The surveillance of progression and assessment of treatment endpoint for nonalcoholic steatohepatitis

Yi-wen Shi, Jian-Gao Fan*

Center for Fatty Liver, Department of Gastroenterology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai Key Lab of Pediatric Gastroenterology and Nutrition, Shanghai, 20092, China.

Running title: Surveillance of NASH

Correspondence: Jian-Gao Fan
fanjiangao@xinhuamed.com.cn

Abstract

Nonalcoholic steatohepatitis (NASH) is the aggressive form of nonalcoholic fatty liver disease (NAFLD), characterized as steatosis-associated inflammation and liver injury. Without effective treatment or management, NASH would develop life-threatening outcomes. In this situation, evaluation and identification of those at-risk for adverse outcomes are important. The key issues in screening NASH patients are the assessment of advanced fibrosis, differentiation of NASH from simple steatosis, and their dynamic changes during follow-up. Currently, the staging for NASH and evaluation of effectiveness for drugs still rely on pathological diagnosis, while liver biopsy brings sample error issues and subjectivity. To address this problem, optimizing the pathological assessment and developing noninvasive surrogate methods for accessible,
accurate, and safe evaluation is of significance. Although noninvasive methods including elastography, serum soluble biomarkers and combined models have been widely studied in the last decade, the application of noninvasive diagnostic measurements in clinical practice is still insufficient. Much work remains to be done in establishing cost-effective strategies both for screening for at-risk NASH and identify the changes of disease severity. In this review, we summarized the current state of the noninvasive methods for detecting steatosis, steatohepatitis and fibrosis of NASH, introduced the noninvasive assessment for screening at-risk patients, and focused on the characteristics should be monitored in the follow-up.

**Keywords:** nonalcoholic steatohepatitis; noninvasive diagnosis; disease progress; risk stratification; treatment efficacy.
1. Introduction

Nonalcoholic fatty liver disease (NAFLD) is a heterogeneous and silently progressive disease, affecting about 32% of the global population\(^1\),\(^2\). With the alarming increase rate of worldwide prevalence and incidence, NAFLD has become one of the most common causes of chronic liver diseases in the majority of the industrialized areas\(^3\),\(^4\). Compared with nonalcoholic fatty liver (NAFL) characterized by bland steatosis, nonalcoholic steatohepatitis (NASH) is a more progressive phenotype of NAFLD, characterized as hepatocyte injury, inflammation and scarring. It was estimated that around 25% of the NAFLD patients will develop NASH, and 20% of patients with NASH will develop cirrhosis and hepatocellular carcinoma (HCC) in the 20 to 30 years\(^5\). In the recent decade, the liver-specific and overall mortality rate of NASH is growing fast, especially in the patients with obesity, type 2 diabetes mellitus (T2DM) and metabolic syndrome\(^6\). Early identification and targeted treatment for NASH are urgently needed to improve patient outcomes.

Currently, diagnosis and evaluation of severity for NASH is still based on liver biopsy-proven histopathological assessment and scoring, relying on invasive liver biopsy. The main scoring systems covered the characteristics of liver fibrosis, inflammation and steatosis of NASH\(^7\),\(^8\). Although a number of noninvasive tests and predictive models exist to characterize fibrotic NASH patients, the diagnostic performance and clinical application are to be improved. Since there are still no NASH-specific drugs with definite efficacy approved by major drug administration agencies worldwide, lifestyle intervention including dietary changes and exercise, with the
purpose of 10% weight loss, is the most effective approaches for the management of fibrotic NASH and underlying cardiometabolic comorbidities. Based on this situation, a number of novel drugs is under discovery and development. Accepted as the “gold” standard for diagnosis of NASH, liver biopsy is essential for both patient enrollment and efficacy assessment for most phase 2b trials and all phase 3 trials, under the approval requirements.

The accurate evaluation of the severity of NASH and the risk for progression of liver cirrhosis and HCC is an essential need in both screening at-risk NASH patients and reflecting the treatment response (including NASH remission and cirrhosis prevention) of novel NASH drugs in clinical trials. In the present review, we will discuss the approaches and the advances of the surveillance of progression and assessment of treatment endpoint for NASH.

2. Risks of NASH progression

*Fibrosis: the determinant factors*

Liver fibrosis is recognized as the determinant of liver-related morbidity and mortality in patients with NAFLD/NASH. Previous evidence showed that significant fibrosis (≥F2) and advanced fibrosis (≥F3) are independently associated with overall mortality, liver transplantation, and liver-specific mortality in patients with NAFLD. In this cohort, both patients with or without NASH, fibrotic NAFLD had lower survival free of liver transplantation rate than non-fibrotic NAFLD. Recently, a meta-analysis added more evidence that the risk of liver-related mortality, all-cause mortality and liver transplant was increased with the increasing biopsy-confirmed fibrosis stage.
According to the Finnish population-based FINRISK and Health 2000 studies with a median follow-up of 12.1 years. Crude incidence of liver-related outcomes in NAFLD was 0.97/1000 person-years, and it was associated with noninvasive fibrosis stage. Moreover, HCC risk was highest with cirrhosis, followed by noncirrhotic fibrosis and comorbid T2DM in biopsy-proven NAFLD cohort. Correspondingly, NASH patients with compensated cirrhosis would develop reduced liver-related complications in the situation of fibrosis regression, which was presented as decrease in the levels of NAFLD fibrosis score (NFS), liver stiffness measurement, hepatic collagen and alpha-smooth muscle actin (α-SMA) expression. In addition, most phase 3 clinical trials of NASH with novel drugs would also target fibrosis with stage≥F2 to prevent fibrosis progression and liver-related events. Therefore, identifying NASH patients with significant fibrosis or advanced fibrosis would be effective in screening high-risk population for progression to liver cirrhosis and HCC.

Inflammation: trigger of fibrosis and carcinogenesis

Patients with simple steatosis are often considered to have similar life expectancy compared with the general population, while patients with NASH are generally considered to be more progressive. Due to the “inflamed” situation, adipose tissue released free fatty acids and toxic lipids, followed by fat accumulation, lipotoxicity, oxidative stress, mitochondrial dysfunction in the hepatocytes, leading to liver fibrogenesis and carcinogenesis. It was reported up to one-third of NASH patients without effective intervention would develop advanced liver fibrosis or cirrhosis, and even HCC. Although in a previous study the characteristic of fibrosis covered the
impact of steatohepatitis on the prognosis of NAFLD patients, persistent hepatocyte injury and chronic inflammation in the liver is one of the driving forces of disease progression and carcinogenesis\textsuperscript{22}. Further study confirmed that NASH had faster fibrosis progression and higher risk for HCC than NAFL, patients with NAFL would progress one fibrosis stage per 14.3 years, while patients with NASH would progress one fibrosis stage per 7.1 years\textsuperscript{23}.

*Metabolic dysfunction: cause or consequence*

Obesity is the most common form of metabolic dysfunction, which is also considered related to the epidemic of NAFLD. Overall obesity increases de novo lipogenesis, and decreases β oxidation of free fatty acids and very low-density lipoprotein (VLDL) secretion, resulting in hepatocytes lipidosis and lipotoxicity. However, it should be noted that a large proportion of patients with NAFLD are lean or nonobese\textsuperscript{24, 25}. About 8-19% of Asians with body mass index (BMI) less than 25 kg/m\textsuperscript{2} are also found to have NAFLD\textsuperscript{26}, and the prevalence of NAFLD in nonobese subjects is as high as 16%\textsuperscript{24}. Actually, obesity defined by BMI is only a crude measurement of obese. Other anthropometric parameters might be useful for diagnosis of central obesity, occult obesity, and sarcopenic obesity. Central adiposity, sarcopenia, dyslipidemia, and insulin resistance are also strongly associated with NASH and related fibrosis in a dose-dependent manner\textsuperscript{27}. The progressive course of NASH has been closely linked to the increasing number of metabolic comorbidities. T2DM had the strongest association with incident HCC in patients with NAFLD\textsuperscript{28-30}. Metabolic syndrome is an independent predictor of all-cause, liver-specific, and cardiovascular mortality in patients with
NAFLD\textsuperscript{31,32}. In contrast, mortality of metabolically-normal NAFLD patients is similar to the cohort without liver disease\textsuperscript{33-35}. Thus, assessing metabolic dysfunctions including insulin resistance may help define the high-risk NASH patients\textsuperscript{36}. In addition, accumulating evidence suggests that NAFLD has complex link with metabolic dysfunctions. NAFLD especially NASH is also associated with an increased risk of incident T2DM and atherosclerotic cardiovascular disease events\textsuperscript{37}.

3. Histopathological surveillance for NASH

\textit{Liver biopsy is imperfect}

In consideration of a chronic progressive disease, the screening of high-risk patients and the surveillance for the development of liver-related complications is urgently needed for the management of NASH. Currently, numbers of novel NASH pharmacological agents are being developed, and monitoring the treatment response also relies on reliable assessment in their clinical trials. Histopathological assessment is always considered as the “gold” standard for the diagnosing and evaluating of NASH severity and fibrosis stage. However, limited by its invasive nature, liver biopsy is not feasible for repeated assessment. The histological evidence from liver biopsies both have moderate accuracy and limited validation, more reliable technique is needed toward an accurate quantification.

Histological classification of NASH is semiquantitative scoring systems. NAFLD Activity Score (NAS) developed by NASH Clinical Research Network (NASH CRN) and steatosis, activity, fibrosis (SAF) scoring system developed by Fatty Liver Inhibition of Progression (FLIP) Pathology Consortium are the two mostly used scoring
systems\textsuperscript{7,38}. Both systems identify the location and the features of fibrosis, numbers of inflammatory foci, numbers of balloon cells and percentage of parenchymal involvement by steatosis. The process of assessments depends on manual and subjective judgment, which results in intra- and inter-observer variability in assessment. Although liver biopsy is generally safe and widely available, histological scoring is also limited by sampling error and ordinal classification. Developing innovative methods including machine learning (ML), artificial intelligence (AI) - based approach, whole-slide image (WSI), etc. to facilitate histopathological assessment may be the key to improve the “gold” standard.

**Novel liver biopsy-based assessment**

Second-harmonic generation (SHG) microscopy is highly sensitive to the collagen fibril/fiber structure, and has enabled the imaging of fibrillar collagen in various tissues. SHG-based novel technology has also been applied for the assessment of hepatic fibrosis in chronic liver diseases\textsuperscript{39}. As one of the SHG-based novel technologies, the assessment of hepatic steatosis by HistoIndex has a good correlation with histopathologists\textsuperscript{40}, and has been applied in a phase 2 clinical trial (MGL-3196, Resmetirom) to evaluated the dynamic change of steatosis during treatment\textsuperscript{41}. Quantification of fibrosis-related parameters (q-FPs) model is developed by Wang et al to assess the characteristics of liver fibrosis in NAFLD. The model containing four q-FPs (the number of collagen strands, strand length, strand eccentricity, and strand solidity) was established in 50 test subjects and validated in 42 validation subjects, providing continuous and quantitative evaluation of fibrosis \textsuperscript{42}. Furthermore, a
combination of qFibrosis, qInflammation, qBallooning, and qSteatosis (qFIBS index) was also developed to add quantitative assessment of the characteristics of NAS (lobular inflammation, ballooning, and steatosis) by using SHG and two-photon excitation fluorescence (TPEF) imaging technology. Now qFIBS was developed and then validated in a cohort of 219 patients with biopsy-proven NAFLD/NASH and showed a robust correlation between NAS and fibrosis stages\textsuperscript{43}. Recently, qFIBS was applied in the phase 2 trial of tropifexor (NCT02855164), to assess the resolution of NASH and fibrosis. qFIBS tested regressive changes in septa morphology and reduction in septa parameters in F3 patients, which would be classified as “unchanged” by traditional scoring, which showed greater sensitivity of qFIBS\textsuperscript{44}.

More advances in machine-learning based approaches are enabling the histopathological monitoring of the progression and retrogression of NASH\textsuperscript{45}. Digital WSI could scan hematoxylin-eosin (HE) stained slides to quantify steatosis by assessing the steatosis proportionate area (SPA), and could scan Elastica van Gieson-stained slides to quantify fibrosis by assessing the collagen and elastin fibers\textsuperscript{46-48}, and is thus regarded as an automated, precise, objective and quantitative assessment of NASH. As the most important feature of NASH, ballooning cells are always too subjective to determine. AI-based technology may be trained to reproducibly quantify ballooned hepatocytes and standardize the evaluation\textsuperscript{49}. ML-based model is also able to assess NASH histological characteristics accurately as well as to assess treatment response. PathAI has concordance with ordinal grades from pathologists in terms of three NAS components. In addition, PathAI detected improvement of DELTA Liver
Fibrosis score in fibrosis responders in the combination group (cilofexor+ firsocostat) in the ATLAS study. AI or ML-based technologies are rapidly emerging and are promising to improve the inadequacies of pathological assessment of fibrotic NASH.

4. Noninvasive markers: more practical for monitor

With the increasing prevalence of NASH, the base of at-risk patients who needed screening is large. Liver biopsy might become one of the important bottlenecks in the diagnosis and monitoring of these patients. Currently, there is an essential need to develop accurate noninvasive tests, markers or models to evaluate NASH severity and to monitor drug efficacy. Based on these needs, researchers have developed some noninvasive assessments including serum biomarker, elastography-based markers, imaging studies, genetic tests and combination with omics.

Noninvasive tests are more acceptable for evaluation of steatosis degree and fibrosis stage than liver biopsy, which would also improve the compliance of screening and monitoring of NAFLD. As histological assessment from liver biopsy is still imperfect, an ideal solution is to link the clinical outcomes such as cirrhosis, HCC, and liver-related complications with novel noninvasive markers. As the surrogate, the feasible approach also lies in correlating the histological severity of NASH and fibrosis stages with quantized noninvasive markers (Table 1).

Serum biomarkers for assessment of steatosis

Currently, the most promising noninvasive diagnosis of hepatic steatosis includes the fatty liver index (FLI), the hepatic steatosis index (HSI), the NAFLD-liver fat score, the visceral adiposity index, the lipid accumulation product (LAP), and the triglyceride
× glucose index. Most of these indexes have been validated in biopsy-proven cohort, or use magnetic resonance spectroscopy (MRS) as a reference. The accuracy of FLI, HIS, LAP and the Zhejiang University index (ZJU) were evaluated in a general population by ultrasonography. Although FLI showed the highest C-statistic (0.85), the relatively low sensitivity of ultrasonography in detecting mild steatosis should be concerned. Although assessing steatosis grade is rather simpler than assessing inflammation or fibrosis, detecting >5% hepatic steatosis by circulating biomarkers alone is quite insufficient. Combinations or equations of biomarkers would add more accuracy in correlation with steatosis. Dallas Steatosis Index (DSI), consisting of age, sex, diabetes, hypertension, race, BMI, serum triglycerides and ALT, is developed from the Dallas Heart Study, containing 737 patients with MRS diagnosed liver fat content. The C-statistic of DSI is up to 0.824, while the diagnostic performance still needs external validation. It should be noted that, in most areas, ultrasound test may be more available than blood-based test. The serum proteins measured in these models are associated with metabolic disorders or insulin resistance, not strictly specific to hepatic fat content, which may be an explanation why these models have insufficient accuracy especially in nonobese or lean subjects. Therefore, these serum-based noninvasive markers might add limit information for surveillance.

_Serum biomarkers for assessment of liver fibrosis_

As the major driver of liver-related outcomes in NAFLD, assessing fibrosis stages is essential for screening at-risk patients. Simple serum biomarker panels as fibrosis-4 index (FIB-4) and aspartate aminotransferase (AST) to Platelet Ratio Index (APRI),
originally developed for chronic viral hepatitis also could be applied in NASH patients. The cut-off value for FIB-4 is 2.67 and 1.30 to rule in and rule out advanced fibrosis in patients with NAFLD, respectively. NAFLD Fibrosis Score (NFS) is developed from liver biopsy-proven NAFLD cohort, with the cut-off values 1.455 and 0.676. BARD score also aims to diagnose advanced fibrosis combining BMI, AST/ALT and diabetes. Both FIB-4 and NFS are relatively easy to perform and recommended by current guidelines to identify patients at low or high risk for advanced fibrosis in patients with NAFLD. These tests have been widely used in multiple scenarios, especially available in primary health care units. However, due to the various etiology of the derivation cohort, the accuracy of these tests needed to be improved when applied in NAFLD cohort. In addition, models developed from biopsy-proven NAFLD cohort often use higher cut-off values than the general population. These lead to inferior diagnostic performance of NFS, FIB-4 and APRI.

For tests including patented markers often contain direct biomarkers of fibrogenesis or fibrinolysis which are from the extracellular matrix (ECM) due to liver injury. Type III collagen and hyaluronic acid (HA) and often involved in these tests. PIIINP could discriminates against advanced fibrosis with the C-statistic 0.82-0.84. Enhanced liver fibrosis (ELF) test, a commercial panel of markers, consisting of serum HA, amino-terminal propeptide of type III procollagen (PIIINP), and tissue inhibitor of metalloproteinase-1 (TIMP-1), was first developed in children with NAFLD and validated in larger cohorts. Recently ELF test was used to assess the fibrosis improvement during aldafermin (NGM282) treatment. Another type III collagen-
based fibrosis algorithm including age, presence of diabetes, PRO-C3 (a marker of type III collagen formation) and platelet count (called ADAPT) also showed better diagnostic performance than APRI, FIB-4 and NFS in predicting advanced fibrosis. FibroMeter consists of age, weight, glucose, AST, ALT, ferritin, and platelets, which has been compared directly with ELF. ELF and FibroMeter had significantly higher C-statistic than NFS and FIB-4 in diagnosing advanced fibrosis, while the C-statistic did not differ significantly between ELF and FibroMeter. FibroTest is also a commercial panel, with C-statistic of 0.75-0.86 for significant fibrosis and 0.81-0.92 for advanced fibrosis. FIBROSpect consisting of alpha 2 macroglubulin, HA, and TIMP-1, is highly sensitive (positive predictive value, NPV 92.5-94.7%) to advanced fibrosis, with a C-statistic 0.856. Hepamet was developed in 2452 biopsy-proven NAFLD patients, and proved to have higher C-statistic than FIB-4 and NFS. Hepamet would also avoid the interference by age, BMI or diabetes. These tests provided more accuracy in predicting advanced fibrosis but also increased costs, and there is still a lack of direct comparison in the same cohort. In general, biomarkers or models detecting advanced fibrosis achieved relatively high NPV while the positive predictive value (PPV) still needed to be improved.

Serum biomarkers for assessment of steatohepatitis

Regarding steatohepatitis, both hepatocyte ballooning and inflammation are the most important features, while current biochemical or imaging measures cannot effectively distinguish NASH from NAFL. Using serum ALT as a predictive marker for diagnosis of steatohepatitis is insufficient. It is reported less than 30% of the NASH patients have
Elevated ALT levels (>35 U/L). If using ALT > 2 times the upper limit of normal (ULN) to diagnose NASH would achieve only 50% sensitivity and 61% specificity. Cytokeratin 18 (CK18) is released into the serum in initiation of apoptosis in the form of CK18-M30 and CK18-M65 fragments. Serum CK-18 is the one that has been the most widely investigated in diagnosis of NASH. Previously CK-18 showed potential predictive value for fibrosis, but then was found to have a better correlation with ALT rather than with steatosis or fibrosis. Other studies with repeated liver biopsy also showed serum CK18 level is associated with NAS ≥ 5 (definite NASH) in patients with NAFLD. Meta-analyses have confirmed that CK-18 could predict steatohepatitis with a C-statistic around 0.8, while the sensitivity of 66%–78% is waited to be improved. Index of NASH (ION) is also developed to diagnose steatosis, which consists of waist-to-hip ratio, triglyceride, ALT, HOMA and gender, but ION showed low sensitivity in an external cohort especially in nonobese subgroup. Although serum level of hypersensitive C-reactive-protein (hs-CRP) was included in the diagnosis of metabolic-dysfunction associated fatty liver disease (MAFLD), the diagnostic value in NASH is quite limited. Recently a study enrolling 100 subjects observed the independent relationships between hs-CRP and NAFLD. More direct and solid evidence is needed to make hs-CRP a diagnostic marker for NASH. Genetic biomarkers both single nucleotide polymorphism and noncoding RNAs are studied for predicting NASH. NASH Score (PNPLA3 genotype, AST, and fasting insulin) and circulating miR-122 showed the potential significance in the histological and molecular process of NASH. Unlike NASH related fibrosis, there are currently no direct biomarkers for
steatohepatitis. Accumulating studies showed that one single biomarker could be insufficient to discriminate bland steatosis and NASH.

**Advances in imaging-based approaches**

Ultrasonography is the most widely used imaging tool due to its accessibility and low cost, while it lacks specificity for identifying the etiology of liver disease. In patients with mild to moderate steatosis, the accuracy of ultrasonography would decrease to around 50%\(^7\). Thus, quantitative ultrasound-based techniques are being developed to improve the diagnosis of hepatic steatosis. Attenuation coefficient (AC) and backscatter coefficient (BSC) has shown to be correlated with the severity of hepatic steatosis. In a biopsy-proven study, AC and BSC achieved an accuracy of 61.7% and 68.3% in predicting the steatosis grades, respectively, significantly higher than traditional ultrasonography\(^7\). Recently, another novel method ultrasound-guided attenuation parameter (UGAP) also showed excellent ability to distinguish mild steatosis (0.922, 95%CI: 0.870-0.973) in non-B non-C chronic hepatitis subjects\(^7\). Controlled attenuation parameter (CAP) is often equipped with the transient elastography (TE) devices for liver fat quantification, which is widely studied and clinically available. CAP showed good sensitivity in detecting mild steatosis (S1), and have excellent diagnostic accuracy in distinguishing S1, S2, and S3 referring to liver biopsy\(^80,\ 81\). In terms of incidence and resolution of steatosis, CAP is also sensitive to reflect the dynamic changes\(^8\). Although CAP reduces the sampling error by increasing the detection volume (3cm\(^3\)), the accuracy would be reduced by the thickness of subcutaneous adipose.
Among magnetic resonance (MRI)-based biomarkers, MRS is sensitive to small amount of hepatic adipose and is recognized as the most accurate noninvasive method to quantify steatosis. MRS is often used as the reference in assessing other noninvasive markers\textsuperscript{52}. Due to the professionalism of MRS, the wide application is limited. Therefore, MRI-proton density fat fraction (PDFF) is more accessible in most tertiary health centers currently. MRI-PDFF could assess the fat content in the whole liver and also allow for the assessment of regions of interest (ROI). Multiple studies have proved the close agreement with histological steatosis grades\textsuperscript{83, 84}. The liver fat content measured by MRS or MRI-PDFF is continuous variable, which could reflect the dynamic changes of hepatic steatosis. MRI-PDFF also showed the absolute and relative the liver fat content. MRI-PDFF showed better diagnostic accuracy than CAP in a head-to-head comparison\textsuperscript{81}.

CT assesses the liver fat content through the absolute attenuation of liver parenchyma value\textsuperscript{85}. CT is more sensitive to moderate-to-severe steatosis than mild steatosis. The sensitivity for detecting grade≥2 steatosis is up to more than 90%. Although CT is not routinely used to identify steatosis, it would be important in detecting incidental steatosis.

Regarding fibrosis, TE is the simplest and the most commonly used noninvasive imaging tool for screening in clinics. The cut-off values of liver stiffness measurement (LSM) by TE identifying advanced fibrosis various with different etiology. For NAFLD, a recent study determined a cut-off of 6.5 kPa to rule out and a cut-off of 12.1 kPa to rule in advanced fibrosis\textsuperscript{86}. In a study enrolling Asian NAFLD patients, the cut-off to
rule out advanced fibrosis is 7.9 kPa and the cut-off to rule in advanced fibrosis is 9.6 kPa\textsuperscript{87}. LSM is sensitive to advanced fibrosis and cirrhosis, while the specificity of ruling out F1 and F2 fibrosis is still to be improved. In addition, LSM would be affected by various factors including obesity, subcutaneous fat thickness, high ALT level and cholestasis\textsuperscript{88}. Agile 3+ and Agile 4 combining LSM with routine clinical parameters were two recently developed models to identify advanced fibrosis and cirrhosis respectively. Both Agile 4 and Agile 3+ showed better diagnostic performance especially positive predictive value than FIB-4 and LSM\textsuperscript{89}. Acoustic radiation force imaging (ARFI) is developed from a chronic hepatitis C patients cohort to diagnose advanced fibrosis. The efficacy of ARFI, supersonic shear imaging (SSI) and TE was compared in a head-to-head study. Similar to TE, the applications of ARFI and SSI are limited in obese subjects, while SSI showed higher accuracy than ARFI for diagnoses of F2 fibrosis\textsuperscript{90}.

Magnetic Resonance Elastography (MRE) is equipped with special hardware added to MRI, to assess liver stiffness. Both MRE and TE had excellent diagnostic accuracy of stage F2-F4 fibrosis with a C-statistic higher than 0.9\textsuperscript{91}. Several studies have revealed a better accuracy of MRE compared with TE\textsuperscript{81, 91, 92}. MRE also had a higher success rate than TE in obese patients (95.8% vs. 88.5%). In a recent meta-analysis, MRE showed higher C-statistic in detecting F\geq2 and F\geq3, while in detecting cirrhosis, MRE was similar with TE and shear wave elastography (SWE)\textsuperscript{93}. The combination of MRI with other imaging tests and biomarkers would increase diagnostic performance. MEFIB is the combination of MRE and FIB-4, which showed a relatively
high PPV of 97.1% in assessing ≥stage F2 fibrosis, with a PPV of 91.0%. The MRI-aspartate aminotransferase (MAST) score was a combination of MRI and NFS, FIB-4 and FibroScan-aspartate aminotransferase (FAST). MAST had a higher C-statistic than that of the components, reducing the number of the patients in “gray zone”\textsuperscript{95}.

5. Dynamic monitoring and prognosis risk assessment

Definition and biomarkers of at-risk NASH

Given the large burden brought by NASH, novel drug development would become a very active field. Currently, the emerging treatment mostly targets hepatic fibrosis and steatohepatitis-associated inflammatory activity. In this case, we should select patients who are at-risk of disease progression to be included in clinical trials, and use effective tests to assess the drug response repeatedly. The Liver Forum defined NAFLD subgroups as the following phenotypes: NAFL, indeterminate NASH, NASH without fibrosis, NASH with early fibrosis, NASH with bridging fibrosis, compensated cirrhosis, and decompensated cirrhosis\textsuperscript{96}. As a number of biopsy-proven studies showed that both fibrosis stage and NAS at baseline were correlated with a higher risk of increased fibrosis stage during follow-up. Recently, Harrison et al. raised the definition of “at-risk NASH”, defined as NAS ≥4 and fibrosis stage ≥2 in patients with NAFLD\textsuperscript{97}. Followed by this definition, many studies raised noninvasive solutions to distinguish these patients.

MACK-3 is a combination of AST, homeostasis model assessment (HOMA) and CK18, showing high accuracy in at-risk NASH patients (NAS ≥4 and F ≥2)\textsuperscript{98}. The cutoffs ≤0.134 and ≥0.550 could be used to rule out and rule in these patients who need
more aggressive drug intervention\textsuperscript{99}. The algorithm ADAPT mentioned before was also effective in detecting these at-risk patients\textsuperscript{100}. In terms of imaging tests, a recent study compared the diagnostic performance in detecting at-risk NASH patients among MEFIB, MAST, and FAST. All three models provide utility in NAFLD risk stratification, while MEFIB showed better performance in detecting at-risk NASH than MAST and FAST\textsuperscript{101}. Direct correlation to the severity of inflammation was previously regarded as the bottleneck of imaging tests, currently corrected T1 (cT1) showed potential in predicting NASH. cT1 had better diagnostic accuracy (0.78 vs. 0.69) in identifying high-risk NASH, compared with MRI-PDFF\textsuperscript{102}. Furthermore, a protein-based signatures of fibrosis could also serve as diagnostic tools. A disintegrin and metalloproteinase with thrombospondin motifs like 2 (ADAMTSL2) protein and an 8-protein panel showed the predictive value for at-risk NASH\textsuperscript{103}.

Biomarkers of treatment response and clinical outcomes

The best clinical outcome to evaluate the efficacy of NASH treatment is the liver-related morbidity and mortality, while the surrogate endpoint is histologic outcome. Current guidelines recommend histological NASH resolution without worsening of fibrosis or regression of fibrosis without worsening of NASH as the treatment endpoint in phase 3 trials of NASH\textsuperscript{11}. The reliance on histologic outcomes for primary trial endpoints may become a barrier to patient enrollment. There is an urgent to develop accurate noninvasive markers dynamic to drug-induced changes. Markers or algorithms which reflect disease severity or long-term prognosis could be utilized as surrogate endpoints for clinical trials in NASH (Figure 1).
A considerate part of the noninvasive markers reported dynamic changes associated with histological changes. Of them, imaging-based tests showed the potential to be the surrogates of histological assessment of steatosis grade and fibrosis stage. As early as in the FLINT trial of obeticholic acid (OCA), MRI-PDFF has served as a surrogate marker of steatosis. Taking a 30% relative reduction in MRI-PDFF as an endpoint, OCA was better than the placebo in achieving the goal. In addition, the non-responders also had lower histological improvement (19% vs. 50%). Patented ELF and PIIINP were also used as serum markers for efficacy in the PIVENS Trial. ELF showed a significant correlation with advanced fibrosis in patients with NASH, while not with the longitudinal changes of fibrosis. As mentioned above, the ML-based methods could translate histological characteristics into continuous variables. For instance, collagen proportionate area (CPA) assessed by digital image analysis may offer a more granular assessment of fibrosis. Small changes detected by CPA might be missed when comparing fibrosis stages. Furthermore, the use of ML-based histological assessment is worth evaluation as a surrogate endpoint in clinical trials.

6. Health-related quality of life and extrahepatic outcomes measures

NASH patients often have concomitant extrahepatic diseases, such as obesity, dyslipidemia, hypertension, T2DM, cardiovascular disease and chronic kidney disease. Recently, the nomenclature of MAFLD raised the attention of metabolic disorders in NAFLD patients. In obese NASH patients, the diagnostic accuracy of noninvasive markers should be modulated. We also investigated the diagnostic value of metabolic disorders in NASH fibrosis. As a significant driver of disease progression, insulin
resistance has been proven to play an essential role in developing steatohepatitis and fibrosis. Although treatment may benefit these comorbidities, there is insufficient evidence to take metabolic comorbidities as one of the trial endpoints. Compared with cirrhotic patients, non-cirrhotic NASH would have a higher non-cirrhotic NASH incidence of cardiovascular disease\textsuperscript{110}. In this case, metabolic-related events should be closely monitored, while longer follow-up periods are required to observe liver-related outcomes.

It is well-known that NAFLD not only increases the risk for development of hepatic and extrahepatic outcomes, but also causes the impairment of health-related quality of life (HRQoL). In comparison with the healthy controls, patients with NAFLD have decreased HRQoL scores and impaired patient-reported outcomes (PRO), even worse than that in patients with other chronic liver diseases\textsuperscript{111}. The changes of HRQoL and PRO score in NAFLD are associated with hepatic disease severity and its improvement after effective treatment. The HRQoL score declines in order from NAFL, to NASH, then advanced fibrosis and cirrhosis in patients with NAFLD. On the other hand, histological improvement such as reduction of steatosis degree, remission of NASH, decreased NAS, and regression of fibrosis stage after multiple new drugs trial for NASH can ameliorate PRO and HRQoL score. Therefore, the evaluation and monitoring of HRQoL and PRO in NAFLD patients should be encouraged in routine diagnosis and treatment. PRO and HRQoL should be regarded as one of the primary endpoints for the management of NASH and related cirrhosis.

**Summary**
The increasing prevalence of NASH has brought large health economic burdens globally, characterized by excess mortality, adverse clinical outcomes and impairment of patient reported outcome (PRO). Since there are still no effective drugs for NASH treatment, many clinical trials for developing novel drugs have been in progress in the last decade. NASH has a heterogeneous collection of metabolic disorders and slowly progressing features of liver diseases. The challenge in monitoring NASH lies in developing techniques to provide dynamic assessment for comprehensive features. Regarding NASH severity evaluation and efficacy assessment, many noninvasive markers and algorithms are developing with significant progress. A number of serum markers, imaging modalities, and noninvasive algorithms are undergoing investigation. Nevertheless, most of their diagnostic performance, accessibility, and cost-effectiveness need to be improved. Furthermore, the monitoring of NASH should also include PROs and extrahepatic diseases especially metabolic disorders. A comprehensive but individualized surveillance would be available for each patient in the near future. We are convinced that given more effort and cooperation from healthcare systems, researchers, pharmaceutical companies and NASH patients, the management and prognosis of NASH patients would benefit from the advances of monitoring and evaluation system.

References


9. Younossi ZM, Corey KE, Lim JK. AGA Clinical Practice Update on Lifestyle


21. Marengo A, Jouness RI, Bugianesi E. Progression and Natural History of


33. Younossi ZM, Otgonsuren M, Venkatesan C, Mishra A. In patients with non-alcoholic fatty liver disease, metabolically abnormal individuals are at a higher risk for mortality while metabolically normal individuals are not. Metabolism. 2013;62:352-360.


43. Liu F, Goh GB, Tiniakos D, Wee A, Leow WQ, Zhao JM, et al. qFIBS: An Automated Technique for Quantitative Evaluation of Fibrosis, Inflammation,


63. Guillaume M, Moal V, Delabaudiere C, Zuberbuhler F, Robic MA, Lannes A, et al. Direct comparison of the specialised blood fibrosis tests FibroMeter(V2G) and Enhanced Liver Fibrosis score in patients with non-alcoholic fatty liver disease from


70. Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of
non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. Ann Med. 2011;43:617-649.


89. Sanyal AJ, Foucquier J, Younossi ZM, Harrison SA, Newsome PN, Chan WK, et


Table 1. Surveillance markers for steatosis, steatohepatitis and fibrosis of NASH

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Assessment</th>
<th>C-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steatosis</td>
<td>Controlled attenuation parameter (CAP)(^{112})</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Dallas Steatosis Index (DSI)(^{53})</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>MRI-proton density fat fraction (PDFF)(^{113})</td>
<td>0.99</td>
</tr>
<tr>
<td>Steatohepatitis</td>
<td>Cytokeratin 18 (CK18)(^{114})</td>
<td>0.83-0.93</td>
</tr>
<tr>
<td></td>
<td>NAFIC score(^{115})</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Corrected T1 (cT1)(^{102})</td>
<td>0.78</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>Fibrosis-4 index (FIB-4)(^{116})</td>
<td>0.75 for SF</td>
</tr>
<tr>
<td></td>
<td>Liver stiffness measurement (LSM)(^{81})</td>
<td>0.86 for SF</td>
</tr>
<tr>
<td></td>
<td>NAFLD fibrosis score (NFS)(^{116})</td>
<td>0.83 for cirrhosis</td>
</tr>
<tr>
<td></td>
<td>Aspartate aminotransferase (AST) to Platelet Ratio Index (APRI)(^{116})</td>
<td>0.70 for SF</td>
</tr>
<tr>
<td></td>
<td>BARD score(^{117})</td>
<td>0.64 for SF</td>
</tr>
<tr>
<td></td>
<td>ELF score(^{63})</td>
<td>0.70 for SF</td>
</tr>
<tr>
<td></td>
<td>FiberMeter(^{63})</td>
<td>0.79 for AF</td>
</tr>
<tr>
<td></td>
<td>Shear wave elastography (SWE)(^{118})</td>
<td>0.86 for AF</td>
</tr>
<tr>
<td></td>
<td>Acoustic radiation force imaging (ARFI)(^{118})</td>
<td>0.77 for AF</td>
</tr>
<tr>
<td></td>
<td>Magnetic resonance elastography (MRE)(^{91})</td>
<td>0.89 for SF</td>
</tr>
</tbody>
</table>

AF, advanced fibrosis; SF, significant fibrosis.
Specific evaluation should be designed for different phases of NASH. Different assessment is also needed for patients at different stages of NASH. MRE, magnetic resonance elastography; MRI-PDFF, magnetic resonance imaging-proton density fat fraction; NASH, nonalcoholic steatohepatitis; NAS, NASH activity score; AEs, adverse events; CAP, controlled attenuation parameter; US, ultrasound; CK18, Cytokeratin 18; FLI, fatty liver index; NLFS, HSI, hepatic steatosis index; FIB-4, fibrosis-4 index; LAP, lipid accumulation product; NAFLD fibrosis score; ELF test, enhanced liver fibrosis (ELF) test; LSM, liver stiffness measurement; MELD score, model for end-stage liver disease score; HVPG, hepatic venous pressure gradient.