Noninvasive Tests in the Assessment of NASH and NAFLD Fibrosis: Now and Into the Future

Leon A. Adams, MBBS, FRACP, PhD1,2 Wah-Kheong Chan, MBBS, MRCP, PhD, AM3

1 Medical School, Faculty of Medicine and Health Sciences, The University of Western Australia, Nedlands, Western Australia, Australia
2 Department of Hepatology, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia
3 Department of Medicine, Gastroenterology and Hepatology Unit, University of Malaysia, Kuala Lumpur, Malaysia


Address for correspondence Leon Adams, MBBS, FRACP, PhD, Medical School, University of Western Australia, QEII Medical Centre, Verdun St, Nedlands, WA 6009, Australia (e-mail: leon.adams@uwa.edu.au).

Serum Biomarkers of NASH

A recent systematic review and meta-analysis examined 219 different diagnostic blood biomarkers for NASH and concluded no test differentiated steatosis from NASH with greater than 80% sensitivity and specificity.1 This highlights a significant gap in clinical care for NAFLD patients and may also reflect underlying limitations in using the comparison gold standard of liver biopsy from which biomarkers are referenced against.2 For example, pathologist agreement for NASH is modest and a diagnosis of NASH may be missed in one-quarter of liver biopsies due to sampling error.2,3

Markers of liver injury in NAFLD have been identified from either large-scale molecular screening or targeted analysis based upon known pathogenic pathways. The utility of a screening approach is exemplified by the NASH ClinLipMet score, which consists of a combination of metabolites and lipids identified by mass spectrometry, in association with
serum insulin, aspartate aminotransaminase (AST), and PNPLA3 genotype. The score predicts NASH with a high degree of accuracy (area under the receiver operator characteristic curve [AUROC] of 0.866); however, complexity and lack of validation have prevented widespread clinical application. Glycoproteomic screening has identified serum Mac-2 binding protein (Mac-2-bp) and its isof orm, Wisteria floribunda agglutinin-positive Mac-2-bp (WFA*M2BP) as accurate markers of liver injury in Asian and Western cohorts of NAFLD patients, although it remains unclear whether they are optimal for determining NASH or liver fibrosis. Lipidomic screening has identified alterations in glycerophospholipids, sphingolipids, sterols, and oxidized fatty acids; however, large-scale validation has not been performed and again the complexity of these markers limits easy translation to use at the bedside.

Hepatocyte apoptosis is upregulated in NASH and results in the release of cytokeratin-18 (CK-18) fragments which are cleaved intrahepatic intermediate filament proteins. The accuracy of CK-18 levels for predicting NASH has been variable. CK-18 levels improve with weight loss following bariatric surgery and NASH resolution in clinical trials; however, it is not sufficiently accurate to guide individual patient management. Serum levels of pathogenic mediators of liver injury including adipocytokines and inflammatory mediators (e.g., C-reactive protein, tumor necrosis factor, interleukin [IL]-6, C-X-C motif ligand 10, IL-1 receptor antagonist [IL-1RA]) have generally had limited accuracy in isolation for the detection of NASH. Subsequently, novel biomarkers have been combined in an effort to increase diagnostic accuracy. The NASH-test2 combines age, sex, α2-macroglobulin, apo A1, haptoglobin, total bilirubin, gamma-glutamyl transferase (GGT), AST, cholesterol, and triglycerides, which provides reasonable estimation of NASH (AUROC 0.80), although it performs less favorably in patients with type 2 diabetes. Other investigators have combined CK-18 and fibroblast growth factor-21 levels with additional markers including adipocyte fatty acid binding protein and IL-1RA, with improved predictive value compared with individual analytes. Further independent validation is required before routine clinical use.

Several studies have examined the utility of serum markers to predict NASH resolution following lifestyle or pharmacological intervention. Serum CK-18 levels and alanine aminotransaminase (ALT) levels parallel improvement in steatosis, lobular inflammation, and ballooning; however, their accuracy is modest for NASH resolution (AUC < 0.80). Similarly, serum levels of WFA*M2BP correlate with change in liver inflammation, ballooning, and fibrosis over 48 weeks; however, its low accuracy for NASH resolution (AUC < 0.60) precludes clinical use. A model including weight loss, type 2 diabetes, ALT normalization, age, and a nonalcoholic fatty liver activity score (NAS) ≥ 5 was highly accurate (AUC 0.94–0.96) in predicting NASH resolution in 261 patients over 12 months. However, the need for a baseline liver biopsy for evaluation of NAS score limits applicability and external validation is also required.

### Imaging Assessment of NASH

NASHMRI, which is derived from several magnetic resonance imaging (MRI) parameters independently associated with NASH, has been shown to be promising for the diagnosis of NASH with an AUROC of 0.83. However, this requires further validation. At present, no imaging modality has been adequately validated to diagnose NASH with the level of certainty required for clinical use. However, improvement in hepatic steatosis ≥ 30% based on MRI-proton density fat fraction following intervention has been shown to be predictive of NASH resolution and has been used a surrogate marker of NASH resolution in clinical trials. Moreover, several noninvasive methods have been proposed for the diagnosis of hepatic inflammation and fibrosis (e.g., the liver inflammation and fibrosis, LIF, score) or fibrotic NASH (e.g., FibroScan-based FAST score, MACK-3) (see section “Emerging Methods”).

### Serum Biomarkers of Fibrosis

“Simple” serum biomarker panels (e.g., FIB-4, AST-Platelet Ratio Index [APRI], NAFLD Fibrosis Score) include variables such as routine liver function tests, platelet count, and clinical parameters including age and body mass index (BMI). These tests are relatively easy to perform and subsequently have been widely validated; however, have indeterminate results in approximately 30% of cases. A meta-analysis of up to 39 studies involving over 9,000 subjects determined summary AUROC values for APRI, FIB-4, and NAFLD Fibrosis Score to be 0.77, 0.84, and 0.84, respectively, for the diagnosis of advanced fibrosis. The negative predictive values (NPVs) of these scores for excluding advanced fibrosis is high (89–93%); however, the positive predictive values (PPVs) are modest (55–67%) potentially leading to false positive results.

More complicated biomarkers, which incorporate direct markers of fibrogenesis and fibri nolysis such as serum tissue metalloproteinases or hyaluronic acid (e.g., Enhanced Liver Fibrosis [ELF] score, Hepascore, FIBROspect II, fibrometerV2), are more costly but are generally more accurate than “simple” biomarkers. Recent algorithms developed and validated in large multiethnic cohorts include HepaMet, composed of age, sex, homeostatic model assessment score, diabetes, AST, albumin, and platelet counts, and ADAPT, composed of Pro-C3, age, diabetes, and platelets. These scores have demonstrated superior accuracy to FIB-4 and NAFLD Fibrosis Score. There is a lack of studies directly comparing propriety algorithms, preventing recommendation of one over the other. Moreover, the use of serum fibrosis scores incorporating direct markers as the first test in sequential testing (see below) is less well studied compared with serum fibrosis scores using indirect markers.

The advantages of serum biomarkers include their accessibility, patient acceptability, reliability, and cost (Table 1). Awareness of potential confounding factors is important when interpreting results; significant hepatic and systemic inflammation may increase serum biomarker levels independently of fibrosis, and biomarkers incorporating bilirubin (Hepascore,
Fibrotest) may be falsely increased in patients with Gilbert’s syndrome or hemolysis. FIB-4 and NAFLD Fibrosis Score become less specific with increasing age, with higher cut-offs proposed to exclude advanced fibrosis in those older than 65 years. The performance of serum biomarkers is also less accurate in patients with diabetes, who represent an “at-risk” population for fibrosis.

**Imaging Assessment of Fibrosis**

Vibration-controlled transient elastography (VCTE) using FibroScan (Echosens) has been shown to be excellent for the diagnosis of advanced fibrosis and cirrhosis. It is based on the principle that liver stiffness increases as fibrosis increases, and that vibration travels faster in stiffer tissue. The speed of a vibration generated by the mechanical actuator is measured by an ultrasound device mounted on the same axis on the examination probe and is translated into stiffness measurement. An examination is considered successful if there were 10 valid measurements, and reliable if the interquartile range to median ratio were ≤ 0.30. The patient should be fasting for 2 hours prior to the examination as liver stiffness measurement (LSM) increases in the immediate post-prandial period. Additional factors which may falsely increase LSM values are outlined in Table 1. A LSM of < 5 kPa indicates that significant fibrosis is unlikely, < 10 kPa indicates that advanced fibrosis is unlikely, > 15 kPa indicates that advanced fibrosis is likely, and > 20 to 25 kPa indicates that clinically significant portal hypertension is likely. The strength of FibroScan is the exclusion of advanced fibrosis with high (< 95%) NPVs for values < 8 kPa, though modest PPVs (68%) with values > 9.6 kPa. The XL probe was introduced to reduce failed and unreliable examinations in obese patients and has similar accuracy as the standard M probe. Although the XL probe results in lower LSM compared with the M probe in the same patient, similar cut-off values can be used for interpretation if the probe selection was based on the device recommendation or the BMI of the patient. Overall, successful and valid scans are achievable in > 90% of patients; however, failed or invalid scans may be seen in up to 30% of obese individuals. In short, VCTE provides a simple, non-invasive method for estimating hepatic fibrosis with immediate results for patient counseling in the clinic. FibroTouch (Hisy) is another transient elastography device for LSM. It has the advantage of using a single probe for patients with different body habitus. It is not as extensively studied as FibroScan but has been shown to have similar accuracy.

**Table 1 Comparative characteristics of different noninvasive methods of fibrosis assessment**

<table>
<thead>
<tr>
<th></th>
<th>Serum biomarkers</th>
<th>Sonographic elastography</th>
<th>Magnetic resonance elastography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient acceptability</td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>Excellent</td>
<td>Good</td>
<td>Excellent</td>
</tr>
<tr>
<td>Availability</td>
<td>Wide</td>
<td>Limited</td>
<td>Limited</td>
</tr>
<tr>
<td>Cost</td>
<td>+</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Accuracy</td>
<td>+++</td>
<td>++++</td>
<td>+++++</td>
</tr>
<tr>
<td>Failure/Unreliable rate</td>
<td>&lt; 1%</td>
<td>20%</td>
<td>1-2%</td>
</tr>
<tr>
<td>Confounders</td>
<td>Systemic inflammation, Gilbert's, hemolysis, acute hepatitis, cholestasis</td>
<td>Acute hepatitis, cholestasis, congestion, focal liver lesions, portal vein thrombosis, BMI</td>
<td>Hepatic iron*</td>
</tr>
</tbody>
</table>

Abbreviation: BMI, body mass index.

*May be overcome with spin-echo echo-planar sequences.
significantly greater unreliable results. Another study on 112 NAFLD patients yielded similar findings. Separately, a study on 100 NAFLD patients who underwent liver biopsy found FibroScan to be more accurate than Elast-PQ for the diagnosis of fibrosis stages ≥ F2 and ≥ F3. Compared with transient elastography, ultrasound elastography techniques has the advantage of anatomical visualization, however, the operator must undergo additional training to have the technical and anatomical expertise to perform the examination.

Magnetic resonance elastography (MRE) uses a special pulse sequence to image the micron-level cyclic displacements caused by propagating waves generated by a hardware. The acquired data are processed by a dedicated software that produces a color elastogram of the examined liver and a stiffness measurement of the selected region of interest. In a pooled analysis of individual patient data of 232 NAFLD patients from 9 studies, MRE at 60 to 62.5 Hz had an AUROC of 0.86, 0.87, 0.90, and 0.91 for the diagnosis of fibrosis stages ≥ F1, ≥ F2, ≥ F3, and F4, respectively. The more advanced three-dimensional (3D) MRE at 40 Hz (AUROC of 0.98) has been shown to be more accurate than the conventional 3D MRE at 60 Hz and 2D MRE at 60 Hz (AUROC of 0.93 and 0.92, respectively) for the diagnosis of fibrosis stage ≥ F3. A study on 127 NAFLD patients who underwent liver biopsy found that MRE was more accurate than FibroScan for the diagnosis of fibrosis stages ≥ F2 (AUROC 0.91 vs. 0.82) and F4 (AUROC 0.97 vs. 0.92), but not fibrosis stages ≥ F1 (AUROC 0.83 vs. 0.78) and ≥ F3 (AUROC 0.89 vs. 0.88).

**Combination Testing**

A current limitation of the noninvasive tests is the presence of an indeterminate range and relatively poor PPVs for the determination of advanced fibrosis or cirrhosis typically between 40 and 60% for cirrhosis. Combining serum-based tests with elastography is attractive given they use different and complimentary patient characteristics to predict fibrosis. Tests may be combined in a synchronous fashion (outlined in – Fig. 1), where concordant positive or negative tests provide a high PPV or NPV for advanced fibrosis, but at the expense of indeterminate or conflicting results in approximately half of patients. Sequential testing typically utilizes a serum test initially, with elastography used in indeterminate or positive serum biomarker results (– Fig. 1). A study of 759 biopsy-proven NAFLD patients found that the use of noninvasive fibrosis scores (i.e., NAFLD Fibrosis Score or FIB-4 score) followed by FibroScan only for patients with indeterminate or high scores was most accurate for the diagnosis of advanced fibrosis when the prevalence of advanced fibrosis is low, such as in the general population. On the other hand, the use of LSM ≥ 10 kPa had the highest diagnostic accuracy in population with higher prevalence of advanced fibrosis. In a separate study on 968 patients with biopsy-proven NAFLD, sequential testing using NAFLD Fibrosis Score or FIB-4 score followed by FibroScan for patients with indeterminate score was similarly found to be better than the tests individually, regardless of obesity and ALT levels, although the accuracy of sequential testing was lower in obese patients. Notably, the accuracy of FibroScan was lower among obese patients irrespective of whether a M or XL probe was used. In another study on patients with chronic liver disease of various etiologies including NAFLD, a stepwise algorithm, starting with the easy liver fibrosis test or eLIFT (which is a composite score of readily available parameters, i.e., age, gender, AST, GGT, platelet, and prothrombin time), followed by FibroMeter, VCTE (which combines in a single formula the blood markers for FibroMeter and FibroScan result) for patients with increased eLIFT score, had a diagnostic accuracy for advanced fibrosis of 83.3% with sensitivity and specificity of 76.1 and 92.2%, respectively.

![Fig. 1](image_url) Predictive and indeterminate values of concurrent and sequential testing strategies using noninvasive fibrosis tests in nonalcoholic fatty liver disease. Figures are based upon use of FIB-4 or NAFLD Fibrosis Score as serum tests and FibroScan as elastography technique.
Monitoring Fibrosis over Time

Experience in evaluating noninvasive tests over time in response to treatment or for evaluating disease progression is sparse. Improvement of liver inflammation with treatment of NASH may in turn lead to reduction in liver elasticity and blood biomarkers and an underestimation of fibrosis stage. Serum biomarkers including ELF, WFA + M2BP, NFS, Fib4, and APRI have poor to modest accuracy (AUC < 0.75) in predicting fibrosis response to pharmacotherapy or lifestyle intervention and cannot be recommended to monitor for short-term (<12 months) treatment response. Both MRE and VCTE had a low degree of accuracy (AUC < 0.65) for predicting fibrosis improvement over 24 weeks in a negative therapeutic trial. The AUC of MRE increased to 0.79 when combined with baseline values, suggesting that it may be useful for monitoring fibrosis response. An algorithm combining platelet count, ALT normalization, and change in hemoglobin A1c had high accuracy (AUC 0.96) for predicting fibrosis improvement following 1 year of lifestyle intervention in NASH patients; however, this requires further validation. In the absence of intervention, Fib-4, APRI, and NAFLD Fibrosis Score are reasonably accurate (AUC 0.80–0.82) in predicting progression to advanced fibrosis, although they do not improve with fibrosis regression. Consequently, they can be used to confirm absence of progression (NPVs 89–90%) rather than detecting progression to advanced fibrosis (PPVs 44–49%).

Prediction of Outcomes

Serum biomarkers, FibroScan and multiparametric MRI, have been demonstrated to be predictive of outcomes in patients with chronic liver disease. The NAFLD Fibrosis Score, FibroMeter, Hepascore, and FibroScan have been shown to be able to stratify NAFLD patients according to liver-related complications and overall mortality. These tests are able to discriminate patients with more severe fibrosis from those with less severe fibrosis, which is the most important determinant for liver-related outcomes and overall survival in NAFLD patients. A separate study found high-risk cases (i.e., those with advanced fibrosis) based on paired FibroScan values, to be predictive of not only liver-related complications, but also cardiovascular events. Recently, the stepwise algorithm using eLIFT and FibroMeterVCTE (see above) has also been shown to be able to stratify patients with chronic liver disease of various etiologies according to liver-related complications and overall mortality. Similarly, a study of 112 patients with chronic liver disease of various etiologies, found a multiparametric MRI (see below) defined LIF score of ≥3, to be significantly associated with the development of liver-related complications or mortality.

Emerging Methods

Fibrotic NASH, defined as NASH with NAFLD activity score ≥4 and fibrosis stage ≥2, is increasingly used as target for enrolment in clinical trial. In line with this, recent noninvasive methods have focused on this diagnostic goal. The FibroScan-based FAST score uses controlled attenuation parameter, LSM, and serum AST level for the diagnosis of fibrotic NASH. The score was developed from a prospective, multicenter study of patients undergoing liver biopsy for suspicion of NAFLD in the U.K. and was validated in seven international cohorts. The AUROC for the diagnosis of fibrotic NASH in the derivation cohort and pooled validation cohort was 0.80 and 0.85, respectively. In the pooled validation cohort, the PPV and NPV were 0.69 and 0.94, respectively, with 30% of patients in the gray zone. The FAST score can help identify patients who are unlikely to have fibrotic NASH and reduce the screen failure rate for clinical trial.

MACK-3 is a score that uses the homeostatic model assessment of insulin resistance, AST, and CK-18 for the diagnosis of fibrotic NASH. The accuracy and performance characteristics are similar to the FAST score with the AUROC for the diagnosis of fibrotic NASH to be 0.85 in both derivation and validation cohorts. In a validation cohort, the PPV and NPV were 0.82 and 0.97, respectively, with 36% of patients in the gray zone. In an external cohort, the MACK-3 had an AUROC of 0.80 for the diagnosis of fibrotic NASH with sensitivity, specificity, PPV, and NPV of 1.00, 0.44, 0.43, and 1.00, respectively. However, 47% patients were in the gray zone. The MACK-3 has the advantage of being a blood-based test with high applicability, but requires fasting serum insulin, which is not routinely tested in clinical practice, and CK-18 which is not widely available.

In a proof-of-concept study on 77 patients with chronic liver disease of various etiology and 7 healthy controls, multiparametric MRI was reported to have an AUROC of 0.94 to distinguish patients with any fibrosis from healthy controls and patients without fibrosis. Multiparametric MRI uses multigradient-echo acquisition to calculate the T2* map, which accurately detects elevated iron concentration. On the other hand, T1 relaxation time map detects extracellular water, which may be increased in the presence of inflammation and fibrosis. However, T1 relaxation time map is also increased with elevated iron concentration, which can be corrected using data from the T2* map. Compared with MRE, this MRI technique has the advantage of not requiring an additional hardware to generate propagating waves and is not affected by adiposity and the presence of ascites. However, it cannot be used in the presence of severely elevated iron concentration. Using optimal cut-offs for corrected T1 or cT1 for the differentiation of the different fibrosis stages, the LIF score, which is a standardized continuous score of 0 to 4, was derived. In a study on 71 patients with suspected NAFLD who underwent liver biopsy, the LIF score had an AUROC of 0.80 to differentiate NASH from simple steatosis and to distinguish significant (defined as activity ≥2 and fibrosis ≥2) from mild NAFLD. The optimal cut-off for diagnosis of NASH or significant NAFLD was same at 1.4. Of note, there was significant overlap in the LIF score between F0 and F1, and between F2 and F3 (based on steatosis, activity, and fibrosis score). The AUROC for diagnosis of cirrhosis was 0.85 and the optimal cut-off was 3.0. In a separate study on adults who underwent liver biopsy, cT1 had an AUROC of 0.72 for the diagnosis of fibrosis.

Noninvasive Tests in the Assessment of NASH and NAFLD Fibrosis

Adam, Chan
stage (based on modified Ishak score) ≥ F1, ≥ F3, or ≥ F5, which was inferior to transient elastography, which had an AUROC of 0.93, 0.84, and 0.86, respectively. Further independent validation in large cohorts of NAFLD patients with a full spectrum of histological disease are required before the accuracy and utility of multiparametric MRI can be determined.

**Conclusion**

NAFLD is highly prevalent and has an attendant risk of liver-related morbidity and mortality, which is predicted by the degree of underlying liver fibrosis. The combination of serum and sonographic elastography techniques are able to accurately detect the presence or absence of advanced fibrosis with few patients requiring liver biopsy for indeterminate results. However, noninvasive tests are currently insufficiently accurate to monitor treatment response, although MRE holds promise that it may fill the gap. Strategies to date have concentrated on detecting patients with advanced fibrosis; however, earlier disease that is expected to progress (e.g., fibrotic NASH) may become the diagnostic target in the future as safe, effective, and affordable treatments become available.

**Main Concepts and Learning Points**

- Serum and imaging methods are not currently recommended for diagnosing NASH due to limited accuracy; however, combination methods offer promise for distinguishing fibrotic NASH.
- Serum biomarker panels and elastography reliably exclude advanced fibrosis, however, have modest predictive value. MRE is currently considered the most accurate elastography technique, however, is limited by expense and availability.
- Concordant concurrent serum marker and transient elastography results provide a high degree of certainty for the presence or absence of advanced fibrosis; however, conflicting and thus indeterminate results occur in up to 50% of patients.
- Serial normal serum biomarker results over time confirm absence of progression to advanced fibrosis; however, longitudinal monitoring is not useful to predict fibrosis progression or response to treatment. MRE offers promise as a tool for monitoring fibrosis change over time, however, requires further validation.

**Conflict of Interest**

Dr. Adams reports personal fees from Pfizer, outside the submitted work. In addition, Dr. Adams has a patent Hepascore issued to Quest diagnostics. Dr. Chan reports grants from Resonance Health and personal fees from Boehringer Ingelheim, outside the submitted work.

**References**

Noninvasive Tests in the Assessment of NASH and NAFLD Fibrosis

Adams, Chan 337


47 Leong WL, Lai LL, Nik Mustapha NR, et al. Comparing point shear wave elastography (ElastPQ) and transient elastography for diagnosis of fibrosis stage in non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2019


51 Petta S, Vanni E, Bugianesi E, et al. The combination of liver stiffness measurement and NAFLD fibrosis score improves the noninvasive diagnostic accuracy for severe liver fibrosis in patients with nonalcoholic fatty liver disease. Liver Int 2015;35(05):1566–1573


56 Jayakumar S, Middleton MS, Lawitz EJ, et al. Longitudinal correlations between MRE, MRI-PDFF, and liver histology in patients
with non-alcoholic steatohepatitis: analysis of data from a phase II trial of selonsertib. J Hepatol 2019;70(01):133–141
61 Kamarajah SK, Chan WK, Nik Mustapha NR, Mahadeva S. Repeated liver stiffness measurement compared with paired liver biopsy in patients with non-alcoholic fatty liver disease. Hepatol Int 2018;12(01):44–55