Noninvasive Assessment of Liver Fibrosis in NAFLD

Arun J. Sanyal,[1] Laurent Castera,[2,3] and Vincent Wai-Sun Wong[4,5]

[1]Division of Gastroenterology, Hepatology and Nutrition, Department of Internal Medicine, Virginia Commonwealth University School of Medicine, Richmond, Virginia; [2]UMR1149 (Center of Research on Inflammation), French Institute of Health and Medical Research, Université Paris Cité, Paris, France; [3]Service d'Hépatologie, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, Clichy, France; [4]State Key Laboratory of Digestive Disease, Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong, China; and [5]Medical Data Analytics Centre, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China

Nonalcoholic fatty liver disease (NAFLD) has emerged as a leading cause of liver-related morbidity and mortality worldwide, afflicting approximately a billion individuals. NAFLD is a slowly progressive disease that may evolve in a subset of patients toward cirrhosis, hepatocellular carcinoma, and end-stage liver disease. Liver fibrosis severity is the strongest predictor of clinical outcomes. The emergence of effective therapeutics on the horizon highlights the need to identify among patients with NAFLD, those with severe fibrosis or cirrhosis, who are the most at risk of developing complications and target them for therapy. Liver biopsy has been the reference standard for this purpose. However, it is not suitable for large-scale population evaluation, given its well-known limitations (invasiveness, rare but severe complications, and sampling variability). Thus, there have been major efforts to develop simple noninvasive tools that can be used in routine clinical settings and in drug development. Noninvasive approaches are based on the quantification of biomarkers in serum samples or on the measurement of liver stiffness, using either ultrasound- or magnetic resonance–based elastography techniques. This review provides a roadmap for future development and integration of noninvasive tools in clinical practice and in drug development in NAFLD. We discuss herein the principles for their development and validation, their use in clinical practice, including for diagnosis of NAFLD, risk stratification in primary care and hepatology settings, prediction of long-term liver-related and non–liver-related outcomes, monitoring of fibrosis progression and regression, and response to future treatment.

Keywords: Cirrhosis; Blood Biomarkers; Transient Elastography; Magnetic Resonance Elastography; Nonalcoholic Fatty Liver Disease.

Nonalcoholic fatty liver disease (NAFLD) has emerged as a leading cause of liver-related morbidity and mortality worldwide.[1] Its prevalence ranges from 20%–30% of the adult population and 10% of children.[2] NAFLD has 2 fundamental phenotypes (ie, nonalcoholic fatty liver and nonalcoholic steatohepatitis (NASH)[3]; NASH is more likely to progress to cirrhosis and end-stage liver disease.[4] With the aging of the affected population and longer exposure to the disease, the burden of disease from cirrhosis due to NAFLD is increasing and expected to increase 2- to 3-fold from 2015 to 2030 in many regions of the world.[5,6]

The public health relevance of NASH has spurred intense research and drug-development efforts. Recent studies demonstrate that several classes of compounds may not only cause resolution of NASH, but also improve fibrosis. The emergence of effective therapeutics on the horizon highlights the need to identify those with NAFLD, especially NASH, who are most at risk of development of cirrhosis or have developed cirrhosis and target them for therapy.

The reference standard for the evaluation of NAFLD and determination of prognosis has been the histological examination of liver tissue sections obtained by a liver biopsy.[7,8] However, liver biopsy is associated with discomfort and occasionally with severe morbidity and even mortality.[9] Further histological assessment is limited by sampling variability and both intra- and interobserver variability.[10] There is also a paucity of trained workforce to perform and interpret liver biopsies. For all these reasons, histological assessment is not suitable for large-scale population-level evaluation of patients with risk factors for NAFLD underscoring the need for simple noninvasive tools (NITs) that can be used in routine clinical settings and in drug development.

There are 2 characteristics of NAFLD that provide important information about the disease. The first is the activity of the disease, which refers to the lipotoxic load and subsequent injury to the liver. Histologically, this manifests as steatosis, hepatocellular ballooning, and

inflammation, the key features of steatohepatitis.[11] Disease activity drives a fibrogenic response in the liver with progressive fibrosis and architectural disruption culminating in cirrhosis. Cirrhosis causes portal hypertension a key driver of hepatic decompensation and death. Fibrosis severity thus reflects proximity to cirrhosis and is the strongest predictor of clinical outcomes.[12,13] There are thus major efforts to develop NITs to evaluate fibrosis in the context of NAFLD.[14] This review summarizes these results and provides a roadmap for future development and integration of these NITs in clinical practice and in drug development.

Scientific Principles of NIT Development

There have been major advances in biomarker science that underlie the development of NITs for specific purposes. These can be broadly categorized as follows (Figure 1). The first and most important guiding principle is the context of use (COU).[15] The COU defines the intended use of the NIT and the decisions that will emanate from the results. Such intended uses could be establishing a diagnosis, prognosis (ie, risk of future outcomes based on the test run today, disease monitoring to capture the course of disease progression, treatment response to evaluate response to therapies with specific mechanisms of action, and ultimately a surrogate endpoint.[16] The COU further defines the population in which the NIT will be used and the setting in which it will be used (eg, in a primary care clinic or advanced tertiary clinic). It also defines how the results will be generated and provided back to the person ordering the test and the decisions that will follow. The COU is thus a critical determinant of the type of studies needed, the populations and settings in which the studies are needed, and the design of such studies for NIT development.

The second critical determinant of NIT development is the analytical robustness of the assays or tools to be used. Strictly speaking, a biomarker is a measure of normal physiology or perturbation of physiology; the biomarker can be measured in 1 or more ways, and the method itself is not the biomarker. For instance, liver stiffness is a physical property of liver tissue that can be measured by different elastography techniques. However, the method determines the fidelity and accuracy of the measurement of the biomarker. Some fundamental steps in NIT development are therefore to have a full understanding of the accuracy, reproducibility (same sample, different machines or operators), and repeatability (same patient test-retest) of the assays. The quality criteria of elastography techniques and conditions of sample collection, transport to the laboratory, and sample handling are all also relevant because they define whether the boundaries within which the test results will be accurate and valid.

The third critical element is clinical performance and utility. Clinical performance refers to the ability of the measure to accurately reflect a biological phenomenon (eg, fibrosis stage), whereas clinical utility takes in to account the benefit to the patient and society vs the harm from misclassification. The clinical utility also takes into account if a given NIT outperforms what is currently available and the overall benefits in terms of improved access to care, avoidance of risky procedures, and eventually better outcomes. These are considered against the risk of misclassification, which could include leaving patients with disease at risk of outcomes and unnecessary additional testing and stress to patients who do not have the condition (eg, cirrhosis).

Regulatory Approval and Availability

In the United States, there are several different pathways by which NITs can be introduced into routine care. First, as a drug development tool, in which an individual sponsor evaluates the use of the NIT for a specific purpose in the context of their drug development efforts. They carry the risks and burden of generating the evidence to support the use of the NIT in the context of the use of their drug. This is written into label language and the NIT is not approved for similar contexts with other drugs. The second mechanism is scientific consensus based on large amount of published peer reviewed literature. The third mechanism is the biomarker qualification pathway in which the Food and Drug Administration approves the use of an NIT for a specific context of use, which includes the circumstances

Figure 1. Scientific principles of NIT development. AUROC, area under the receiver-operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

Context of use	Analytic robustness	Clinical validity and utility	
• Purpose of use • Population where it will be used • Setting in which it will be used • What the read-out will be • How the read-out will inform clinical decisions	• What is measured • How it is measured • Conditions within which it can be measured • Accuracy of measurement • Repeatability • Reproducibility • Day-to-day and diurnal variation	• Sensitivity • Specificity • AUROC • Optimal cutpoint • Specificity at 90% sensitivity • Sensitivity at 90% specificity	• PPV and NPV at varying prevalence • Benefits of correct identification • Harm from misclassification

Table 1. Checklist for Design and Interpretation of NIT Studies for NASH

Biomarker properties
    Single analyte or multiple analytes
    Single or multiple modalities
    Assay robustness and conditions within which samples remain viable for the assay
    Imaging quality criteria
    Assay/imaging repeatability and reproducibility
    Diurnal variance
    Systematic differences based on race, ethnicity, gender, age, etc.

Study population
    Does it reflect the population where the NIT is intended to be used?
    Probability of ascertainment bias
    Is there a balanced distribution of disease severity (spectrum bias)?

Study design
    Time lag from biopsy to sample collection (ideally <90 d)
    In comparative studies, are all NITs tested on same sample?
    If diurnal variability exists, is sample collection time and procedure standardized
    Tracking sample chain of custody and handling/storage/transport conditions
    Avoid running reference standard (liver biopsy) on the basis of results of the NIT being tested
    Are all histological sections reviewed using a standard protocol and, ideally, need 3+ pathologists to come to consensus; how are
        pathologists masked to clinical and NIT data and to each other's report?
    Sample loading on analytic instruments (randomized or in batches)
    Power analysis for sensitivity and for specificity

Data analysis
    Report sensitivity and specificity at various cutpoints
    Identify optimal cutpoint
    Report high sensitivity and high specificity cutpoints
    AUROC comparison with existing tools to demonstrate superiority
    Independent validation of novel biomarkers from model building to validation cohort vs bootstrap validation
    Identify predictive values at various population prevalence of disease
    Confirm predictive values in selected populations with specific prevalence of disease

AUROC, area under the receiver-operating characteristic curve; NIT, noninvasive test.

in which it will be used, the population in which it will be used, the fidelity of the readout, the clinical decision making based on the readout, and the potential benefits and harm to patients. The contexts of use can range from diagnostic to treatment response based on the BEST (Biomarkers, EndpointS, and other Tools) classification. A surrogate endpoint requires the highest and most robust level of evidence. Ideally, it must be in the biological pathway leading to an outcome, and a certain degree of change should translate into a predictable degree in change in the risk of a relevant outcome.

In Europe, there is no unified approval process for NITs. Vibration-controlled transient elastography (VCTE) is widely available, whereas few countries have access to patented serum tests of fibrosis. Magnetic resonance elastography (MRE) access remains limited.

In other areas including Latin America, Africa, the Middle East, and the Asia-Pacific region, there is no unified approval process for NITs. VCTE is relatively available in many Latin American and Asia-Pacific countries, whereas few countries have access to patented serum tests of fibrosis. MRE can be reimbursed in Japan and Singapore, but access in other countries remains limited.

How to Design and Report an NIT Validation Study

The first step in designing an NIT development study is to consider if the biomarker assay is reliable (Table 1). The analytic robustness and variability are essential determinants to define how much of a change is a true and reliable change. When biomarker panels are being tested, the analytical characteristics of each analyte and modality should ideally be known before embarking on an NIT validation study. Unfortunately, this is often not the case.

NIT validation can be performed both in prospective and in retrospective studies. The study population should reflect the intended-use population for the NIT as defined by its COU.[15] The outcome of interest against which the performance of the NIT will be tested must be clearly defined. In the context of NASH, this is usually a clinical outcome or its surrogate (ie, histological assessment of disease activity or fibrosis), the current reference standard for the purpose.[8] For fibrosis, one must further distinguish between biomarkers that measure susceptibility to develop fibrosis (eg, *PNPLA3* or *HSD17B13* gene mutation), fibrogenesis (eg, α-smooth

muscle actin or PRO-C3), fibrosis burden (eg, Enhanced Liver Fibrosis [ELF] test or elastography), or fibrolysis (eg, PRO-C6).[17–19] This determines the reference standard that the NIT is designed to reflect.

There are several potential sources of bias in the design of NIT studies, and interested readers are referred to several reviews on the subject.[20,21] Ascertainment bias is defined by the nature of the clinic and may enrich or de-enrich the study population with the disease phenotype of interest. Even more important is spectrum bias[22]; simply put, if most of the study population has the phenotype of interest, any NIT will have a high rate of detection and vice versa. It is therefore important to power the study both to rule in and to rule out the disease with high specificity and sensitivity.[23] The process for sample handling and laboratory testing or the protocols for imaging should be harmonized across sites and the concordance across observers documented.

From an analytical point of view, the reference standard is very important. This is often histological scoring of fibrosis in NASH that is limited by sampling error, reading variability, and a limited dynamic range of fibrosis. There are several options for evaluation of a new biomarker when the reference standard is imperfect including consensus by expert opinion, development of a subjective or derived score that the reference standard is correct, or a covariance model with other noninvasive tests measuring the same biological process.[24,25] The time gap between the liver biopsy and the collection of the biosample or performance of test is another important variable and should be kept to a minimum. For assessment of prognostic biomarkers in NASH, the ideal study is to relate NITs to future development of clinical outcomes and the underlying fibrosis stage. The ELF test is the only approved prognostic NIT for NASH, and its approval was based on such an approach.[19]

There are also ongoing attempts to combine multimodality, multianalyte biomarkers to create composite models (eg, the Fibrometer-VCTE and MEFIB.[26,27] A key consideration in such models is to ascertain the contribution of each modality to the overall performance of the model. For models dependent on ratios of analytes, additional analyses on the diagnostic performance of the ratio when the numerator changes or when the denominator changes are required to fully understand how the biomarker performs.

Finally, given the plethora of published papers on NITs that do not meet quality metrics and standards for rigor of the science or transparency of reporting, there is an urgent need for high-quality studies that can be used as high-level evidence to support clinical decision making and for drug development. Several standards for such reporting exist and we recommend that all papers on NITs meet the STARD or TRIPOD standards for reporting of the data.[28,29]

Specific Use

How to Diagnose NAFLD?

The first step in the evaluation is to confirm the diagnosis of a chronic liver disease. In the case of NAFLD, the diagnosis is based on the demonstration of hepatic steatosis and exclusion of other chronic liver diseases, excessive alcohol consumption, and secondary causes of hepatic steatosis. In routine practice, abdominal ultrasonography is the most commonly performed test to detect hepatic steatosis based on bright liver echotexture, deep attenuation of the ultrasound signal, and vascular blunting.

A number of steatosis scores such as the fatty liver index, hepatic steatosis index, U.S. fatty liver index, and NAFLD ridge score were developed for the prediction of hepatic steatosis.[30] These scores are all based on liver enzymes and metabolic risk factors and thus represent association rather than direct measurement of hepatic steatosis. The accuracy may be insufficient for individual case management, but as their calculation is essentially free of charge, the scores can easily be used in existing databases to study the epidemiology of NAFLD.

Controlled attenuation parameter measurement by VCTE estimates the degree of hepatic steatosis through the determination of ultrasound attenuation in the liver. It is supposed to be more sensitive than abdominal ultrasonography in detecting mild steatosis, but head-to-head comparisons are lacking. Its area under the receiver-operating characteristic curve for various steatosis grades is around 0.8.[31,32] Magnetic resonance imaging proton density fat fraction (MRI-PDFF) is considered the gold standard to quantify hepatic steatosis but is often reserved for research settings because of cost and availability.

Risk Stratification in the Primary Care and Nonhepatology Setting

Although most studies on NITs came from tertiary hepatology centers, it is important to recognize that most patients with chronic liver disease are seen at primary care and nonhepatology settings. This is particularly true for NAFLD and alcohol-related liver disease, which often call for a multidisciplinary approach.[33] The need for liver assessment is usually prompted by abnormal liver blood tests, a history of liver disease, or risk factors of liver disease such as harmful drinking, metabolic conditions, or at-risk behaviors.

Availability, test performance, and cost govern the choice of NITs in primary care and nonhepatology settings. Simple fibrosis scores such as the Fibrosis-4 index (FIB-4) and NAFLD Fibrosis Score (NFS) can be calculated with routine clinical and laboratory parameters and thus add little to the healthcare expense (Table 2).[30] Overall, these scores have a high negative predictive

Table 2. Simple Fibrosis Scores for Primary Care and Nonhepatology Settings and Specific Fibrosis Biomarkers

Test	Components or Mechanism	Accuracy	Monitoring	Limitations
Simple fibrosis scores				
AST-to-ALT ratio	AST, ALT	AUROC 0.66–0.74 for F3–F4	Insufficient data	False positive in patients with normal ALT
APRI	AST, platelets	AUROC 0.74 for F3–F4	Insufficient data	False positive in patients with alternative causes of thrombocytopenia
Fibrosis–F4 index (FIB-4)	Age, AST, ALT, platelets	AUROC 0.84 for F3–F4	Rising FIB-4 over time associated with increased risk of cirrhosis, decompensation, and hepatocellular carcinoma	Low accuracy overall in patients younger than 35 y; low specificity in those older than 65 y
NAFLD fibrosis score	Age, body mass index, hyperglycemia, AST, ALT, platelets, albumin	AUROC 0.82 for F3–F4	Insufficient data	Low accuracy overall in patients younger than 35 y; low specificity in those older than 65 y; restricted use to NAFLD
Blood biomarkers				
Single marker				
PRO-C3	Reflect true synthesis of type III collagen	AUROC 0.73 for F3–F4	Modest-to-moderate correlation with changes in histological fibrosis in clinical trials	Most data from NAFLD
WFA$^+$-M2BP	Changes in N-glycosylation of Mac-2 binding protein during liver injury	AUROC 0.82 for F3–F4	Insufficient data	Costly; not widely available
Combination markers				
ELF score	PIIINP, hyaluronic acid, TIMP-1	AUROC 0.83 for F3–F4	Associated with liver-related outcomes; monitoring role to be determined	Less useful for early fibrosis; costly; not widely available
FibroTest	GGT, total bilirubin, $\alpha2$ macroglobulin, apolipoprotein AI, haptoglobin	AUROC 0.88 for F3–F4	Correlates with fibrosis improvement after treatments for chronic viral hepatitis	Less useful for early fibrosis; costly; most data from viral hepatitis
FibroMeter	Depending on the versions for viral hepatitis or NAFLD, may include age, sex, body weight, prothrombin index, ALT, AST, GGT, ferritin, glucose, platelets, urea	AUROC 0.94 for F2–F4	Data on correlation with histological changes and clinical outcomes scarce	Costly; not widely available
Hepascore	Age, sex, total bilirubin, GGT, hyaluronic acid, $\alpha2$ macroglobulin	AUROC 0.81 for F3–F4	Associated with liver-related outcomes; monitoring role to be determined	Not widely available
Imaging biomarkers				
Ultrasound elastography				
VCTE	Measures the velocity of an elastic shear wave that propagates across the liver	AUROC 0.9 for F3–F4	Associated with liver-related outcomes; some correlation with regression and progression of fibrosis and portal hypertension but confounded by the degree of necroinflammation	Confounded by active hepatitis, food intake, congestive heart failure, biliary obstruction, amyloidosis, and possibly the degree of hepatic steatosis; less applicable and reliable in severe obesity

Table 2. Continued

Test	Components or Mechanism	Accuracy	Monitoring	Limitations
Point shear-wave elastography	Detects localized tissue displacement by ultrasound wave	AUROC 0.8–0.9 for F3–F4	Insufficient data	Reliability criteria not well defined; probably affected by the same confounders as VCTE, though success rate is higher than VCTE in obese patients
2-dimensional shear-wave elastography	Captures propagation of shear waves in real time	AUROC 0.80–0.98 for F3–F4	Insufficient data	Probably affected by the same confounders as VCTE
Magnetic resonance elastography	Images propagation of shear waves in the liver	AUROC 0.89–0.96 for F3–F4	Associated with liver-related outcomes; some correlation with regression and progression of fibrosis and portal hypertension but confounded by the degree of necroinflammation	Probably affected by the same confounders as VCTE and iron content; costly; not widely available; some patients may have contraindications to magnetic resonance imaging

ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; AUROC, area under the receiver-operating characteristic curve; ELF, Enhanced Liver Fibrosis; GGT, γ-glutamyltransferase; NAFLD, nonalcoholic fatty liver disease; PIIINP, procollagen III amino-terminal peptide; TIMP-1, tissue inhibitor of metalloproteinases 1; VCTE, vibration-controlled transient elastography; WFA+-M2BP, *Wisteria floribunda* agglutinin-positive Mac-2 binding protein.

value of over 80%–90% in excluding advanced fibrosis, but their positive predictive value is modest at best, especially when applied in low-risk populations in whom the pretest probability of advanced fibrosis is low (Figure 2).[34] False positive results are also more common in individuals with normal alanine aminotransferase (ALT) levels, suggesting that their application should be restricted to patients with known or risk factors of liver disease.[35] Patients with normal fibrosis scores have very low risk of cirrhotic complications and hepatocellular carcinoma and can be safely managed by primary care.[36]

Patients with abnormal fibrosis scores need further assessments to confirm the presence of advanced fibrosis. The choice would depend heavily on the local healthcare setting and availability of tests. A number of studies have confirmed the feasibility of incorporating VCTE examination in diabetes practices.[37] The ELF test, a proprietary panel of 3 specific fibrosis biomarkers, is available in the United Kingdom and is recommended by the National Institute for Health and Care Excellence for fibrosis assessment. The U.S. Food and Drug Administration has also approved the ELF test as a prognostic biomarker in NASH.

Several prospective studies have examined the adoption of NITs in primary care. In the Camden and Islington NAFLD Pathway, primary care physicians performed FIB-4 as the first step in patients with raised ALT or fatty liver on ultrasonography, followed by the ELF test if FIB-4 showed indeterminate results of 1.30–3.25.[38] The pathway increased the detection of advanced fibrosis by 5-fold and reduced the number of inappropriate referrals to hepatologists. The SEAL

program in Germany screened people older than 35 years of age with the aspartate aminotransferase (AST)-to-ALT ratio followed by the AST-to-platelet ratio index (APRI).[39] Among 11,859 participants, 488 (4.1%) had abnormal APRI, and 45 incident cases of advanced fibrosis were identified. The high percentage of false positives and missed appointments highlights the need to refine the clinical care pathway further.

Risk Stratification in Hepatology Settings

While primary care serves as initial screening to exclude advanced liver disease, hepatologists need to determine the severity of liver disease with a greater accuracy, predict future liver-related complications and make treatment decisions. Liver fibrosis is a dynamic process. Therefore, hepatologists also need NITs to monitor for disease progression and treatment response. Simple fibrosis scores outlined in the last section still have value in hepatology settings if not already done. Numerous studies have demonstrated the role of a 2-step approach.[40] In essence, the first test (usually simple fibrosis score) serves to exclude patients with a low likelihood of advanced fibrosis and thus enrich a cohort with a higher pretest probability so that the second test will achieve a much higher positive predictive value in confirming the presence of advanced fibrosis.

Specific fibrosis biomarkers are primarily blood or imaging based. Most blood biomarkers are not accurate enough when used in isolation, so it is necessary to combine several biomarkers in patented formulae to improve the diagnostic accuracy, as is true for the ELF test, FibroTest, and FibroMeter (Table 2). Other blood
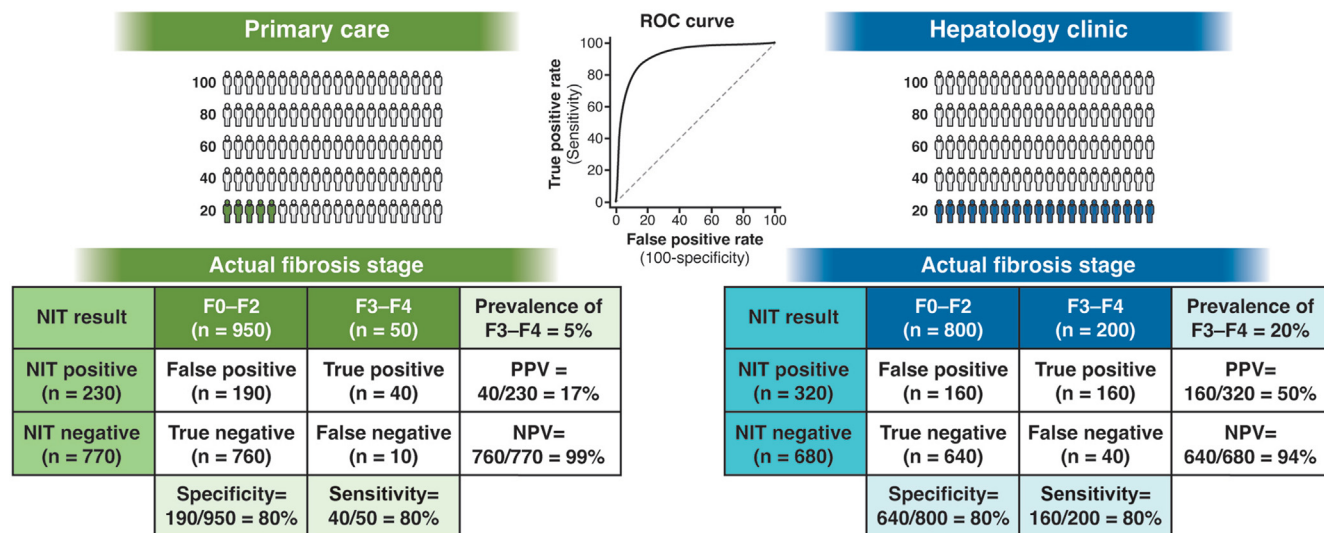
Figure 2. Impact of the prevalence of F3–F4 fibrosis on the performance of NITs. In this example, the sensitivity and specificity of an NIT remain static at 80% regardless of the context. In primary care settings (left) with a prevalence of F3–F4 fibrosis of 5%, the number of patients with false positive results increases with the total number of patients with F0–F2 fibrosis, and this leads to a much lower positive predictive (PPV) value than what would be observed in specialist settings (right), where most noninvasive tests were developed and validated. NPV, negative predictive value; ROC, receiver-operating characteristic.

biomarkers such as the PRO-C3 collagen neo-epitope biomarker and *Wisteria floribunda* agglutinin-positive Mac-2 binding protein have been used in isolation for fibrosis assessment with reasonable performance.

In hepatology settings, imaging-based methods including ultrasound elastography (VCTE, point shear-wave elastography, 2-dimensional shear-wave elastography) and MRE are usually used for better estimation of the fibrosis stage.[30] In head-to-head comparisons, MRE has a higher applicability and accuracy than VCTE, but its wider use is limited by cost and availability.[41]

The latest Baveno VII consensus recommends the use of the rule of 5 to assess the severity of liver disease based on VCTE results.[42] A liver stiffness of <10 kPa and ≥15 kPa can be used to exclude and rule in compensated advanced chronic liver disease, respectively. Patients with liver stiffness <15 kPa and normal platelets are unlikely to harbor clinically significant portal hypertension (CSPH) (defined as hepatic venous pressure gradient ≥10 mm Hg), whereas one can assume CSPH when liver stiffness is over 25 kPa. In addition, patients with liver stiffness <20 kPa and normal platelets seldom have large varices and can avoid endoscopy. Though simple to use, the rule does not consider that optimal cutoffs may differ among different etiologies. The accuracy of the rule for CSPH may also be lower in obese patients with NAFLD.

A few scores have combined imaging biomarkers and simple laboratory assays to improve the diagnostic accuracy. In NAFLD, it is important to diagnose fibrotic NASH (NAFLD activity score ≥4 and fibrosis stage ≥2) because this is the inclusion criteria in most biopsy-based clinical trials, and it is believed that the same NITs can be used to identify patients for treatment when a drug is approved for NASH in the future. The FAST

(FibroScan-AST) score combines controlled attenuation parameter and liver stiffness measurement by VCTE and AST level and has an area under the receiver-operating characteristic curve of around 0.8 for the detection of fibrotic NASH.[43] The MAST (MRI-AST) score is based on a similar concept and replaces VCTE with MRI-PDFF and MRE.[44] The MEFIB index combining MRE and FIB-4 is able to achieve a positive predictive value of >90% in ruling in F2–F4 fibrosis in NAFLD.[27]

Genetic Biomarkers and Risk Scores

Genetic determinants of liver fibrosis have been more extensively evaluated in patients with NAFLD and alcohol-related liver disease.[45] In the past decade, a series of genomic and candidate gene studies have identified a number of single nucleotide polymorphisms (SNPs) conducive to the development of hepatic steatosis and liver injury, among which gene polymorphisms of *PNPLA3*, *TM6SF2*, *GCKR*, and *MBOAT7* have been most extensively evaluated across multiple ethnicities. In contrast, the protein-truncating variant of *HSD17B13* reduces the risk of NAFLD, alcohol-related liver disease, and cirrhosis.[46] The distribution of some of these genetic variants, most notably that of *PNPLA3*, in part explains the ethnic and geographical patterns of liver disease.[47] These include the higher prevalence of fatty liver among Hispanics and lower prevalence among Blacks in the United States, and the relatively high prevalence of fatty liver in East Asia despite a lower metabolic burden.

There have been early efforts to combine these known genetic markers for disease prediction. For example, the Polygenic Risk Score-5 combines the 5 SNPs mentioned previously to predict hepatocellular

carcinoma in a European NAFLD cohort and the general population in the UK Biobank.[48] Likewise, a combination of *PNPLA3*, *SUGP1-TM6SF2*, and *HSD17B13* gene polymorphisms and diabetes predicted cirrhosis risk in heavy drinkers.[49] However, although the discrimination is certainly there, these genetic scores miss a significant proportion of patients with advanced liver disease, and their incremental value over existing clinical risk factors and NITs of fibrosis remains uncertain. Besides, existing works largely focused on SNPs; the role of other genetic markers such as epigenetic changes deserves further studies.

Predictive Biomarkers

Available evidence suggests that NITs can have a role in predicting long-term outcomes in NAFLD patients, including liver-related events, hepatocellular carcinoma (HCC), and mortality.[50,51] Among non-patented serum markers, FIB-4, NFS, and APRI have been the most extensively studied.[52–58] A recent meta-analysis (13 studies in 9001 NAFLD patients)[59] demonstrated their ability to predict liver-related morbidity and mortality, with a level of performance that met or exceeded that of a liver biopsy.[53] For mortality, FIB-4 (area under the curve [AUC], 0.67–0.82) and NFS (AUC, 0.70–0.83) outperformed APRI (AUC, 0.52–0.73) in all studies. Another meta-analysis (4 studies in 6324 patients) reported their ability to predict all-cause mortality.[60] It should be kept in mind, however, that most of these studies were from secondary/tertiary referral centers, retrospective, and with highly variable follow-up (1–20 years). Thus, these data cannot necessarily be extrapolated to primary care even though the predictive value of these NITs for mortality has been reported in the general population.[61,62] Among patented tests, the ELF test, at a cutoff of 11.27, predicted the onset of clinical events in the simtuzumab trial (250 compensated cirrhotic NAFLD patients and 31 months median follow-up).[19] In a combined analysis of the simtuzumab trial and the phase 3 STELLAR trials testing selonsertib, the ELF test predicted progression to cirrhosis in F3 patients and cirrhotic complications in F4 patients; its change also correlated with changes in patient-reported outcomes.[63]

VCTE is the elastography technique with the highest level of evidence. Liver stiffness measurement (LSM) by VCTE has been consistently shown to predict hepatic decompensation, HCC, and liver-related mortality.[54,64–66] For instance, in a large multicenter study (1039 NAFLD patients with a median follow-up of 3 years),[66] baseline LSM was independently associated with occurrence of hepatic decompensation (hazard ratio [HR], 1.03; 95% confidence interval [CI], 1.02–1.04; $P < .001$), HCC (HR, 1.03; 95% CI, 1.00–1.04; $P < .003$), and liver-related death (HR, 1.02; 95% CI, 1.02–1.03; $P < .005$). In contrast, LSM could not predict cardiovascular events or extrahepatic cancers,[65,66] and results for prediction of overall mortality have been conflicting.[54,65–67]

The ability of LSM by MRE to predict liver-related outcomes in NAFLD has been reported less extensively.[68–70] In a recent meta-analysis (6 retrospective studies in 1707 patients with a median follow-up of 3 years),[71] MRE was able to predict liver-related events (HR, 15.9; 95% CI, 9.32–27.2; $P < .001$) for LSM >8 kPa, compared with those with LSM <5 kPa.

Finally, several composite scores, such as Agile 3+ (age, sex, AST, ALT, platelet count, diabetes status, and VCTE)[72] and MEFIB (MRE and FIB-4)[71] or the European Association for the Study of the Liver algorithm (FIB-4 followed by VCTE)[73] have been shown to predict liver-related-events.

Further prospective studies in large multicentric cohorts with longer follow-up are needed, and the use of an artificial intelligence–based algorithm may be helpful to improve prediction in the future.

Disease Monitoring Biomarkers

Dynamic changes in NITs over time may be used for several purposes: (1) to monitor progression of fibrosis to cirrhosis or its regression; or (2) to refine stratification of risk of liver-related events. In a retrospective longitudinal study in 292 NAFLD patients with paired liver biopsies (median time interval 2.6 years), changes over time of APRI, FIB-4, and NFS were significantly associated with fibrosis progression to advanced fibrosis (C-statistic of 0.82 for APRI, 0.81 for FIB-4, and 0.80 for NFS).[7] However, in another recent retrospective study in 133 NAFLD patients with paired biopsy or VCTE with a longer follow-up (12.6 ± 8.5 years), changes of FIB-4, NFS, and APRI were only weakly associated with disease progression.[74] Finally, the ELF test, at a cutoff of 9.76 (sensitivity 77%, specificity 66%), could predict progression to cirrhosis in patients with F3 fibrosis.[19]

As for fibrosis regression, in 200 NAFLD patients enrolled in a placebo-controlled trial with paired liver biopsies 72 weeks apart, reductions in APRI but not in FIB-4 or NFS were significantly correlated with fibrosis improvement at week 72 ($P = .012$).[75] Similarly, among 1135 NAFLD patients with cirrhosis enrolled in 2 large placebo-controlled trials with paired liver biopsies 48 weeks apart, those with cirrhosis regression (16%) had a significant decrease in ELF test and LSM by VCTE over time and a decrease in liver-related events.[76]

Data for refining risk stratification remain limited. In a large population-based Swedish study (40,729 individuals), FIB-4 progression over a mean time of 2.4 years, from a low-risk (<1.30) or intermediate-risk (1.30–2.67) group to a high-risk group (>2.67), was associated with an increased risk of severe liver disease, defined by International Classification of Diseases codes (adjusted HR, 7.99 and 8.64, respectively).[36] In 533

NAFLD patients with advanced fibrosis, increase in LSM by VCTE (>20% from baseline), over a median time of 37 months, was independently associated with hepatic decompensation, HCC, overall mortality, and liver-related mortality (HR, 1.96).[66]

Further prospective studies are needed to assess the impact of dynamic changes in NITs on long-term outcomes. Whether decrease in NITs values over time are associated with improvement in long-term outcomes remains to be demonstrated. As shown in viral hepatitis, inflammation, affecting NITs using transaminases, such as FIB-4, NFS, and APRI, as well as LSM, whatever the technique, is a major confounding factor.[50] In practice, VCTE is probably the most attractive tool in terms of efficacy and cost-effectiveness. The optimal time frame remains to be defined, but VCTE once a year seems reasonable.

Pharmacodynamic of Response Biomarkers

In the hierarchy of intended uses, treatment response biomarkers require a high burden of evidence compared with diagnostic, prognostic, and disease monitoring biomarkers that are often foundational to the development of treatment response biomarkers.

The biological plausibility of the treatment response biomarker is a fundamental element of NIT development. The treatment response biomarker should be in the disease pathway in which it is studied and reflects target engagement by a drug based on its mechanism of action or reflect improvement in underlying biology of the disease (eg, overall fibrosis burden in the liver. Failure to consider this is a common error (eg, the use of FIB-4, which is a surrogate measure of fibrosis burden and not a measure of fibrogenesis to evaluate the antifibrogenic response to treatment).

The analytical aspects of treatment response biomarkers include its dynamic range and sensitivity to detect a change. Here, the reproducibility and repeatability data are again relevant because they provide the confidence limits to define when a change can be considered a true change from baseline. From an NIT performance perspective, such information is required to further define how much of a change in NIT value is linked to a certain change in likelihood of a clinical outcome or change in reference surrogate for the outcome (eg, histological fibrosis stage for NASH). These can be used to identify specific cutoffs based on the measurement of a biomarker at a specific point in time after initiation of therapy to identify treatment response vs nonresponse. There are no approved treatment response NITs at this time, but it is hoped that the large outcomes trials being completed will provide this information.

A major challenge in development of treatment response NITs for fibrosis, which is itself a surrogate measure of outcomes risk, is the degree of variability in assessment of fibrosis stage using conventional approaches on one hand and the relative inability of NITs to be concordant to such measures of fibrosis. Recently, multiple NITs were measured in the context of a clinical trial and related to histological measures of treatment response and shown as a heat map. These identified not only full responders who had both histological and NIT response, but also populations that were responders by NIT criteria but not histological criteria. Long-term trials are needed to determine if these individuals indeed have clinical responses consistent with a treatment response or if they track the changes in histology.

Several classes of NITs are currently being evaluated as treatment response biomarkers. In early-phase trials, depending on the mechanism of action, a decrease in liver fat and ALT are commonly used tests. A 30% decrease in MRI-PDFF is often considered an industry standard as a predictor of histological response.[77] This is likely to be mechanism of action specific and it is unlikely that primary antifibrotic therapies will show such a response. An alternate approach is to use a waterfall plot to demonstrate the proportion, degree, and distribution of change in MRI-PDFF, which provides a more holistic view of the overall de-fatting response of a drug with an mechanism of action expected to reduce hepatic steatosis. Similarly, a 17-unit or greater decrease in ALT has also been used to define treatment response in studies.[78]

There are no validated measures of fibrosis change in early phase trials that can provide assurance of success in more advanced phase trials. In a secondary analysis of the 72-week interim data of the phase 3 REGENERATE study, fibrosis improvement during obeticholic acid treatment was associated with a reduction in ALT, AST, FIB-4, FibroTest, and ELF test , though cutoffs and diagnostic performance for fibrosis change were not defined.[79] Recently, a 0.5-unit decrease in the ELF score has been suggested as a treatment response biomarker and is undergoing additional validation.[76] Whether a reduction in the results of an NIS4 or NIS2 test to values below the threshold for at-risk NASH or a similar reduction of the ELF test, Agile 3 or 4 results in a reduction in clinical outcomes are important future studies. The utility of PRO-C3, a measure of fibrogenesis,[80] as a treatment response biomarker also requires further validation.

While a decrease in 2-dimensional MRE measures of liver stiffness have been used to also assess treatment response,[81] the background variance in reproducibility and repeatability requires large changes to be considered a true change especially in tightly powered trials. A corrected T1 measurement has also been used to evaluate treatment response.[82] Additional data, particularly after correction for steatosis, are needed to determine if these reflect improvement in fibrosis and inflammation and in the long-term reduced risk for outcomes.
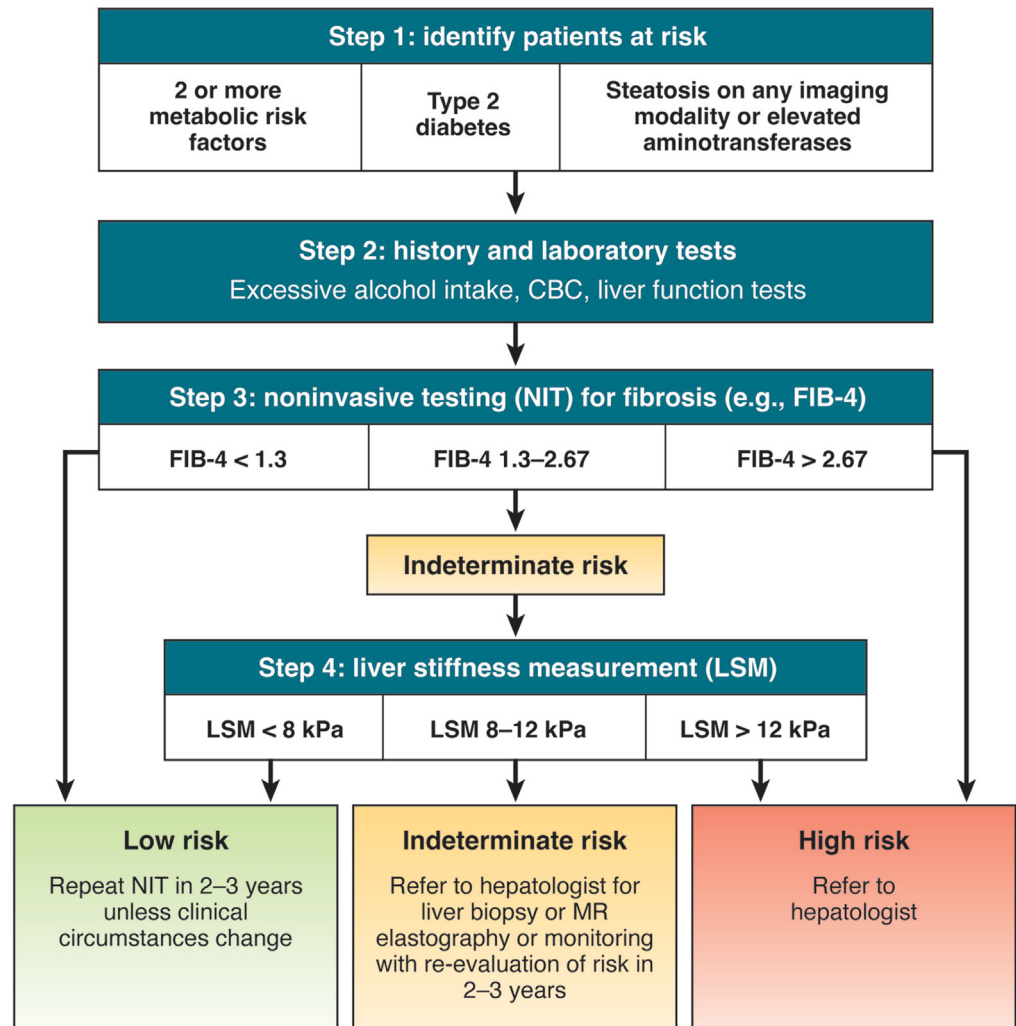
Figure 3. The American Gastroenterological Association clinical care pathway for the risk stratification of NAFLD. Modified with permission from Kanwal et al.[83] CBC, complete blood count; MR, magnetic resonance.

Clinical Algorithms for Patient Evaluation

The American Gastroenterological Association clinical care pathway focuses on patients at risk of NAFLD, though similar concepts can be applied to patients with other chronic liver diseases (Figure 3).[83] In essence, the pathway starts from identifying patients with confirmed or suspected liver disease based on risk factors and abnormal liver tests, with an emphasis on the need to involve primary care physicians and nonhepatologists in case identification. Simple NITs such as FIB-4 are the preferred initial assessment because of their availability, low cost, and high negative predictive value in excluding advanced fibrosis and future liver-related events. Low-risk patients (eg, FIB-4 of <1.3) can be safely managed by primary care, with NIT repeated in 2–3 years unless clinical circumstances change. Patients with indeterminate results (eg, FIB-4 of 1.3–2.67) can undergo a more specific NIT such as specific blood biomarker of fibrosis, VCTE, or MRE depending on the local setting and test availability, whereas high-risk patients (eg, FIB-4 of >2.67 or the second test suggests advanced fibrosis) should be referred to hepatologists. It is recognized that

a second test may not be available in some settings. In that case, it is reasonable to refer patients with indeterminate simple fibrosis score results to hepatologists for further evaluation.

The American Gastroenterological Association clinical care pathway has recently been validated in 2322 participants who underwent VCTE during the 2017–2018 National Health and Nutrition Examination Survey.[84] Overall, 82% of the participants had low FIB-4, among whom 90% had liver stiffness <8 kPa, supporting the notion of using FIB-4 as the first step. In contrast, among participants with indeterminate and high FIB-4, 13% and 33% had liver stiffness ≥8 kPa. Thus, although the clinical care pathway is feasible, one should recognize the implications of false positive results and the corresponding healthcare burden.

The 2021 update of the European Association for the Study of the Liver Clinical Practice Guidelines on NITs for evaluation of liver disease severity and prognosis offers a similar but slightly different approach.[50] In low-prevalence populations, the role of NITs is to exclude, rather than to diagnose advanced fibrosis. Again, FIB-4 of <1.3 and liver stiffness <8 kPa by VCTE are sufficient to classify patients as low risk. In patients with liver
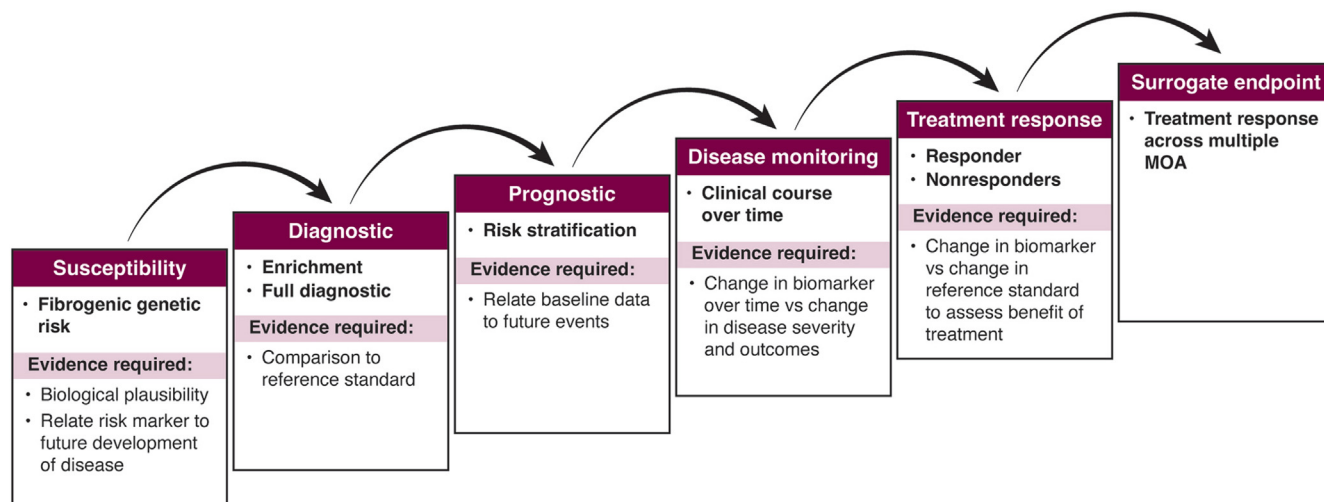
Figure 4. Roadmap for future development and integration of NITs in clinical practice and drug development in NAFLD. MOA, mechanism of action.

stiffness ≥ 8 kPa, the European guidelines recommend the use of a patented serum test. If the serum test is also abnormal (ie, concordant results with VCTE), advanced fibrosis is highly likely, and the patient can be managed as such. If patented serum tests are unavailable or there is discordance between serum tests and VCTE, liver biopsy should be considered.

Future Directions

In conclusion, we have tried in this review to provide a roadmap for future development and integration of NITs in clinical practice and in drug development in NAFLD (Figure 4). The field of NIT development for fibrosis remains a very active one with considerable room for innovation to guide future drug development and clinical care paradigms. For drug development, the needs of early-phase trials are distinct from those of advanced-phase trials. In early-phase trials, which are typically short in duration, it is more likely that fibrogenic drive will decrease, rather than the burden of fibrosis, which may take longer. There is a clear cut need to develop NITs that capture the inflammatory-fibrogenic drive with greater precision, and specific studies of such markers such as PRO-C3 and cT1 or 2- or 3-dimensional MRE for this purpose are needed.

In more advanced-phase trials, it will be important to define if the short-term changes noted with NITs in early-phase trials are valid predictors of long-term decrease in burden of fibrosis measured by decreased progression to cirrhosis, a generally accepted surrogate for drug development. Further, NITs will need to be related to an actual decrease in the fibrosis distribution and burden and eventually clinical outcomes. Demonstration of biological plausibility along with such performance characteristics across multiple mechanisms of action will be needed to establish a surrogate endpoint for clinical trials. These are critically needed to accelerate

drug development and the development of combination therapies. Also, once multiple therapies are approved, it will be needed to define complete responders vs partial responders and nonresponders to guide further therapeutics.

The integration of NITs into routine clinical practice will require translation of findings into routine clinical settings and their validation in such settings. Other key future directions in routine care are the establishment of optimized care pathways to guide management strategies. It is now recognized that NASH is not only an isolated liver disease, but also a liver disease that occurs in the context of multiple end organ diseases that share common biological pathophysiology. This is evidenced by substantial collinearity between the incidence of liver and several nonhepatic outcomes at each stage of NASH.[12] Future directions in clinical settings will need to include either common measures of multiple end organ disease or separate measures of key end organ dysfunction that represent competing threats to life and contribute to the risk of death. This is likely to dictate the choice of initial and subsequent add-on therapies.

In summary, the development of NITs for NASH remains a very active area of investigation. There are several major initiatives such as NIMBLE (Non-Invasive Biomarkers of Metabolic Liver Disease) and LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) that are expected to generate key data to support specific contexts of use for several NITs. These, along with NIT development in the context of large clinical trials and investigator-initiated studies, are likely to transform the drug development approaches and routine clinical care of patients with NAFLD.

References

1. Younossi ZM. The epidemiology of nonalcoholic steatohepatitis. Clin Liver Dis (Hoboken) 2018;11:92–94.

2. Teng ML, Ng CH, Huang DQ, et al. Global incidence and prevalence of non-alcoholic fatty liver disease. Clin Mol Hepatol 2023;29:S32–S42.

3. Neuschwander-Tetri BA, Clark JM, Bass NM, et al. Clinical, laboratory and histological associations in adults with nonalcoholic fatty liver disease. Hepatology 2010;52:913–924.

4. Kleiner DE, Brunt EM, Wilson LA, et al. Association of histologic disease activity with progression of nonalcoholic fatty liver disease. JAMA Netw Open 2019;2:e1912565.

5. Estes C, Razavi H, Loomba R, et al. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology 2018;67:123–133.

6. Estes C, Anstee QM, Arias-Loste MT, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. J Hepatol 2018;69:896–904.

7. Siddiqui MS, Yamada G, Vuppalanchi R, et al. Diagnostic accuracy of noninvasive fibrosis models to detect change in fibrosis stage. Clin Gastroenterol Hepatol 2019;17:1877–1885.e5.

8. Center for Drug Evaluation and Research. Nonalcoholic Steatohepatitis With Compensated Cirrhosis: Developing Drugs for Treatment Guidance for Industry. Rockville, MD: Food and Drug Administration; 2019.

9. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. Hepatology 2009;49:1017–1044.

10. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol 2020;73:1322–1332.

11. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313–1321.

12. Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med 2021;385:1559–1569.

13. Taylor RS, Taylor RJ, Bayliss S, et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. Gastroenterology 2020;158:1611–1625.e12.

14. Anstee QM, Castera L, Loomba R. Impact of non-invasive biomarkers on hepatology practice: past, present and future. J Hepatol 2022;76:1362–1378.

15. Center for Drug Evaluation and Research. Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff (Draft Guidance). Rockville, MD: Food and Drug Administration; 2018.

16. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring, MD: Food and Drug Administration; Bethesda, MD: National Institutes of Health; 2016.

17. Luo Y, Wadhawan S, Greenfield A, et al. SOMAscan proteomics identifies serum biomarkers associated with liver fibrosis in patients with NASH. Hepatol Commun 2021;5:760–773.

18. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al. Vibration-controlled transient elastography to assess fibrosis and steatosis in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2019;17:156–163.e2.

19. Sanyal AJ, Harrison SA, Ratziu V, et al. The natural history of advanced fibrosis due to nonalcoholic steatohepatitis: data from the simtuzumab trials. Hepatology 2019;70:1913–1927.

20. Center for Drug Evaluation and Research. Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment: Guidance For Industry. Rockville, MD: Food and Drug Administration; 2018.

21. Center for Devices and Radiological Health. Statistical Guidance on Reporting Results From Studies Evaluating Diagnostic Tests. Rockville, MD: Food and Drug Administration; 2007.

22. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926–930.

23. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform 2014;48:193–204.

24. Ray P, Le Manach Y, Riou B, et al. Statistical evaluation of a biomarker. Anesthesiology 2010;112:1023–1040.

25. Emerson SC, Waikar SS, Fuentes C, et al. Biomarker validation with an imperfect reference: Issues and bounds. Stat Methods Med Res 2018;27:2933–2945.

26. Cales P, Boursier J, Oberti F, et al. FibroMeters: a family of blood tests for liver fibrosis. Gastroenterol Clin Biol 2008;32:40–51.

27. Jung J, Loomba RR, Imajo K, et al. MRE combined with FIB-4 (MEFIB) index in detection of candidates for pharmacological treatment of NASH-related fibrosis. Gut 2021;70:1946–1953.

28. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1–W12.

29. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.

30. Wong VW, Adams LA, de Ledinghen V, et al. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol 2018;15:461–478.

31. Karlas T, Petroff D, Sasso M, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. J Hepatol 2017;66:1022–1030.

32. Petroff D, Blank V, Newsome PN, et al. Assessment of hepatic steatosis by controlled attenuation parameter using the M and XL probes: an individual patient data meta-analysis. Lancet Gastroenterol Hepatol 2021;6:185–198.

33. Wong VWS, Zelber-Sagi S, Cusi K, et al. Management of NAFLD in primary care settings. Liver Int 2022;42:2377–2389.

34. Mahady SE, Macaskill P, Craig JC, et al. Diagnostic accuracy of noninvasive fibrosis scores in a population of individuals with a low prevalence of fibrosis. Clin Gastroenterol Hepatol 2017;15:1453–1460.e1.

35. Wong VW, Wong GL, Woo J, et al. Impact of the new definition of metabolic associated fatty liver disease on the epidemiology of the disease. Clin Gastroenterol Hepatol 2021;19:2161–2171.e5.

36. Hagström H, Tälback M, Andreasson A, et al. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. J Hepatol 2020;73:1023–1029.

37. Lee HW, Wong GL, Kwok R, et al. Serial transient elastography examinations to monitor patients with type 2 diabetes: a prospective cohort study. Hepatology 2020;72:1230–1241.

38. Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. J Hepatol 2019;71:371–378.

39. Labenz C, Arslanow A, Nguyen-Tat M, et al. Structured early detection of asymptomatic liver cirrhosis: results of the population-based liver screening program SEAL. J Hepatol 2022;77:695–701.

40. Boursier J, Guillaume M, Leroy V, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. J Hepatol 2019; 71:389–396.

41. Park CC, Nguyen P, Hernandez C, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. Gastroenterology 2017;152:598–607.e2.

42. de Franchis R, Bosch J, Garcia-Tsao G, et al. Baveno VII - renewing consensus in portal hypertension. J Hepatol 2022; 76:959–974.

43. Newsome PN, Sasso M, Deeks JJ, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020;5:362–373.

44. Noureddin M, Truong E, Gornbein JA, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. J Hepatol 2022;76:781–787.

45. Scott E, Anstee QM. Genetics of alcoholic liver disease and non-alcoholic steatohepatitis. Clin Med (Lond) 2018;18:s54–s59.

46. Abul-Husn NS, Cheng X, Li AH, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med 2018;378:1096–1106.

47. Yip TC, Vilar-Gomez E, Petta S, et al. Geographical similarity and differences in the burden and genetic predisposition of NAFLD. Hepatology 2022;77:1404–1427.

48. Bianco C, Jamialahmadi O, Pelusi S, et al. Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using polygenic risk scores. J Hepatol 2021;74:775–782.

49. Whitfield JB, Schwantes-An TH, Darlay R, et al. A genetic risk score and diabetes predict development of alcohol-related cirrhosis in drinkers. J Hepatol 2022;76:275–282.

50. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol 2021;75:659–689.

51. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1264–1281 e4.

52. Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2013;145:782–789.e4.

53. Sebastiani G, Alshaalan R, Wong P, et al. Prognostic value of non-invasive fibrosis and steatosis tools, hepatic venous pressure gradient (HVPG) and histology in nonalcoholic steatohepatitis. PLoS One 2015;10:e0128774.

54. Boursier J, Vergniol J, Guillet A, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. J Hepatol 2016;65:570–578.

55. Hagstrom H, Nasr P, Ekstedt M, et al. Accuracy of noninvasive scoring systems in assessing risk of death and liver-related endpoints in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2019;17:1148–1156.e4.

56. Onnerhag K, Hartman H, Nilsson PM, et al. Non-invasive fibrosis scoring systems can predict future metabolic complications and overall mortality in non-alcoholic fatty liver disease (NAFLD). Scand J Gastroenterol 2019;54:328–334.

57. Ioannou GN, Green P, Kerr KF, et al. Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. J Hepatol 2019; 71:523–533.

58. Younes R, Caviglia GP, Govaere O, et al. Long-term outcomes and predictive ability of non-invasive scoring systems in patients with non-alcoholic fatty liver disease. J Hepatol 2021; 75:786–794.

59. Lee J, Vali Y, Boursier J, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: a systematic review. Liver Int 2021;41:261–270.

60. Cianci N, Subhani M, Hill T, et al. Prognostic non-invasive bio-markers for all-cause mortality in non-alcoholic fatty liver disease: a systematic review and meta-analysis. World J Hepatol 2022;14:1025–1037.

61. Kim D, Kim WR, Kim HJ, et al. Association between noninvasive fibrosis markers and mortality among adults with nonalcoholic fatty liver disease in the United States. Hepatology 2013; 57:1357–1365.

62. Unalp-Arida A, Ruhl CE. Liver fibrosis scores predict liver disease mortality in the United States population. Hepatology 2017;66:84–95.

63. Younossi ZM, Anstee QM, Wai-Sun Wong V, et al. The association of histologic and noninvasive tests with adverse clinical and patient-reported outcomes in patients with advanced fibrosis due to nonalcoholic steatohepatitis. Gastroenterology 2021;160:1608–1619.e13.

64. Munteanu M, Pais R, Peta V, et al. Long-term prognostic value of the FibroTest in patients with non-alcoholic fatty liver disease, compared to chronic hepatitis C, B, and alcoholic liver disease. Aliment Pharmacol Ther 2018;48:1117–1127.

65. Shili-Masmoudi S, Wong GL, Hiriart JB, et al. Liver stiffness measurement predicts long-term survival and complications in non-alcoholic fatty liver disease. Liver Int 2020; 40:581–589.

66. Petta S, Sebastiani G, Vigano M, et al. Monitoring occurrence of liver-related events and survival by transient elastography in patients with nonalcoholic fatty liver disease and compensated advanced chronic liver disease. Clin Gastroenterol Hepatol 2021;19:806–815.e5.

67. Braude M, Roberts S, Majeed A, et al. Liver stiffness (Fibroscan(R)) is a predictor of all-cause mortality in people with non-alcoholic fatty liver disease. Liver Int 2023;43:90–99.

68. Gidener T, Ahmed OT, Larson JJ, et al. Liver stiffness by magnetic resonance elastography predicts future cirrhosis, decompensation, and death in NAFLD. Clin Gastroenterol Hepatol 2021;19:1915–1924.e6.

69. Han MAT, Vipani A, Noureddin N, et al. MR elastography-based liver fibrosis correlates with liver events in nonalcoholic fatty liver patients: a multicenter study. Liver Int 2020;40:2242–2251.

70. Higuchi M, Tamaki N, Kurosaki M, et al. Longitudinal association of magnetic resonance elastography-associated liver stiffness with complications and mortality. Aliment Pharmacol Ther 2022; 55:292–301.

71. Ajmera V, Kim BK, Yang K, et al. Liver Stiffness on magnetic resonance elastography and the MEFIB index and liver-related outcomes in nonalcoholic fatty liver disease: a systematic review and meta-analysis of individual participants. Gastroenterology 2022;163:1079–1089.e5.

72. Pennisi G, Enea M, Pandolfo A, et al. AGILE 3+ score for the diagnosis of advanced fibrosis and for predicting liver-related events in NAFLD. Clin Gastroenterol Hepatol 2023; 21:1293–1302.e5.

73. Boursier J, Hagstrom H, Ekstedt M, et al. Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events. J Hepatol 2022;76:1013–1020.

74. Balkhed W, Aberg FO, Nasr P, et al. Repeated measurements of non-invasive fibrosis tests to monitor the progression of non-alcoholic fatty liver disease: a long-term follow-up study. Liver Int 2022;42:1545–1556.

75. Chalasani N, Abdelmalek MF, Loomba R, et al. Relationship between three commonly used non-invasive fibrosis biomarkers and improvement in fibrosis stage in patients with non-alcoholic steatohepatitis. Liver Int 2019;39:924–932.

76. Sanyal AJ, Anstee QM, Trauner M, et al. Cirrhosis regression is associated with improved clinical outcomes in patients with nonalcoholic steatohepatitis. Hepatology 2022; 75:1235–1246.

77. Tamaki N, Munaganuru N, Jung J, et al. Clinical utility of 30% relative decline in MRI-PDFF in predicting fibrosis regression in non-alcoholic fatty liver disease. Gut 2022;71:983–990.

78. Loomba R, Sanyal AJ, Kowdley KV, et al. Factors associated with histologic response in adult patients with nonalcoholic steatohepatitis. Gastroenterology 2019;156:88–95.e5.

79. Rinella ME, Dufour JF, Anstee QM, et al. Non-invasive evaluation of response to obeticholic acid in patients with NASH: results from the REGENERATE study. J Hepatol 2022; 76:536–548.

80. Nielsen MJ, Veidal SS, Karsdal MA, et al. Plasma Pro-C3 (N-terminal type III collagen propeptide) predicts fibrosis progression in patients with chronic hepatitis C. Liver Int 2015; 35:429–437.

81. Loomba R, Lawitz E, Mantry PS, et al. The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: a randomized, phase 2 trial. Hepatology 2018;67:549–559.

82. Harrison SA, Rossi SJ, Paredes AH, et al. NGM282 improves liver fibrosis and histology in 12 weeks in patients with nonalcoholic steatohepatitis. Hepatology 2020;71:1198–1212.

83. Kanwal F, Shubrook JH, Adams LA, et al. Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. Gastroenterology 2021; 161:1657–1669.

84. Udompap P, Therneau TM, Canning RE, et al. Performance of American Gastroenterological Association Clinical Care Pathway for the risk stratification of patients with nonalcoholic fatty liver disease in the US population. Hepatology 2023;77:931–941.

Correspondence

Address correspondence to: Arun Sanyal, MD, Virginia Commonwealth University School of Medicine, Department of Internal Medicine, 1220 East Broad Street, Richmond, VA 23298. e-mail: arun.sanyal@vcuhealth.org; or Laurent Castera, MD, PhD, Service d'hépatologie, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, 100 boulevard du Général Leclerc, 92110 Clichy, Paris, France. e-mail: laurent.castera@bjn.aphp.fr; or Vincent Wai-Sun Wong, MD, State Key Laboratory of Digestive Disease, Institute of Digestive Disease, Medical Data Analytics Centre, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, 9/F, Prince of Wales Hospital, Shatin, Hong Kong, China. e-mail: wongv@mect.cuhk.edu.hk