently suggested that SS could predict the occurrence of complications [329]. Thus the potential of LS values for predicting clinical outcomes seems to be greater than that of liver biopsy, probably LS measures ongoing pathophysiological processes and functions that a biopsy cannot.

Similarly, serum biomarkers such as FibroTest [154,234,330], ELF® [235,239], APRI and FIB-4 [222,331], as well as for models based on standard laboratory tests [332,333] have been shown to have prognostic value in various chronic liver diseases.

Recommendations

- There is increasing evidence for the prognostic value of non-invasive tests, particularly LS measurement using TE, in patients with cirrhosis (A1)
- Increase of LS values over time could be associated with a worse prognosis in patients with fibrosis or cirrhosis (A2)

Conflict of interest

Laurent Castera:
- Grant and research support: none
- Advisory Boards: none
- Speaking and teaching: AbbVie, Biopredictive, Bristol-Myers Squibb, Echosens, Gilead, Merck and Janssen

Henry Lik Yuen Chan:
- Grant and research support: unrestricted grant from Roche
- Advisory Boards: Abbvie, Bristol Myers Squibb, Gilead, Janssen, Merck, Novartis, Roche
- Speaking and teaching: Abbvie, Bristol Myers Squibb, Echosens, Gilead, Glaxo-Smith-Kline, Merck, Novartis, Roche

Please cite this article in press as: EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol (2015), http://dx.doi.org/10.1016/j.jhep.2015.04.006
Clinical Practice Guidelines

Marco Arrese:
- Grant and research support: Fondo Nacional de Desarrollo Cientifico y Tecnologico (FONDECYT 1110455) and the Comision Nacional de Investigacion Cientifica y Tecnologica (grant PFB 12/2007, Basal Centre for Excellence in Science and Technology) from the government of Chile
- Advisory Boards: none
- Speaking and teaching: none

Nezam Afshali:
- Grant and research support: Merck, Vertex, Gilead, AbbVie, BMS
- Advisory Boards: Merck, Gilead, Echosens, Glaxo Smith Kline, Ligand, Springbank, Kadmon, Jannsen, AbbVie, Achillion, Sandhill Scientific
- Speaking and teaching: none

Pierre Bedossa:
- Grant and research support: none
- Advisory Boards: none
- Speaking and teaching: Gilead, Genfit, Echosens, Intercept

Mireen Friedrich-Rust:
- Grant and research support: Echosens, Siemens Medical, Supersonic Imagine
- Advisory Boards: none
- Speaking and teaching: Echosens, Siemens Medical

Kwang-Hyub Han:
- Grant and research support: Medigen co., Eisai co. and KOWA co.
- Advisory Boards: Eisai co. and KOWA co.
- Speaking and teaching: nothing to disclose

Massimo Pinzani:
- Grant and research support: none
- Advisory Boards: Intercept, Abbvie, USB Cell Tech, Silence Therapeutics
- Speaking and teaching: Gilead, BMS, Jansen, Norgine, Core, MSD, Echosens

References


Clinical Practice Guidelines


Please cite this article in press as: EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol (2015), http://dx.doi.org/10.1016/j.jhep.2015.04.006
Clinical Practice Guidelines


Please cite this article in press as: EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol (2015), http://dx.doi.org/10.1016/j.jhep.2015.04.006