Non-invasive assessment of liver fibrosis

Vasilios Papastergiou, Emmanuel Tschochatzis, Andrew K. Burroughs
Royal Free Hospital and UCL, London, UK

Abstract

The presence and degree of hepatic fibrosis is crucial in order to make therapeutic decisions and predict clinical outcomes. Currently, the place of liver biopsy as the standard of reference for assessing liver fibrosis has been challenged by the increasing awareness of a number of drawbacks related to its use (invasiveness, sampling error, inter-/intraobserver variability). In parallel with this, noninvasive assessment of liver fibrosis has experienced explosive growth in recent years and a wide spectrum of noninvasive methods ranging from serum assays to imaging techniques have been developed. Some are validated methods, such as the Fibrotest/Fibrosure and transient elastography in Europe, and are gaining a growing role in routine clinical practice, especially in chronic hepatitis C. Large-scale validation is awaited in the setting of other chronic liver diseases. However, noninvasive tests used to detect significant fibrosis and cirrhosis, the two major clinical endpoints, are not yet at a level of performance suitable for routine diagnostic tests, and there is still no perfect surrogate or method able to completely replace an optimal liver biopsy. This article aims to review current noninvasive tests for the assessment of liver fibrosis and the perspectives for their rational use in clinical practice.

Keywords liver fibrosis, non-invasive, transient elastography, serum biomarkers, liver biopsy

Ann Gastroenterol 2012; 25 (2): 218-231

Introduction

Detection and quantification of hepatic fibrosis represents a longstanding challenge in Hepatology. Currently, accurate assessment of liver fibrosis has become increasingly important in order to make therapeutic decisions, determine prognosis and to follow-up disease progression. Hepatic fibrogenesis is a dynamic process reflecting an imbalanced extracellular matrix turnover. Evaluating the evolution of fibrosis over time can be therefore more important than a “once only” diagnosis. Recent evidence suggesting that liver fibrosis can be reversible [1,2], further emphasizes the importance to monitor fibrosis over time, other than simply diagnosing its presence and staging its severity. Accurate, reproducible and easily applied methods are therefore required for the assessment of hepatic fibrosis. The present article reviews the noninvasive means for assessing liver fibrosis and the existing evidence of their clinical validation in different settings of chronic liver disease. Clinical needs and realistic endpoints are taken into account. Furthermore, critical issues regarding evaluation of performance and validation of any new surrogates as well as future perspectives for their clinical applicability are discussed.

Liver biopsy: a debated gold standard

Historically, liver biopsy has been considered to be the gold standard for assessing the degree of liver fibrosis. Histology of the biopsy specimen allows clinicians to obtain diagnostic information not only on fibrosis but also on inflammation, necrosis, steatosis and hepatic deposits of iron or copper. However, increasing awareness of several drawbacks of liver biopsy has repeatedly questioned its accuracy and value in clinical practice. Clearly, liver biopsy samples are an extremely small (1/50000) part of the liver and therefore sampling error can occur even with optimal specimens [3,4]. The optimal size for staging fibrosis is 20 mm of length with 11 portal tracts [5], but this might need more than one pass with a biopsy needle to be achieved [6], thus increasing the likelihood of adverse events. Moreover, despite the availability of widely-validated standardized fibrosis scoring systems, accuracy of histological examination can be inherently compromised by a significant intraobserver and interobserver variability [7,8]; even an optimal (25 mm long) biopsy specimen has a 25% rate of discordance for fibrosis staging [3] which is further compounded by operator-related expertise and
specialization of the pathologist [9]. Last but not least, liver biopsy is an invasive procedure with a small but significant risk of procedure-related morbidity (pain occurring in 20% and major complications such as intraperitoneal bleeding and hemobilia in 0.5%) [4], with a mortality rate of 0.009-0.12% [10]. Thus, liver biopsy can be poorly tolerated by patients, especially if it needs to be repeated. Transjugular liver biopsy, which is safer and better tolerated, is only available in specialized centers, but it does allow several passes without added complications, and the measurement of hepatic venous pressure gradient (HVPG) [11,12]. Optimal technique for such a biopsy has been published [13]. Due to these limitations, consideration of liver biopsy as the "gold standard" has declined to "best available" standard [14], and has been challenged by the recent increasing availability and validation of noninvasive methods to assess liver fibrosis.

### Non-invasive methods to assess liver fibrosis

The large number of studies evaluating methods to detect and quantify fibrosis is shown in Table 1; these can be schematically divided in two main types: serum markers and imaging modalities.

#### Serum markers

Development of serum markers is in constant evolution offering an attractive alternative to liver biopsy for both patients and physicians. These markers are classified as direct (or class I) which represent extracellular matrix components (reflecting the pathophysiology of liver fibrogenesis); and indirect (or class II) which use routine laboratory data (reflecting the consequences of the liver damage). Direct and indirect markers may be used alone or, more commonly, in combination to produce composite scores.

#### Individual surrogate markers of liver fibrosis

Hepatic stellate cells (HSC) represent the dominant profibrogenic hepatic cell type. Activation and subsequent transformation in fibrogenic and contractile myofibroblasts is

---

**Table 1** Overview of non-invasive methods for the evaluation of liver fibrosis Q1. Each Table should include all abbreviations at the bottom.

<table>
<thead>
<tr>
<th>a. Direct serum markers/panels</th>
<th>b. Indirect serum markers/panels</th>
<th>c. Patented serum panels</th>
<th>d. Imaging modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyaluronate [16]</td>
<td>AST/ALT ratio [28]</td>
<td>Fibrotest® [58]</td>
<td>Transient elastography (Fibroscan®) [84]</td>
</tr>
<tr>
<td>Procollagen type III amino-terminal peptide (PIIINP) [20]</td>
<td>FIB-4 [47]</td>
<td>Enhanced Liver Fibrosis score (ELF®) [76]</td>
<td></td>
</tr>
<tr>
<td>Metalloproteinases (MMP)-1 and MMP-2 [21,22]</td>
<td>Lok index [52]</td>
<td>Fibrometers® [71]</td>
<td></td>
</tr>
<tr>
<td>Tissue inhibitors of the metalloproteinases (TIMPs) [21]</td>
<td>Fibrosis Probability Index (FPI) [36]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transforming growth factor-β1 (TGF-β1) [23]</td>
<td>Goteborg University Cirrhosis Index (GUCI) [37]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP3 [44]</td>
<td>Virahep-C model [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microfibril-associated glycoprotein 4 (MFAP-4) [24]</td>
<td>SHASTA index [39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BAAT [134]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAFLD fibrosis score [136]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BARD [139]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:**

- AST, aspartate transaminase; ALT, alanine transaminase; PGA, prothrombin time; gamma-glutamyl transpeptidase, apolipoprotein A1; APRI, aspartate transaminase to platelet ratio index; BAAT, body mass index, age, alanine transaminase, triglycerides; NAFLD, nonalcoholic fatty liver disease; MR, magnetic resonance; CT, computed tomography
the key event leading to extracellular matrix (ECM) deposition and increased intrahepatic resistance to blood flow (portal hypertension). A variety of “direct” serum markers reflecting ECM turnover (fibrogenesis and fibrinolysis) and/or fibrogenic cell changes have been developed, and used clinically (15-24) (Table 1a). Among them, hyaluronate is the best validated single marker that most accurately predicts advanced fibrosis both in chronic hepatitis C (CHC) (15,16), and other liver diseases (25,26). Given its high negative predictive value (NPV) (98-100%) it could be used on its own in clinical practice for the exclusion of advanced fibrosis [27].

Individual “indirect” serum markers include simple routine blood tests (Table 1b). Aspartate aminotransferase (AST) to alanine aminotransferase (ALT) ratio index (APRI) ≥ 1 has shown good specificity (although relatively insensitive) to detect cirrhosis in patients with CHC with reported positive predictive values (PPV) and NPV ranging from 73.7-100% and 46.7-53.2% respectively [28,29]. However, its usefulness was not confirmed when validation was assessed in independent patient cohorts [30]. Its use for diagnosing cirrhosis in primary biliary cirrhosis [31] and primary sclerosing cholangitis [32] demonstrated a poor clinical outcome of patients with cirrhosis and high AST/ALT ratio [32,33]; there was an estimated 5% (95%CI: 1-8%) increase in hazard of dying per 0.10 increase in AST/ALT ratio in patients with non-alcoholic cirrhosis [33]. Simple fibrosis markers based on routine blood tests also include prothrombin index [34] and platelet count [35].

Indices combining indirect and direct markers of liver fibrosis

Due to the poor accuracy of individual markers to assess liver fibrosis, algorithms or indices combining panels of markers have been developed and widely validated, with reportedly “sufficient” diagnostic accuracy. Some panels are protected by patents and are commercially available with proprietary bundle assays, whereas others are freely available [36-39]. The ALT to platelets ratio index (APRI) is calculated as (AST/upper limit of normal range)/platelet count (10^9/L) x 100 [40]. APRI and Fibrotest are the most extensively studied serum markers. A meta-analysis from 2007 [41] showed that with a cut-off value of 0.5, APRI had 81% sensitivity, but only 50% specificity in predicting significant fibrosis (Metavir ≥ F2); with a cut-off value of 1, the sensitivity and specificity for predicting cirrhosis were 76% and 71% respectively. In a meta-analysis comprising over 8,700 patients [42], the summary of areas under receiver operating characteristic (AUROC) values of APRI for significant fibrosis (F2 or more), severe fibrosis (F3-F4) and cirrhosis (F4) were 0.77, 0.80 and 0.83 respectively. The sensitivity and specificity values for F2 fibrosis or more of an APRI threshold of 0.7 were 77% and 72% respectively, and 61% and 64% when a threshold of 1.0 was used. For cirrhosis, the sensitivity and specificity of an APRI threshold of 1.0 were 76% and 72%. The above data show only a moderate degree of accuracy of APRI for diagnosing CHC-related fibrosis, which is not sufficiently good for a routine diagnostic test.

The Forns index [based on 4 routine variables: age, platelet count, cholesterol and γ-glutamyl-transferase (γ-GT)] has been assessed [43] and later validated in patients with CHC [44,45] and non-hepatitis C liver diseases [45] as well as in human immunodeficiency virus (HIV)/hepatitis C virus (HCV) co infected patients [46]. Using two cut-off values [a lower (4.2) to exclude, and a higher (6.9) to confirm ≥F2 fibrosis], the index showed a good diagnostic performance (AUROC: 0.81-0.86) in CHC patients, with the lower cut-off having 96% NPV to exclude F2 or more fibrosis [43]. Lack of information regarding cirrhosis and a significant rate of unclassified cases are the main limitations. The FIB-4 score combines platelet count, ALT, AST and age and was initially developed for use in HCV/HIV co infection where it correctly classified 87%, and avoided biopsy in 71% of the validation set, with an AUROC of 0.765, sensitivity 70% and specificity 97% for discriminating Ishak 0-3 from 4-6 [47]. In a subsequent analysis including a large cohort of HCV-monoinfected patients, FIB-4 enabled good discrimination of both severe fibrosis (AUROC 0.85) and cirrhosis (AUROC 0.91) [48]. More recently this marker has been assessed in patients with chronic hepatitis B (CHB) [49,50] with reported 71% sensitivity and 73% specificity for diagnosing ≥F2 fibrosis [49]. Moreover, it has been shown to be reliable in the setting of non-alcoholic fatty liver disease (NAFLD); using a cut-off value of 1.3 the sensitivity and specificity for predicting advanced (F3-F4) fibrosis were 74-85% and 65-71% respectively, and 34% and 98% when a FIB-4 threshold of 2.67 was used [51]. The Lok index is an evolution of the APRI combining platelet count, INR and AST/ALT ratio. This index uses two cut-off values: 0.2 to rule out cirrhosis and 0.5 to confirm cirrhosis, whereas values between these cut-offs are considered indeterminate. In a cohort study of 1,141 patients with CHC, the first report documented an AUROC of 0.78-0.81 to detect cirrhosis, concluding that the model might obviate a liver biopsy in 50% of cases [52]. A similar performance was observed in a later study but without a clear advantage in comparison to the APRI [53]. This index could be used for screening for cirrhosis but the laboratory differences in measuring INR [54] may reduce its applicability.

Patented serum marker panels/models

The Fibrotest® (Fibrosure® in the USA) is the most widely validated indirect serum marker panel, extensively studied in CHC [44,55-58] but also in CHB [59,60], HCV/HIV co infection [61] and NAFLD [62,63]. It is computed using five parameters, namely total bilirubin, haptoglobin, gamma-glutamyl-transpeptidase, a2-macroglobulin and apolipoprotein-A. In a systematic review including 9 studies (1679 patients), an excellent discrimination was found for identifying cirrhosis (summary AUROC=0.90), but a lesser ability to identify significant (≥F2) fibrosis (AUROC=0.81). However, the conclusion was that noninvasive tests are not ready to replace liver biopsy [64]. In a large metaanalysis performed by the group that invented the test and included 6378 subjects, the mean standardized AUROC for diagnosing
significant (≥F2) fibrosis was 0.84 (95%CI:0.83-0.86) without differences between different etiologies of chronic liver disease [65]. Interestingly, in a later study including 2411 patients, performance of Fibrotest was good in all etiologies for both ≥F2 and F4 detection (standardized AUROC=0.73), except for ≥F2 in NAFLD (standardized AUROC=0.64) [62]. The Fibroindex® was developed for patients with CHC and uses platelet count, AST and IgG levels. It showed good diagnostic accuracy and high positive predictive values for significant fibrosis. Changes in the Fibroindex correlated significantly with variations in fibrosis stage before and after administration of antiviral therapy [66]. Importantly, performance of Fibroindex, AST/ALT ratio and Forns, were poor when evaluated in patients with CHC but normal transaminase levels [67]. The Hepascore®, a score of 6 individual markers ranging from 0.00 to 1.00 [68], showed good discrimination (AUROC of 0.81 for significant fibrosis and 0.88 for cirrhosis) when assessed in a cohort of 512 chronically HCV-infected patients [69]. Fibrometers® are a family of 6 blood tests: one for staging and one for quantifying liver fibrosis in each of the 3 main causes of liver disease [chronic viral hepatitis, alcoholic liver disease (ALD) and NAFLD]. The Tests aim to relate to a morphometric quantitation of the fibrotic area, and the results are validated through an expert system that detects erroneous results [70,71]. Reported AUROC ranges from 0.85-0.89 for ≥F2 fibrosis and 0.91 for cirrhosis in patients with CHC [44,71], although validation in large and independent cohorts is still lacking. Fibropect® includes hyaluronate, TIMP-1 and α-2-macroglobulin and is validated in CHC. The AUROC for significant fibrosis was 0.82 in the initial study [72] and ranged from 0.82 to 0.87 in subsequent studies, showing 71.8-93% sensitivity, 66-73.9% specificity and an overall test accuracy ranging from 73.1 to 80.2% [73-75]. Again this falls short of sufficient diagnostic accuracy. Enhanced Liver Fibrosis (ELF®) is a panel of direct noninvasive markers that includes age, hyaluronan, type III collagen and TIMP-1. In a cohort of more than 1000 patients with chronic liver disease, the algorithm detected ≥F2 fibrosis (sensitivity, 90%) and accurately detected the absence of fibrosis (NPV for F2 or more fibrosis 92%), suggesting that it could be used to screen a range of chronic liver diseases [76].

Limitations

Although noninvasive, easy to repeat and highly applicable, serum markers have obvious limitations. Their main disadvantage is represented by their low accuracy to detect intermediate stages of fibrosis as compared to cirrhosis [77,78]. Another drawback is the potential lack of liver-specificity: serum levels of direct markers such as hyaluronate can be affected by renal and/or liver failure, extrahepatic sites of fibrogenesis or postprandial state [79]. Similarly, interpretation of the Fibrotest results should be done critically taking into account that hemolysis (decrease in haptoglobin), Gilbert's syndrome (increase in bilirubin) or inflammatory states (increase in α2-macroglobulin) can be associated with erroneous results. Furthermore, inter-laboratory reproducibility of some parameters such as transaminase levels, INR or platelet count is questionable. Patented [80] and more complex [63] serum panels have shown slightly better accuracy than simple ones (APRI, FIB-4, Forns), and acceptable inter-laboratory reproducibility for clinical practice [70,81]. On the other hand, simple serum markers are cost-free, easy to calculate and widely available almost everywhere. The methodological quality of the studies assessing diagnostic accuracy of noninvasive serum markers represents a further issue. Several aspects are at fault, related to aspects of the study's design, methods of sample recruitment, the execution of the tests, and the completeness of the study report, as outlined by the QUADAS tool [82]. This is further discussed later.

Imaging techniques

Classical imaging techniques, including ultrasonography (U/S), computed tomography (CT) and magnetic resonance imaging (MRI) are used in clinical practice for the detection of advanced liver disease (mainly cirrhosis) either directly (by detecting overt morphological changes of the cirrhotic liver) or indirectly by detecting signs of portal hypertension (enlarged spleen, collateral vein circulation, etc.). Interestingly, in a recent paper [83], the U/S evaluation of the liver surface was shown to be highly accurate for diagnosing clinically doubtful cirrhosis. There is room for further development in this area. However, the necessity to accurately identify lesser degrees of liver fibrosis has led to the development of novel imaging modalities which primarily aim to assess liver fibrosis [84-87]. All equipment is expensive and this is a major drawback, unless the technology is incorporated into current imaging modalities.

Transient elastography (TE)

Transient Elastography (TE) (Fibroscan®, Echosens®, Paris, France) is probably the most widely used noninvasive method in Europe, both for CHC [55,88-90] and other liver diseases [91-94]. Briefly, vibrations of mild amplitude and low frequency are transmitted by the transducer inducing an elastic shear wave that propagates within the liver. Pulse-echo ultrasonic acquisitions are performed to follow the shear wave and measure its speed, which is directly related to the tissue stiffness (the harder the tissue, the faster the shear propagates). Results are expressed in Kilopascals (Kpa) and correspond to the median value of ten validated measurements ranging from 2.5 to 75 Kpa [95], with 5.5 Kpa reported to define normality [96]. The volume of liver tissue evaluated by TE approximates a cylinder 4x1 cm which is at least 100 times bigger than a liver biopsy. Moreover, TE is painless and rapid (<5 min) and thus highly acceptable for patients. The diagnostic performance of TE has been...
widely addressed [64,97-99]. Diagnostic accuracy of TE for cirrhosis is good with pooled estimated sensitivity and specificity approaching 90%, whereas for the detection of significant fibrosis the diagnostic value of TE is substantially lower with pooled estimates of sensitivity and specificity ranging to 70-80% [97-99]. In a recent meta-analysis on TE, we included 40 studies after application of strict inclusion criteria (use of liver biopsy as the reference standard, time interval between TE and liver biopsy <3 months, and adequate methodological quality) and used pre-test and post-test probabilities as a more practical and informative way to report the results [97]. We concluded that TE could be used as a good screening test for cirrhosis, with 90% probability of correctly diagnosing cirrhosis (when pre-test probability is 50%) but only 78-85% of correctly diagnosing lesser stages of fibrosis following a "positive" measurement. Interestingly, a "negative" measurement (i.e. excluding the diagnosis) was found to be less informative as cirrhosis might still be present in 16%, or ≥F2 fibrosis in 20% of patients. For a higher pre-test index of suspicion (pre-test probability=75%), the probability of a correct diagnosis following a "positive" measurement exceeded 90% for all fibrosis stages, but with a "negative" measurement diagnosis was still wrong for 24-44% of patients. Despite this reduced accuracy for the diagnosis of intermediate fibrosis stages, TE appears to be more accurate than serum biomarkers in predicting cirrhosis, and the most accurate method for early detection of cirrhosis [80]. However, reported diagnostic threshold (cut-off) values are subject of wide inter-study variability, and there are overlaps between and within stages, varying from 4 to 10.1 Kpa for the diagnosis of significant fibrosis, and 9 to 26.5 Kpa for cirrhosis [97]. Use of ranges of values rather than single cut-offs is probably more practical for the interpretation of the results [95] but this requires an international consensus. Despite the wide acceptance and incorporation in clinical practice, this method is not free of limitations. Acute liver injury [100-102], congestive heart failure [103] as well as a postprandial measurement can be associated with an overestimation of the measured liver stiffness. The degree of liver inflammation also affects TE values [104] limiting the validity of assessing regression of fibrosis, as inflammation also improves. Moreover, failure to obtain interpretable measurements represents a further problem in the use of TE. According to the recommendations of the manufacturer, the following parameters should be achieved in order for the results to be reliable: 1) number of valid shots ≥10; 2) the interquartile range of measurements should not exceed 30% of the median value; and 3) successful measurements should represent at least 60% of the total number of acquisitions [95]. However, according to the results of a recent prospective study including 19 French centers, results of liver stiffness measurements were uninterpretable in up to 22% of patients, thus significantly affecting diagnostic performance of TE [105]. TE is difficult in patients who are obese or who have narrow intercostal spaces and impossible in patients with ascites. Indeed, in a 5-year study based on more than 13000 examinations, increased waist circumference and limited operator experience were found to be the most important determinants of TE failure and unreliable results [106]. Recently, a novel Fibroscan probe (XL probe) has been shown to increase diagnostic accuracy in obese patients [107].

**Acoustic radiation force impulses (ARFI)**

Use of ARFI has been recently suggested to be a valid method to assess liver fibrosis. This imaging technology permits evaluation of liver stiffness in a region of interest (ROI) involving mechanical excitation of tissue by the use of short-duration (~262 μs) acoustic pulses while performing a real-time B-mode conventional hepatic U/S. Results are expressed in (m/s). Although the volume of liver explored is smaller than that of TE (10 mm long x 6 mm wide) a critical advantage is the possibility to choose the representative area of interest avoiding large vessels and ribs. Diagnostic accuracy of ARFI is good and comparable to TE for cirrhosis [85,108-111] with an AUROC of 0.89 (vs. 0.80 for TE, P=0.09) according to a recent study [112]. In this same comparative study including 139 patients with CHC, ARFI was more accurate than TE for the noninvasive staging of both significant (≥F2) (AUROC: 0.86 vs. 0.78, P=0.024) and severe (F3-F4) fibrosis (AUROC: 0.94 vs. 0.83, P=0.02). Importantly, no cases of invalid measurements were recorded vs. 6.5% of unreliable results in patients undergoing TE (P=0.029). Moreover, in contrast to TE, liver steatosis does not seem to influence ARFI [113]. Another advantage is that it can be easily incorporated into a modified U/S machine. Further validation including inter- and intra-operator reproducibility is required before ARFI can be used in routine clinical practice.

**Magnetic resonance (MR) elastography**

Application of elastography to MRI uses a modified phase-contrast method to evaluate the propagation of the shear waves within the liver. Advantages include the potential to analyze the whole parenchyma, as well as the applicability for patients with obesity or ascites. In a study including 96 patients with CHC, discriminative performance of MR-elastography was significantly better than that of TE for the detection of ≥F2 fibrosis (AUROC: 0.99 vs. 0.84, p<0.05) [86]. Limitations are represented by the high costs and the fact it is too time-consuming for widespread implementation in clinical practice.

**Sequential algorithms**

Recently, research has focused on the sequential use of two or more methods in order to increase diagnostic accuracy [44,114,115]. Castera et al showed that the combination of TE and Fibrotest resulted in excellent diagnostic accuracy for detecting both significant fibrosis (AUROC 0.88) and cirrhosis (AUROC 0.95) [55]. Sebastiani et al proposed a
performing well in diagnosing cirrhosis but only moderately been specifically tested in this setting (Table 3) with most syndrome [130-132]. A series of noninvasive methods have is considered to be the hepatic component of the metabolic pattern [129] (i.e., frequent perisinusoidal and perivenular patients) may be related to the different fibrosis distribution The higher cut-off values obtained in ALD (compared to HCV accuracy of TE in distinguishing different stages of fibrosis. Studies including 103 [127] and 174 [128] patients showed high for cirrhosis and significant fibrosis respectively) correctly classified the majority of 217 CHB patients independently of ALT values [123]. Further studies in larger and independent cohorts are needed to optimize the use of TE in CHB.

Several noninvasive markers have been used to assess liver fibrosis in patients with ALD [124-126] (Table 3). Two studies including 103 [127] and 174 [128] patients showed high accuracy of TE in distinguishing different stages of fibrosis. The higher cut-off values obtained in ALD (compared to HCV patients) may be related to the different fibrosis distribution pattern [129] (i.e., frequent perisinusoidal and perivenular fibrosis), which highlights the need for disease-specific cut-off values. NAFLD is now a major cause of liver disease as it is considered to be the hepatic component of the metabolic syndrome [130-132]. A series of noninvasive methods have been specifically tested in this setting (Table 3) with most performing well in diagnosing cirrhosis but only moderately well for detecting intermediate stages of fibrosis [51,133-142]. Although simple noninvasive markers have been shown to be reliable in NAFLD (with FIB-4 being the most accurate) [51], a recent study on 242 NAFLD patients showed better accuracy of complex models (Hepascore, Fibrotest, FIB-4) as compared to simple ones (APRI, BARD) for the detection of advanced fibrosis [63]. Validation of ELF®, in an independent cohort of 196 NAFLD-patients revealed good accuracy with AUROC of 0.90, 0.82 and 0.76 for severe, moderate and absent fibrosis respectively [135]. Recently, TE proved to be accurate in the discrimination of advanced NAFLD-fibrosis, showing small biopsy size rather than TE to account for most cases of discordance between TE and liver biopsy [143]. Interest is steadily increasing for using noninvasive markers in the setting of cholestatic liver diseases as variables included (e.g.; γ-GT, cholesterol, apolipoprotein-A1) are largely influenced by cholestasis.

Use of noninvasive markers in different clinical settings

Most of the data mentioned so far regards patients with CHC in whom several noninvasive markers have been validated and used in clinical practice (Table 2).

Regarding CHB, the number of studies is smaller. In a comparative study including 110 CHB patients, Fibrotest and APRI achieved the highest diagnostic accuracy for detecting significant fibrosis whereas their combination in a single algorithm yielded an accuracy of 93% [118]. In a study where FIB-4 was used to diagnose mild fibrosis in CHB, a cut-off ≤1.45 to differentiate moderate from severe fibrosis had a sensitivity of 71.1% and specificity of 73.1% [49]. TE has also shown good ability to identify HBV-related fibrosis. Probably because of the frequent macronodular pattern of cirrhosis arising from HBV (associated with lesser amount of fibrosis), cut-offs used in CHB are generally lower than those in CHC, ranging from 7.2-7.5Kpa for ≥F2 fibrosis and from 9-13.4 Kpa for cirrhosis [91,119-121]. Combination of Forns with TE increases diagnostic accuracy in CHB [122]. Importantly, the influence of increases in transaminases levels should be considered when interpreting results of TE [100]. Interestingly, a dual cut-off algorithm (>13.1 Kpa positive, ≤9.4 Kpa negative and >9.4Kpa positive, ≤6.2 Kpa negative for cirrhosis and significant fibrosis respectively) correctly classified the majority of 217 CHB patients independently of ALT values [123]. Further studies in larger and independent cohorts are needed to optimize the use of TE in CHB.

Several noninvasive markers have been used to assess liver fibrosis in patients with ALD [124-126] (Table 3). Two studies including 103 [127] and 174 [128] patients showed high accuracy of TE in distinguishing different stages of fibrosis. The higher cut-off values obtained in ALD (compared to HCV patients) may be related to the different fibrosis distribution pattern [129] (i.e., frequent perisinusoidal and perivenular fibrosis), which highlights the need for disease-specific cut-off values. NAFLD is now a major cause of liver disease as it is considered to be the hepatic component of the metabolic syndrome [130-132]. A series of noninvasive methods have been specifically tested in this setting (Table 3) with most performing well in diagnosing cirrhosis but only moderately accurate. Studies including 103 [127] and 174 [128] patients showed high accuracy of TE in distinguishing different stages of fibrosis. The higher cut-off values obtained in ALD (compared to HCV patients) may be related to the different fibrosis distribution pattern [129] (i.e., frequent perisinusoidal and perivenular fibrosis), which highlights the need for disease-specific cut-off values. NAFLD is now a major cause of liver disease as it is considered to be the hepatic component of the metabolic syndrome [130-132]. A series of noninvasive methods have been specifically tested in this setting (Table 3) with most performing well in diagnosing cirrhosis but only moderately

Repeated measurements of liver fibrosis and prediction of clinical outcomes

A major advantage of noninvasive markers is given by their easy reproducibility over time. Longitudinal assessment of noninvasive markers could allow clinicians not only to monitor disease progression but also to determine the effect on liver fibrosis of antiviral therapy, reduction in alcohol intake (in ALD), or weight loss (in NAFLD), with virtually no cost in terms of safety and patient acceptance. Although data is limited and mainly restricted to patients with viral hepatitis (where there is a specific therapy and a

Annals of Gastroenterology 25
<table>
<thead>
<tr>
<th>Method</th>
<th>Included serum markers</th>
<th>Etiology</th>
<th>AUROC</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥F2</td>
<td>F4</td>
<td>≥F2</td>
<td>F4</td>
</tr>
<tr>
<td>APRI [40]</td>
<td>AST, platelets</td>
<td>HCV</td>
<td>0.69-0.88</td>
<td>61-0.94</td>
<td>57-89</td>
<td>61-88</td>
<td>38-57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBV</td>
<td>0.69-0.72</td>
<td>64-0.76</td>
<td>42.9</td>
<td>63-86</td>
<td>53.8</td>
</tr>
<tr>
<td>Fibrotest® [58]</td>
<td>γ-GT, haptoglobin, bilirubin, A1 apolipoprotein, a2-macroglobulin</td>
<td>HCV</td>
<td>0.74-0.89</td>
<td>0.71-0.92</td>
<td>50-87</td>
<td>70-80</td>
<td>57.9-93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBV</td>
<td>0.77-0.85</td>
<td>0.76-0.87</td>
<td>55.6</td>
<td>90</td>
<td>38</td>
</tr>
<tr>
<td>Forns Index [43]</td>
<td>Age, γ-GT, cholesterol, platelets</td>
<td>HCV</td>
<td>0.60-0.86</td>
<td>0.63-0.72</td>
<td>NA</td>
<td>61.2-95</td>
<td>35.9-53.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBV</td>
<td>0.63-0.72</td>
<td>0.81</td>
<td>78.3</td>
<td>66-94.7</td>
<td>44-90.5</td>
</tr>
<tr>
<td>FIB-4 [47]</td>
<td>Age, ALT, AST, platelets</td>
<td>HCV</td>
<td>0.82-0.89</td>
<td>0.79-0.91</td>
<td>37.6-74.3</td>
<td>80.1-98.2</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBV</td>
<td>0.74-0.86</td>
<td>0.80-0.93</td>
<td>71</td>
<td>81</td>
<td>38</td>
</tr>
<tr>
<td>Fibrometer® [137]</td>
<td>Platelets, prothrombin time, macroglobulin, AST, hyaluronate, age, urea</td>
<td>HCV</td>
<td>0.78-0.89</td>
<td>0.91</td>
<td>80.5-89</td>
<td>84.1-89.9</td>
<td>77.6-82.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBV</td>
<td>0.74</td>
<td>0.89</td>
<td>94.1</td>
<td>87.6</td>
<td>85.4</td>
</tr>
<tr>
<td>Hepascore® [68]</td>
<td>Age, sex, a2-macroglobulin, hyaluronate, bilirubin, γ-GT</td>
<td>HCV</td>
<td>0.62</td>
<td>0.85-0.94</td>
<td>53.8-82</td>
<td>65-92</td>
<td>70-88</td>
</tr>
<tr>
<td>Enhanced Liver Fibrosis Score (ELF®) [76]</td>
<td>N-terminal propeptide of collagen type III, hyaluronate, TIMP-1, age</td>
<td>HCV/HBV</td>
<td>0.77-0.87</td>
<td>0.87-0.90</td>
<td>90</td>
<td>84-89.8</td>
<td>70-88</td>
</tr>
<tr>
<td>Fibroindex® [66]</td>
<td>Platelet count, AST and γ-globulin</td>
<td>HCV</td>
<td>0.86</td>
<td>NA</td>
<td>36</td>
<td>NA</td>
<td>27.5</td>
</tr>
<tr>
<td>Fibrospect® [72]</td>
<td>a-2-macroglobulin, hyaluronate and TIMP-1</td>
<td>HCV</td>
<td>0.82-0.97</td>
<td>NA</td>
<td>71-83</td>
<td>66-73.9</td>
<td>60.9-82.6</td>
</tr>
<tr>
<td>MP3 [44]</td>
<td>MMP-3 and TIMP-1</td>
<td>HCV</td>
<td>0.82</td>
<td>NA</td>
<td>65</td>
<td>NA</td>
<td>89</td>
</tr>
<tr>
<td>Fibrosis Probability Index (FPI) [36]</td>
<td>Age, past alcohol intake, AST, cholesterol and HOMA-IR</td>
<td>HCV</td>
<td>0.77</td>
<td>NA</td>
<td>42-85</td>
<td>48-98</td>
<td>87</td>
</tr>
<tr>
<td>Goteborg University Cirrhosis Index (GUCI) [37]</td>
<td>AST, INR and platelet count</td>
<td>HCV</td>
<td>NA</td>
<td>0.85</td>
<td>NA</td>
<td>80</td>
<td>NA</td>
</tr>
<tr>
<td>Lok Index [52]</td>
<td>Platelet count, AST/ALT ratio and INR</td>
<td>HCV</td>
<td>NA</td>
<td>0.78-0.81</td>
<td>NA</td>
<td>37-92</td>
<td>NA</td>
</tr>
<tr>
<td>Vira-Hep C [38]</td>
<td>AST, platelet count, alkaline phosphatase and age</td>
<td>HCV</td>
<td>0.83</td>
<td>NA</td>
<td>51-90</td>
<td>54-90</td>
<td>53-75</td>
</tr>
</tbody>
</table>

*reported values are indicative and based on published studies including more than 100 patients
NA, Not available; AUROC, area under receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; AST, aspartate aminotransferase; γ-GT, gamma glutamyl transferase; ALT, alanine aminotransferase; TIMP, tissue inhibitors of matrix metallopeptinases; MMP, metalloproteinases; HOMA-IR, homeostasis model assessment of insulin resistance; INR, international normalized ratio.
clear definition of response) use of noninvasive surrogates can detect improvements in patients successfully treated for CHC [159]. Indeed, both serum markers (including Forn's, Fibrotest, FIB-4, Fibrometer and APRI) and TE measurements have been shown to decrease significantly after viral clearance, correlating with a decrease in the fibrosis stage, an outcome not observed in nonresponders [160-163]. High rates of improvement and even regression of liver fibrosis following successful antiviral treatment [2] are questioning the need for a follow-up liver biopsy. Longitudinal measurements using noninvasive methods could allow an effective monitoring of the evolution of fibrosis, with each patient serving as his/her own control, as the initial measurement, will be the comparator for worsening or improvement, as shown in the liver transplant setting [158].

The potential ability of noninvasive methods to identify patients at risk of disease progression and therefore to "predict" clinical events, is another area recently explored. In a study including 165 patients with cirrhosis, liver stiffness values <19 Kpa were highly predictive of the absence of large varices [164], whereas in a follow-up study including 866 HCV-positive patients the 3-year cumulative probability of developing HCC was 0.4% in patients with TE values <10 Kpa vs. 38% in those with liver stiffness measurements >25 Kpa [165]. Fontana et al showed, using the population of the HALT-C trial, that baseline hyaluronate, TIMP-1, PIINP and YKL-40 levels combined with simple laboratory parameters (including baseline bilirubin, albumin and INR) were all associated with clinical outcomes in patients with disease progression (P<0.0001) [166]. In the same study, baseline hyaluronate and platelet count were best at predicting histological progression (AUROC=0.663). In a large multicenter study including 600 cirrhotics, combination of the Lok Index and Forn's score was efficient to exclude (NPV >90%) the presence of clinically relevant esophageal varices (defined as large varices or small varices with red signs or in Child C cirrhosis) allowing to avoid about 33% of upper GI endoscopies [167]. An ELF score >8.34 was significantly associated with adverse liver-related events occurring in 61 out of 457 patients followed up for a median of 7 years [168]; a unit change in ELF was associated with a doubling of risk of liver-related outcomes. In a cohort of 1457 CHC patients, both increasing values of TE and Fibrotest had good prediction of 5-year survival [169]. Lastly, the prognostic accuracy of TE

### Table 3 Indicative performance characteristics of noninvasive methods to assess advanced (F3-F4) liver fibrosis in patients with NAFLD/NASH

<table>
<thead>
<tr>
<th>Score [References]</th>
<th>Serum markers/Fibroscan</th>
<th>Cut-off</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>NPV (%)</th>
<th>PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB-4 [51,141]</td>
<td>Age, ALT, AST, platelets</td>
<td>1.3</td>
<td>74-85</td>
<td>65-71</td>
<td>73-96</td>
<td>22-72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.67</td>
<td>34</td>
<td>98</td>
<td>59-93</td>
<td>60-93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.25</td>
<td>26</td>
<td>98</td>
<td>85</td>
<td>75</td>
</tr>
<tr>
<td>BAAT [134]</td>
<td>Age, BMI, ALT, triglycerides</td>
<td>2</td>
<td>71</td>
<td>80</td>
<td>86</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.676</td>
<td>33-51</td>
<td>98</td>
<td>85-86</td>
<td>82-90</td>
</tr>
<tr>
<td>BARD [51,136,139,142]</td>
<td>BMI, AST/ALT, diabetes</td>
<td>2-4</td>
<td>NA</td>
<td>NA</td>
<td>96</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>51-89</td>
<td>44-77</td>
<td>45-95</td>
<td>27-81</td>
</tr>
<tr>
<td>ELF® [135]</td>
<td>N-terminal propeptide of collagen type III, hyaluronate, TIMP-1</td>
<td>-1.0281</td>
<td>90</td>
<td>75</td>
<td>96</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2112</td>
<td>80</td>
<td>90</td>
<td>94</td>
<td>71</td>
</tr>
<tr>
<td>Fibrotest® [121]</td>
<td>Alpha-2-macroglobulin, apolipoprotein A1, haptoglobin, bilirubin, γ-GT</td>
<td>0.30</td>
<td>92</td>
<td>71</td>
<td>98</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.70</td>
<td>25</td>
<td>97</td>
<td>89</td>
<td>60</td>
</tr>
<tr>
<td>*Fibrometer® [137]</td>
<td>Platelets, prothrombin time, macroglobulin, AST, hyaluronate, age, urea</td>
<td>NA</td>
<td>79</td>
<td>96</td>
<td>92</td>
<td>88</td>
</tr>
<tr>
<td>Fibroscan® [122]</td>
<td>Transient elastography</td>
<td>7.9</td>
<td>91</td>
<td>75</td>
<td>52</td>
<td>79</td>
</tr>
</tbody>
</table>

*Performance characteristics for the detection of significant (≥F2) fibrosis
ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; IFG, impaired fasting glycaemia; TIMP, tissue inhibitor metalloproteinases; γ-GT, gamma glutamyl transpeptidase; NA, not available; NPV, negative predictive value; PPV, positive predictive value
was assessed in the setting of HIV/HCV co-infection [170] and in PBC [171]. However, to be clear as to whether these measurements should be used in clinical practice, prognostic models such as Child-Pugh and MELD or quantitative morphometry [172,173], rather than traditional qualitative histological staging, should represent the comparator when evaluating prognosis [174].

### Methodological limitations in the validation of noninvasive tools and future perspectives

A crucial issue in the evaluation of novel noninvasive tools is their validation against the currently available gold standard (i.e. the liver biopsy). Current methodology for the evaluation of the effectiveness of a test relies on the estimation of the AUROC, plotting sensitivity and 1-specificity. This represents the probability that a surrogate will correctly rank two randomly chosen patients: one with “normal” and one with “diseased” liver biopsy; and the perfect test will score 1.00. However, because liver biopsy is known to be an imperfect reference standard, AUROC determinations >0.90 are not achievable even for a marker that perfectly measures fibrosis [175]. This is indeed true for non-invasive tests which for their development were independent from liver histology, such as transient elastography. However, serum markers have been developed and calibrated with direct reference to a set of liver biopsies. Therefore, the perfect serum marker in this case would replicate the “gold” histological standard and should theoretically be able to reach an AUROC of 1, replicating even the misclassifications of liver biopsy [176]. As AUROC does not reflect the accuracy of a diagnostic test, (but just assesses discrimination in comparison to the reference standard), NPV/PPV and/or pre- and post-test probabilities depending on the prevalence of the target condition and clinical context in which the test is used (e.g. excluding any fibrosis, excluding any cirrhosis etc.) should be used instead. Given the practical impossibility to histologically analyze the whole liver (which would represent the ideal histological standard), novel clinical reference standards are needed in order for the field to move forward. Importantly, histological scores of liver fibrosis are ordinal categories with no quantitative relationship between them and therefore inappropriate to use as continuous variables [177]. Noninvasive markers of liver fibrosis should be ideally validated against quantitative histological measures. In this setting, we have described collagen proportionate area (CPA, i.e.: proportion of the area of the biopsy occupied by collagen measured by computer-assisted segmentation of digital images) [172] which could represent a better candidate to compare with noninvasive methods [178]. Moreover, HVPG is a well validated surrogate marker that could be used as an alternative reference standard, given its ability to predict overall liver-related outcomes [12,179,180]. Notably, in patients with recurrent HCV infection after liver transplantation, CPA correlated significantly with HVPG and performed better than both Ishak staging and HVPG for the prediction of clinical decompensation (AUROC=0.97) [181].

Unfortunately, enthusiasm for the development of noninvasive surrogates is not always supported by proper validation and sufficient strength of evidence. In a recent meta-analysis on TE, we exposed a series of issues such as invalidated stiffness cut-offs and low methodological quality: maximum interval between TE and liver biopsy was found to exceed 3-6 months in many studies, whereas only 6 studies reported the characteristics, and the proportion with optimal characteristics for both histological and TE measurements [97]. A further issue arising when assessing noninvasive markers is the so called “spectrum bias”: this is over-representation of extreme stages of fibrosis (F0 and F4). In a study population an excess of patients with severe fibrosis will automatically generate higher sensitivity and specificity values, than in a population including only lesser and adjacent stages of fibrosis (F1 and F2). In addition, the following areas should also be targeted for future research:

- Improvement and large-scale validation of promising novel imaging technologies (U/S characteristics, ARFI, MR-elastography).
- Switch from cross-sectional to longitudinal studies aiming to validate the performance of noninvasive surrogates to monitor fibrosis progression (or regression) over time.
- Definition of the role of combined vs. sequential noninvasive test approaches to optimize accuracy.
- Validation of noninvasive methods for screening fibrosis and cirrhosis in at-risk groups (e.g. diabetic patients) and in the general population.
- Validation in special populations (e.g. pediatric subjects).

### Conclusions

In the few past years, the field of noninvasive assessment of liver fibrosis has experienced explosive growth, resulting in a greatly expanded toolbox of biomarkers. Largely validated methods such as the Fibrotest/Fibrosure, and TE in Europe, are gaining a place in routine clinical practice, especially in CHC, whereas large-scale validation is pending for other highly prevalent diseases including CHB, NAFLD and ALD. Currently, use of liver biopsy solely to stage fibrosis in all patients is difficult to justify, considering it is invasive and has sampling variability. However, a diagnostic biopsy is still appropriate and allows a quantitative morphometric assessment of fibrosis, and if transjugular, the measurement of HVPG. The baseline biopsy associated with a noninvasive test, should give the best baseline for further monitoring with noninvasive tests. However, despite the satisfactory performance of noninvasive tests to detect both significant fibrosis and, even more, cirrhosis (i.e., the two major clinical endpoints), this does not reach sufficient diagnostic accuracy with current tests. There is still no perfect surrogate or method able to completely replace liver biopsy. Therefore, information deriving from both noninvasive methods and liver biopsy should be integrated in a complementary approach for
long-term management of chronic liver disease. Algorithms combining the most validated noninvasive methods could also be used as initial screening tools, avoiding liver biopsy, especially if cirrhosis or minimal to no fibrosis, is predicted by these tests. In cases of indeterminate results liver biopsy can be performed to confirm the exact stage of fibrosis. Recently, the APASL consensus on liver fibrosis has recommended the use of a stepwise algorithm of using noninvasive markers, concluding that this approach may reduce the need for liver biopsy by 30% [182]. Thus, incorporation of noninvasive tools into clinical guidelines may not be too far away, leading to wider use in clinical practice. Noninvasive methods will become clinically applicable, especially if specific antifibrotic treatments, become available; until then a liver biopsy will retain its role in hepatology.

Acknowledgements

V. Papastergiou is supported by the Hellenic Association for the Study of the Liver.

References


77. Rockey DC, Bissell DM. Noninvasive measures of liver fibrosis. Hepatology 2006;43:S113-S120.


122. Wong GL, Wong VW, Choi PC, Chan AW, Chan HL. Development of a non-invasive algorithm with transient elastography (Fibroscan) and serum test formula for advanced liver fibrosis in chronic hepatitis B. Aliment Pharmacol Ther 2010;31:1095-1103.


Non-invasive assessment of liver fibrosis

fibrosis markers identify patients with mild and progressive hepatitis C recurrence after liver transplantation. *Gastroenterology* 2010;138:147-158 e141.


179. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: In search of a pathophysiological classification of cirrhosis. *Hepatology* 2010;51:1445-1449.

