Liver International

WILEY

Liver INTERNATIONAL

MINI REVIEW

Liver Biopsy in Metabolic Dysfunction-Associated Steatohepatitis Clinical Trials: Potential Pitfalls and Solutions

Eric J. Lawitz[1] | Fred Poordad[1] | Zachary D. Goodman[2]

[1]The Texas Liver Institute, University of Texas Health San Antonio, San Antonio, Texas, USA | [2]Center for Liver Disease, Fairfax, Virginia, USA

Correspondence: Eric J. Lawitz (lawitz@txliver.com)

Received: 14 January 2026 | **Revised:** 7 July 2026 | **Accepted:** 11 July 2026

Handling Editor: Luca Valenti

Keywords: clinical trials | histology | liver biopsy | MASH | serious adverse events

ABSTRACT

Liver biopsy is the reference standard for the evaluation of inflammation and fibrosis in MASH and is the primary endpoint in steatohepatitis clinical trials. An optimal liver biopsy is generally considered to be obtained with a 16-gauge needle, measure at least 2 cm in length and contain at least 10 complete portal tracts, although these features have not been prospectively validated as mandatory thresholds in MASH trials. Published clinical trials for MASH have not consistently reported sufficient details of liver biopsy quality (needle gauge, length, number of portal tracts) at baseline or post-treatment, and biopsy quality metrics are rarely reported separately by treatment arm. Placebo response rates in clinical trials for MASH varied widely (2% to 34%) and this variation could influence the outcomes of clinical trials. Future MASH trials should report biopsy quality metrics (needle gauge, core length and portal tract count) by treatment arm in supplementary appendices of those publications, treat the 16-gauge/≥ 2 cm/≥ 10 portal tract criteria as preferred acquisition standards rather than required thresholds, and may consider pre-specified sensitivity analyses restricted to biopsies meeting adequacy criteria, recognising that such exclusions may alter the disease spectrum and must be interpreted cautiously. Biopsy quality is a plausible contributor to variability in placebo responses and trial outcomes, but direct evidence that needle gauge, core length, or portal tract number independently affect histologic response rates in randomised MASH trials remains limited. As digital pathology/AI tools mature and NIT-based surrogate endpoints advance toward regulatory qualification, ensuring high quality histologic reference standards remains essential.

1 | Introduction

A meta-analysis of studies published from 1990 to 2019 suggests that the global prevalence of Metabolic Dysfunction-Associated Steatohepatitis (MASH; previously NASH or Non-alcoholic steatohepatitis) in the global population among persons ≥ 20 years of age was approximately 5.3% (standard error = 2.6%) [1]. The prevalence of nonalcoholic fatty liver disease has increased in population studies between 2016 and 2019 compared to studies between 1990 and 2006 [1]. Using data from the Global

Abbreviations: AASLD, American Association for the Study of Liver Diseases; cm, centimetre; cT1, MRI-corrected T1; DP/AI, digital pathology/artificial intelligence; EASL, European Association for the Study of the Liver; ELF, enhanced liver fibrosis; FAST, FibroScan aspartate aminotransferase; FDA, Food and Drug Administration; FIB-4, Fibrosis-4 Index; LLB, laparoscopic liver biopsy; LSM, liver stiffness measurement; MALO, major adverse liver outcomes; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease; mITT, modified intent-to-treat; mm, millimetre; MRE, magnetic resonance elastography; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NFS, non-alcoholic fatty liver disease fibrosis score; NIMBLE, Non-invasive Biomarkers for Metabolic Liver Disease; NITs, non-invasive tests; SAEs, serious adverse events; US, United States; VCTE, vibration-controlled transient elastography.

Lay Summary

The liver biopsy (microscopic evaluation of a piece of liver taken with a needle) is the best way to determine if a patient has liver disease. Clinical trials should report the details of the liver biopsy (width of the needle used to get the biopsy, and the length of the liver biopsy) in order to ensure that the groups of patients have similar details. In addition, non-invasive methods of determining liver disease (blood tests and measurement of liver stiffness) should be compared against liver biopsy results so we can confirm they give similarly reliable information without requiring a needle biopsy.

Key Points

- Liver biopsy remains the reference standard for the evaluation of inflammation and fibrosis in MASH and is the basis for accelerated regulatory approval of MASH therapies.
- An optimal liver biopsy is generally considered to be obtained with a 16-gauge needle, measure at least 2.0 cm in length and to contain at least 10 complete portal tracts, although these thresholds have not been prospectively validated in MASH trials.
- Published clinical trials for MASH have not consistently reported biopsy quality (needle gauge, core length, portal tract count) at baseline or post-treatment, and these metrics are rarely reported separately by treatment arm.
- Future MASH trials should report biopsy quality metrics (needle gauge, core length and portal tract count) by treatment arm and treat the 16-gauge/≥ 2.0 cm/≥ 10 portal tract criteria as the preferred acquisition target when clinically feasible, and may consider prespecified sensitivity analyses restricted to biopsies meeting adequacy criteria, recognising that such exclusions can alter the disease spectrum and must be interpreted cautiously.
- Digital pathology/AI tools and non-invasive tests are advancing, but their performance still depends on high-quality histologic reference standards.

Burden of Disease Study 2021, the age-standardised incidence rate of Metabolic Dysfunction-associated Steatotic Liver Disease (MASLD) per 100 000 people rose from 475.54 in 1990 to 593.28 in 2021 [2]. The age-standardised prevalence rate of MASLD per 100 000 people is projected to rise to 928.10 by 2045 [2]. In the United States, the number of cases of MASH are predicted to increase from 17 million in 2016 to 27 million in 2030, an increase of 59% [3]. The percentage of MASH cases with F3/F4 fibrosis is also predicted to increase from 21% in 2016 to 29% in 2030 [3]. These projections for liver disease burden do not account for the increasing use of glucagon-like peptide-1 (GLP-1) agonists or the thyroid hormone receptor beta agonist.

Histology via liver biopsy remains the reference standard for the grading and staging of MASH according to the American Association for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL) [4, 5]. The principal goal of the liver biopsy is to establish or confirm a histologic diagnosis or to generate a differential diagnosis of the aetiology of liver disease [6]. Specifically in MASH, the liver biopsy allows for the evaluation of the defining histologic features including steatosis, inflammation, cellular ballooning and fibrosis [4]. Elastography and serum tests do not provide the entirety of data obtained from tissue analysis. The liver biopsy has several challenges with respect to endpoints in clinical trials (its invasive nature; inter- and intra-reader variability; partial representation of the entire liver; the absence of harmonised guidance for reading liver biopsies by regulatory authorities).

The liver biopsy is the key criterion for conditional approval by regulatory authorities for treatment of non-cirrhotic MASH in clinical trials. Therapies can be granted accelerated approval under 21 Code of Federal Regulations Part 314, Subpart H by the United States Food and Drug Administration (FDA), which permits approval of medications based on surrogate markers to provide patients earlier access to promising therapies. The surrogate endpoints for treatment of MASH are improvement in liver fibrosis with no worsening of steatohepatitis, or the resolution of steatohepatitis with no worsening of liver fibrosis. This accelerated approval pathway requires additional clinical data after initial approval to confirm that the therapy meets clinically meaningful endpoints (decreases in death from any cause; liver transplants; hepatic decompensation; progression to cirrhosis). Currently, two therapies (resmetirom and semaglutide) are conditionally approved by the US FDA for the treatment of MASH under this accelerated approval pathway. Several additional therapies are in ongoing phase 3 trials.

2 | Importance of the Sample Quality of the Liver Biopsy

Results support the routine use of 16-gauge rather than 18-gauge biopsy needles for routine ultrasound assisted percutaneous liver biopsies [7]. Laparoscopic liver biopsies (LLB) with a 16- or 18-gauge needle are indicated for complex or high-risk patients (morbid obesity targeted multi-site sampling, or technically unfeasible) [8]. A recent international multidisciplinary Delphi consensus on LLB (45 experts across six continents) standardised its application in these scenarios and emphasised that direct visual inspection of the liver surface allows targeted, multi-site sampling and the retrieval of high-quality specimens. For a laparoscopic liver biopsy, at least 2–3 complete tissue cores with each core ≥ 1.5 cm in length and containing more than 10 portal tracts would be considered adequate [8]. Based on available observational and methodological studies, an optimal biopsy for MASH assessment is generally considered to be obtained with a 16-gauge needle, measuring ≥ 2 cm in length and containing at least 10 complete portal tracts. These features have not yet been validated as mandatory quality thresholds in clinical trials for MASH. To highlight the impact of tissue quantity, when using a lower minimum cutoff of only ≥ 6 portal tracts, the percentages of cases that would have met the criterion of complete portal tracts obtained with two 1.5 cm biopsy cores were 95% (37/39) for a 16-gauge needle compared to only 70% (32/46)

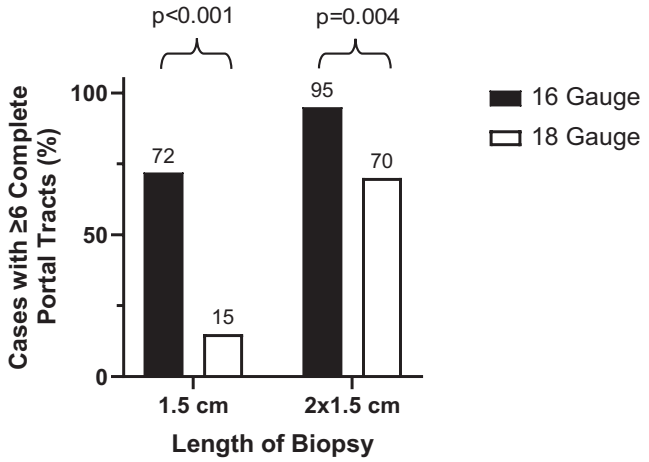

FIGURE 1 | Comparison of the percent of liver biopsies that would have contained ≥6 complete portal tracts in 1.5 cm or two 1.5 cm liver biopsies obtained with 16-gauge or 18-gauge needles. Based on the number of portal tracts in 1.0 cm biopsies, the number of biopsies that would have met the criterion of ≥6 complete portal tracts was calculated for a single 1.5 cm biopsy core compared with 2 passes of 1.5 cm each that would have achieved a 3.0 cm biopsy core [7].

for an 18-gauge needle ($p = 0.004$) (Figure 1) [7]. The percentage of cases that would have met the criterion of ≥6 complete portal tracts for only one 1.5 cm biopsy core was 72% (28/39) for a 16-gauge needle compared to 15% (7/46) for an 18-gauge needle ($p = 0.001$). Biopsies taken with an 18-gauge needle were significantly narrower ($p < 0.001$), more fragmented ($p < 0.001$) and smaller in cross-sectional area ($p < 0.001$) than biopsies taken with a 16-gauge needle resulting in fewer complete portal tracts. These results suggest that liver biopsies should be obtained with 16-gauge needles and not be ≤1.5 cm in length. It should be noted that these data on needle gauges were derived from a liver biopsy cohort of mixed etiologies including MASH. While the principles of tissue adequacy are considered generalisable across liver diseases, direct validation in homogenous populations of MASH is lacking.

Shorter biopsies also resulted in a significant increase in the percentage of patients reported to have mild grades of inflammation: 86.6% (139/161) in 1 cm long biopsies versus 60.2% (97/161) in 1.5 cm long biopsies and 49.7% (80/161) in ≥3 cm long biopsies ($p < 0.001$) [9]. Similarly, shorter biopsies resulted in a significant increase of patients staged as having mild fibrosis: 80.1% (129/161) in 1 cm long biopsies versus 68.3% (110/161) in 1.5 cm long biopsies and 59% (95/161) in ≥3 cm long biopsies ($p < 0.002$) [9]. A biopsy length equal to or longer than 2.0 cm and containing at >10 portal tracts was considered necessary for accurate evaluation. These data were derived from patients with chronic viral hepatitis and not MASH.

3 | Liver Biopsy Complications and Safety

The safety profile in patients who underwent a percutaneous liver biopsy with a 16-gauge needle was similar in those with an 18-gauge core liver biopsy. One hundred and fifty patients were randomised to undergo a percutaneous liver biopsy with a 16-gauge or an 18-gauge core liver biopsy [10]. Patients who underwent a liver biopsy with a 16-gauge needle had a similar safety profile compared to those with an 18-gauge needle with respect to pain scores 1 h after the biopsy ($p = 1.0$), 3 h after the biopsy ($p = 0.82$) and 24 h after the biopsy ($p = 1.0$) [10]. Most (64%–70%) of the pain scores after the liver biopsy were mild with 16-gauge or 18-gauge needles [10].

The use of ultrasound assistance has benefits in percutaneous liver biopsies. In a prospective study of 166 patients with hepatitis C virus who underwent either ultrasound assisted biopsy or non-assisted biopsy, ultrasound assisted biopsies caused significantly less biopsy pain (36.4% vs. 47.3%; $p < 0.0001$) and significantly less pain-related morbidity (1.8% vs. 7.7%, $p < 0.05$) [11]. There was no significant difference in diagnostic yield between ultrasound assisted and unassisted biopsies (98% vs. 94%, $p = 0.15$).

Overall, percutaneous liver biopsy is considered a safe procedure with few serious complications. In a meta-analysis of 51 studies with results from 12 481 patients published before October 2018, major complications were rare [12]. Of the 12 481 patients, 63 (0.5%) had moderate or severe post-procedure pain that resolved with analgesic medication, and 45 (0.4%) developed bleeding requiring transfusion. Three patients (0.02%) had vasovagal reactions requiring medication and two patients (0.02%) developed pneumothorax. One patient with a history of hepatocellular carcinoma with liver cirrhosis died 2 h after the biopsy [12].

Even in patients with compensated cirrhosis and chronically infected with hepatitis C virus, the rate of liver biopsy complications was also rare [13]. The rate of serious adverse events (SAEs) after a liver biopsy was 1.1% (29/2740). The most common SAEs were bleeding (0.6%; 16/2740), severe pain (0.26%; 7/2740) and punctured gall bladder (0.07%; 2/2740). An SAE of marked hypotension, pneumothorax, syncope and dehydration occurred in one patient each [13].

4 | Variability of Liver Biopsies

In a large dataset of paired liver biopsies from a clinical trial, inter-reader agreement was poor and intra-reader variability was modest [14]. The kappa values for inter-reader agreement on histologic endpoints from 339 patients with paired biopsies were generally 0.4 or less, while intra-reader agreement ranged from 0.645 to 0.836. In a separate study of digitised slides evaluated independently by nine internationally recognised expert liver pathologists, substantial divergence was observed in the identification of ballooned hepatocytes [15]. The kappa value for overall interobserver agreement on the presence or absence of ballooning was 0.197, improving only to 0.362 when a five-cell or greater threshold was applied. Notably, applying a size-based threshold (greater than two or three times the size of normal hepatocytes) was insufficient to meaningfully improve interobserver agreement. However, the intraclass correlation coefficient for consistency in ballooning burden was higher at 0.718, indicating moderate agreement when ballooning was assessed as a continuum rather than as a binary determination. These findings are particularly relevant given that hepatocyte ballooning is a key histologic feature underlying the surrogate endpoints used for accelerated approval where its unequivocal resolution is required

to demonstrate therapeutic efficacy. While the data highlight a degree of subjectivity that warrants careful consideration, Brunt and colleagues concluded that ballooning exists along a spectrum and proposed the development of a concordance atlas to train artificial intelligence-based imaging technologies that could reproducibly quantify ballooned hepatocytes, potentially standardising the assessment of therapeutic efficacy in future clinical trials [15]. Such AI-assisted approaches may help reconcile the observed interpretive variability with the rigour required by current FDA regulatory endpoints.

These methodological concerns have had real consequences in drug development. A phase 2 clinical trial for MASH assessing seladelpar was stopped due to safety concerns based on liver histology at the end of treatment [16, 17]. Screening and baseline biopsies were assessed by a single pathologist while the end of treatment biopsies were assessed by 2 pathologists blinded to the treatment groups. Subsequent adjudication by a panel of expert pathologists concluded that the histologic concerns at the end of treatment were also present at baseline. Relying on a single pathologist may not provide the most reliable results due to intra-observer variability. If multiple pathologists had reviewed the samples, it might have been recognised that the concerning findings of new histologic worsening were already present at baseline. This could have prevented halting a program that ultimately resulted in an approved therapy of seladelpar for primary biliary cholangitis. These events underscore the importance of employing consensus-based, multi-reader approaches for histologic assessment at all study timepoints.

5 | Non-Invasive Methods Compared to Liver Biopsy

Non-invasive tests (NITs) clearly have a role in the assessment of liver fibrosis. NITs have been widely adopted to determine the stage of liver fibrosis or stiffness in chronic liver disease, in diagnosing and risk-stratifying cirrhosis and predicting the complications of cirrhosis [18]. NITs include blood and imaging-based assessments [18, 19]. Before considering a liver biopsy to assess fibrosis staging in patients with chronic liver disease, the AASLD recommends initially using blood and imaging-based NITs to detect significant fibrosis (F2-4), advanced fibrosis (F3-4), or cirrhosis (F4) [20]. In order to stage fibrosis, the AASLD recommends starting with a blood-based NIT such as the Fibrosis-4 Index (FIB-4) or the Nonalcoholic Fatty Liver Disease Fibrosis Score (NFS) followed by confirmatory testing with an imaging-based NIT such as liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE) or Magnetic Resonance Elastography (MRE) [21]. The most validated and commonly used blood-based NITs for assessing liver fibrosis are the Fibrosis-4 Index and the Enhanced Liver Fibrosis score, whereas the best-validated elastography-based NITs are VCTE and MRE [18].

EASL also recommends a multi-step approach to detect fibrosis in adults with MASLD. First, a blood-based NIT such as FIB-4 should be used. Then an imaging-based test should be used to determine the fibrosis stage [5]. In most cases, liver biopsies are not required for the clinical management of patients with chronic liver disease. However, liver biopsies are still required for the definitive diagnosis of steatohepatitis.

Composite NIT scores incorporating VCTE with clinical and laboratory data are emerging as prognostic tools. The Agile 3+ and Agile 4 scores have been shown to accurately predict major adverse liver outcomes (MALO) in MASH, exceeding those of LSM alone, FIB-4, or FAST score [22]. A recent head-to-head comparison of VCTE and histology in 3532 patients with MASLD found that both methods had similar accuracy in predicting liver-related events, suggesting that VCTE-based assessment may serve as a viable alternative to histologic staging for prognostic purposes [23]. These findings raise the question of whether histologic endpoints remain necessary for outcome prediction. Liver biopsy continues to be required for the definitive diagnosis of steatohepatitis and for current regulatory approval pathways. The requirement for paired liver biopsies also imposes substantial operational burden on MASH drug development. Screen failure rates of 70%–80% have been reported in MASH trials, driven in large part by the absence of ballooning on screening biopsies [24]. The invasive nature of the procedure contributes to patient reluctance, slow enrollment and dropout, collectively increasing trial costs and restricting the generalisability of trial results.

6 | Digital Pathology/Artificial Intelligence (DP/AI) Models

Digital pathology and artificial intelligence models can be used to automatically detect, localise, quantify and score histologic parameters [25]. In addition, DP/AI models have the potential to reduce the impact of scoring variability in clinical trials for MASH [25, 26]. DP/AI models using whole-slide imaging systems could provide a more reliable assessment of treatments in clinical trials of MASH and be included in clinical practice. The utility of machine learning for quantitative liver histology assessment and disease monitoring in MASH was demonstrated in a study in which 5139 liver biopsy slides from 3018 patients enrolled in three randomised controlled trials for advanced fibrosis were evaluated by both three experienced hepatobiliary pathologists and a machine learning model. There was a high degree of concordance between the histological parameters estimated by the AI model and an experienced pathologist for steatosis and fibrosis [27]. However, the concordance was lower for hepatocellular ballooning and lobular inflammation. In addition, a stringent machine-learning fibrosis score minimised the placebo response rate in the large number of patients treated with placebo in two clinical trials. Once the AI model is trained, the system is expected to be highly reproducible [27]. As with pathologist assessments, however, adequate liver biopsy material remains essential to optimise the performance of DP/AI models. If inadequate biopsies are used as the reference standard during validation and calibration of DP/AI tools, systematic errors will be embedded in the resulting models. Such errors are difficult to detect post hoc and may propagate into downstream clinical and regulatory decisions. This underscores why biopsy adequacy standards must precede, not follow, the deployment of these emerging tools.

7 | Quality of Liver Biopsies in Clinical Trials

The response rates in the placebo groups in clinical trials of treatments for MASH have varied widely ranging from 2% to 34% (Table 1). In general, details of the gauge of the needle

TABLE 1 | Quality of liver biopsies in clinical trials in MASH.

	Active groups					Placebo groups				
Intervention references	**Baseline biopsy: needle gauge; biopsy length (cm)**	**Post-treatment biopsy: needle gauge; biopsy length (cm)**	**Response rate (%): resolution of MASH[a]**	**Response rate (%): improvement of fibrosis[b]**	**Response rate (%): improvement of MASH[c]**	**Baseline biopsy: needle gauge; biopsy length (cm)**	**Post-treatment biopsy: needle gauge; biopsy length (cm)**	**Response rate (%): resolution of MASH[a]**	**Response rate (%): improvement of fibrosis[b]**	**Response rate (%): improvement of MASH[c]**
Pegozafermin [28]	NA; NA	NA; NA	37%; 23%; 26%	22%; 26%; 27%	NA	NA; NA	NA; NA	2%	7%	NA
Resmetirom [29]	16-gauge when possible; ~2.2 cm	16-gauge when possible; ~2.2 cm	25.9%; 29.9%	24.2%; 25.9%	NA	16-gauge when possible; ~2.2 cm	16-gauge when possible; ~2.2 cm	9.7%	14.2%	NA
Tirzepatide [30]	NA; 2.89 cm, 2.76 cm, 2.74 cm	NA; 2.89 cm, 2.76 cm, 2.74 cm	44%, 56%, 62%	NA	NA	NA; 3.00 cm	NA; 3.00 cm	10%	NA	NA
Denifanstat [31]	NA; NA	NA; NA	26%	NA	38%	NA; NA	NA; NA	11%	NA	16%
Survodutide [32]	16-gauge needle if possible; ≥ 2.0 cm	16-gauge needle if possible; ≥ 2.0 cm	NA	NA	47%; 62%; 43%	16-gauge needle if possible; ≥ 2.0 cm	16-gauge needle if possible; ≥ 2.0 cm	NA	NA	14%
Pioglitazone [33]	NA; NA	NA; NA	NA	NA	58%	NA; NA	NA; NA	NA	NA	17%
Efruxifermin [34]	NA; NA	NA; NA	NA	39%; 41%	NA	NA; NA	NA; NA	NA	20%	NA
Obeticholic Acid [35]	NA; 36% < 1.5 cm	NA; 41% < 1.5 cm	NA	NA	45%	NA; 17% < 1.5 cm	NA; 13% < 1.5 cm	NA	NA	21%
Lanifibranor [36]	NA; NA	NA; NA	NA	NA	48%; 55%	NA; NA	NA; NA	NA	NA	33%
Semaglutide [37]	NA; NA	NA; NA	62.9%	36.8%	NA	NA; NA	NA; NA	34.3%	22.4%	NA

Note: The results from the references on the interventions in clinical trials in MASH are summarised. The details of the liver biopsies at baseline and at post-treatment from the Supplementary Appendices of those publications are summarised if available.

Abbreviation: NA, not available.

[a]Resolution of MASH (absence of hepatocellular ballooning, the absence or mild inflammation and a reduction in the non-alcoholic fatty liver disease [NAFLD] activity score) with no worsening of fibrosis.

[b]Improvement of fibrosis by at least one stage with no worsening of the NAFLD activity score.

[c]Improvement of MASH (≥ 2 point reduction in NAFLD activity score) with no worsening of fibrosis.

used to obtain the liver biopsies and the lengths of the liver biopsies at baseline and at the end of treatment were not reported in the clinical trials of treatments for MASH. The percentages of biopsies obtained with a 16-gauge needle were not reported in any of these clinical trials. In the clinical trial of obeticholic acid, 36%–41% of the liver biopsies were < 1.5 cm in the groups that received obeticholic acid compared to 13%–17% of those in the placebo group [35]. In the clinical trial of tirzepatide, the lengths of the liver biopsies were 2.74–2.89 cm in the groups that received tirzepatide compared to 3.00 cm in the placebo group [30]. Published clinical trials to date do not provide sufficient information to assess the adequacy of liver biopsies either at baseline across treatment groups or when comparing pre-treatment and post-treatment samples. There is sporadic reporting of biopsy needle gauge and core length across trials. There are no studies in MASH that directly link gauge or length to histologic response rates after randomisation. The included trials differed in the definitions of endpoints, the requirements for biopsy, the durations of the trials and the study populations. For example, the endpoint in some trials was the resolution of MASH with no worsening of fibrosis while the endpoints in other trials were the improvement of fibrosis with no worsening of the NAFLD activity score, or the improvement of MASH with no worsening of fibrosis. The impact of biopsy quality on placebo response should therefore be regarded as hypothesis-generating rather than a proven determinant of heterogeneity.

8 | Emerging Biopsy-Sparing and Biopsy-Free Trial Frameworks

Several recent and ongoing MASH programs are incorporating NIT-based enrichment and endpoint strategies to reduce dependence on serial biopsies. Systematic reviews and meta-analyses suggest that MRE-based endpoints may exhibit placebo responses with lower between-study heterogeneity than histologic endpoints, supporting their potential reproducibility as trial endpoints. VCTE-based liver stiffness, while strongly supported for prognostic value and now under FDA qualification as a surrogate endpoint, has substantially higher measurement variability than MRE-based techniques.

Our focus in this review is on biopsy quality because liver histology remains central to accelerated approval in non-cirrhotic MASH, but we anticipate that validation of NIT-based surrogate endpoints will gradually shift development programs toward biopsy sparing or biopsy-free designs.

In the Foundation for the National Institutes of Health's Non-Invasive Biomarkers for Metabolic Liver Disease (NIMBLE) project, multiple biomarkers have met the prespecified criteria described in the Letter of Intent by the US FDA's Center for Drug Evaluation and Research for further qualification [38, 39]. The Enhanced Liver Fibrosis (ELF) test and the FibroMeterVCTE met the prespecified criteria for the diagnosis of clinically significant liver fibrosis (≥stage 2), advanced fibrosis (≥stage 3) and cirrhosis (stage 4). In addition, the NIS4 blood-based diagnostic panel met the prespecified criteria for further qualification efforts for diagnosis of MASH. These biomarkers may become 'reasonably likely' surrogate endpoints for clinical trials in adults with non-cirrhotic MASH and moderate to advanced fibrosis. This NIMBLE project which links changes in liver stiffness to the risk of all-cause mortality and liver-related events, explicitly acknowledges the potential for non-invasive tests to replace histologic assessment as the primary endpoint in accelerated-approval MASH trials. If fully qualified, VCTE-based liver stiffness could substantially reduce reliance on serial liver biopsies for both enrollment and efficacy evaluation, thereby lowering patient burden, improving recruitment and enabling more representative, adequately powered studies in MASH [40].

9 | Discussion

Histology via liver biopsy is still the reference standard for the evaluation of inflammation and fibrosis in MASH and constitutes the primary endpoints in clinical trials of treatments for MASH. The US FDA has developed the following guidance for developing medications for non-cirrhotic non-alcoholic steatohepatitis [41]:

1. Conditional approval of a medication for non-cirrhotic non-alcoholic steatohepatitis is based on achievement of either of two histologic end points (improvement in the stage of liver fibrosis or resolution of MASH) which are likely to result in clinical benefit.
2. Full approval of a medication is based on the reduction of clinical outcomes (death from any cause; liver transplantation; hepatic decompensation events; progression to cirrhosis).

Placebo response rates using histology in clinical trials for MASH have varied widely. The placebo response rates ranged from 2% for resolution of MASH to 33% for improvement of MASH (Table 1). This variation is largely driven by the specific histologic outcome assessed, as pooled analyses have demonstrated substantially different placebo rates across endpoints [42]. MASH is a dynamic disease capable of both spontaneous regression and progression in the absence of pharmacologic intervention [43]. Differences in trial duration, patient population characteristics and biopsy assessment methods likely contribute further to the heterogeneity observed across individual trials.

The impact of the definition of MASH resolution on placebo response rates was recently highlighted in a cross-trial analysis comparing the MAESTRO-NASH and ESSENCE trials using aligned endpoints and statistical methods [44]. The ESSENCE trial reported MASH resolution rates of 63% with semaglutide and 34% with placebo. When the MAESTRO-NASH endpoint definition (≥ 2-point reduction in NAFLD Activity Score [NAS]) was applied to ESSENCE, the placebo response fell to 19%, closer to the 9% observed in MAESTRO-NASH. The proportion of patients achieving a two-point NAS reduction was 25% in both placebo arms. These findings suggest that the higher placebo response of 34% in ESSENCE could have been attributable to low NAS in the reassessed baseline biopsies rather than true histologic improvement. More broadly, the reliance on binary histologic endpoints to measure what are continuous biological processes limits the sensitivity to treatment effects. Imprecision in histologic

scoring introduces regression to the mean and misclassification bias at both baseline and post treatment histologic results, which can dilute true effect sizes and inflate apparent placebo responses [45]. These observations support the continued development of non-invasive biomarkers as alternatives to histologic endpoints in clinical studies. This highlights how biopsy assessment at baseline including tissue adequacy, scoring thresholds and pathologist variability can influence apparent placebo response rates and potentially obscure the magnitude of treatment effect. Standardised reporting of biopsy quality metrics would help distinguish genuine placebo responses from suboptimal or variable baseline histologic assessment.

Currently, MASH trials employ centralised or consensus pathology review and robust randomisation procedures, which should minimise systematic imbalances in biopsy quality between arms. However, because needle gauge, core length and portal tract counts are rarely reported by treatment group, it is not possible to exclude subtle imbalances in tissue adequacy. Our intention is not to imply that such imbalances commonly occur, but to underscore that transparent reporting would allow this assumption to be empirically verified.

To allow for a more balanced comparison of arms of a MASH clinical study, consideration could be given to pre-specified sensitivity analyses that exclude liver biopsies deemed insufficient in quality or quantity. This analysis would report outcomes only for samples both pre and post treatment that meet predetermined biopsy criteria. We recognise that excluding low-quality biopsies may preferentially remove patients with milder inflammatory activity or less advanced disease, thereby altering the underlying disease spectrum and potentially inflating observed effect sizes. Such approaches must therefore be clearly labelled as sensitivity analyses and interpreted cautiously. Ideally, patient selection and biopsy quality should be optimised at enrollment through standardised acquisition protocols and real-time adequacy checks, rather than relying on post hoc exclusions at analysis.

Based on available literature, narrow (18-gauge) and short ($\leq$1.5 cm) biopsy cores often result in low-quality samples with low numbers of portal tracts (Figure 1). Core samples $\leq$2.0 cm may erroneously underestimate inflammation and fibrosis. An acceptable alternative may be to obtain a cumulative biopsy length of 4 to 6 cm with an 18-gauge needle, which might be sufficient to determine staging, although this approach has not been validated in a MASH trial setting. Both 16-gauge and 18-gauge needles for liver biopsies have demonstrated similar safety profiles.

The emergence of AI-assisted pathology systems offers a complementary approach to addressing inter-reader variability, and clinical validation studies have demonstrated that AI tools can achieve reproducibility exceeding published interpathologist agreement for MASH histologic components [38]. Notably, lower placebo response rates have been observed with AI-assisted scoring compared to manual pathologist scoring. The standardisation of interpretation may be as important as the standardisation of biopsy acquisition. The performance of AI models remains fundamentally dependent on adequate biopsy material. Unleashing the full impact of artificial intelligence and digital pathology can only be accomplished if adequate liver biopsy material is delivered to optimise the results.

The quality of the liver biopsy will also be important in validating non-invasive tests. The algorithm for the evaluation of patients at risk for NAFLD in primary care from the AASLD recommends primary risk assessment with the FIB-4 index and secondary risk assessment with VCTE or ELF. In specialty care, MRE or MRI-corrected T1 (cT1) can be considered for further assessment. Liver biopsy should be considered in patients with diagnostic uncertainty or discordant NITs [46]. The November 2025 AASLD Practice Guidance update now recommends identifying candidates for resmetirom and semaglutide treatment using NITs such as VCTE (8–15 kPa), MRE (3.1–4.4 kPa), or ELF (9.2–10.5) rather than liver biopsy, which is impractical in routine clinical care [47]. It is important that calibration of NITs occurs against high-quality histologic reference standards. If the sample quality of the liver biopsy used as the reference standard is variable, then it will be difficult to optimally validate non-invasive liver disease assessments, which will lead to errors in diagnosis and staging of liver disease, both in clinical research and in clinical practice.

In conclusion, as the MASH therapeutic landscape evolves with newly approved and emerging therapies, the rigour applied to liver biopsy acquisition, quality assessment and reporting must keep pace. We propose three recommendations for future MASH clinical trials. First, at minimum, all trials should report quality metrics including needle gauge, core length and number of complete portal tracts by treatment arm in supplementary appendices of their publications. This is a low burden change that allows readers to evaluate balance across arms. Second, when clinically feasible, we suggest adopting preferred biopsy acquisition standards including the use of a 16-gauge needle, a core length $\geq$2.0 cm and $\geq$10 complete portal tracts. These thresholds represent reasonable extrapolations from mixed aetiology and viral hepatitis cohorts, not prospectively validated MASH populations, and should not be treated as mandatory criteria without further validation. Third, investigators may optionally include pre-specified, clearly labelled sensitivity analyses restricted to adequate biopsies, provided the primary analysis remains based on the full randomised population and results are interpreted cautiously, as such exclusions may shift the underlying disease spectrum. These measures would not only reduce a possible source of heterogeneity in placebo response rates but would also provide the high-quality histologic reference standards needed to validate non-invasive tests and AI-assisted pathology systems. As NIT based surrogate endpoints advance toward regulatory qualification and biopsy sparing trial designs become feasible, ensuring the accuracy of the histologic benchmarks against which these tools are calibrated will be essential. Optimising the liver biopsy today is an investment in the reliability of the non-invasive tools of tomorrow.

Acknowledgements

We thank Frank J. Dutko, Ph.D. for writing assistance and editorial support.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

Eric J. Lawitz = Educator/Speaker: Gilead Sciences, Ipsen, Novo Nordisk Inc., Madrigal; Consultant: 89Bio Inc., AstraZeneca, Atea Pharmaceuticals, Boehringer Ingelheim, Bristol Myers Squibb, Corcept Therapeutics, Eli Lilly and Co., Inventiva, Merck & Co. Inc., Novo Nordisk Inc., Organovo, Regeneron, Sagimet Biosciences. Fred Poordad and Zachary Goodman = None.

Data Availability Statement

The authors have nothing to report.

References

1. Z. M. Younossi, P. Golabi, J. M. Paik, A. Henry, C. Van Dongen, and L. Henry, "The Global Epidemiology of Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH): A Systematic Review," *Hepatology* 77, no. 4 (2023): 1335–1347.

2. C. Kan, K. Zhang, Y. Wang, et al., "Global Burden and Future Trends of Metabolic Dysfunction-Associated Steatotic Liver Disease: 1990-2021 to 2045," *Annals of Hepatology* 30, no. 2 (2025): 101898.

3. C. Estes, Q. M. Anstee, M. T. Arias-Loste, et al., "Modeling NAFLD Disease Burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the Period 2016-2030," *Journal of Hepatology* 69, no. 4 (2018): 896–904.

4. F. Kanwal, B. A. Neuschwander-Tetri, R. Loomba, and M. E. Rinella, "Metabolic Dysfunction-Associated Steatotic Liver Disease: Update and Impact of New Nomenclature on the American Association for the Study of Liver Diseases Practice Guidance on Nonalcoholic Fatty Liver Disease," *Hepatology* 79, no. 5 (2024): 1212–1219.

5. EASL-EASD-EASO, "EASL-EASD-EASO Clinical Practice Guidelines on the Management of Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD)," *Journal of Hepatology* 81, no. 3 (2024): 492–542.

6. Z. D. Goodman, "Role of Liver Biopsy in Clinical Trials and Clinical Management of Nonalcoholic Fatty Liver Disease," *Clinics in Liver Disease* 27, no. 2 (2023): 353–362.

7. T. Palmer, I. Georgiades, D. Treanor, et al., "Improved Tissue Sections for Medical Liver Biopsies: A Comparison of 16 vs 18 g Biopsy Needles Using Digital Pathology," *Journal of Clinical Pathology* 67, no. 5 (2014): 415–419.

8. S. Y. Lei, Y. Han, D. C. Rockey, et al., "An International Multidisciplinary Consensus Statement on Laparoscopic Liver Biopsy," *Cell Reports Medicine* 7, no. 6 (2026): 102831, https://doi.org/10.1016/j.xcm.2026.102831.

9. G. Colloredo, M. Guido, A. Sonzogni, and G. Leandro, "Impact of Liver Biopsy Size on Histological Evaluation of Chronic Viral Hepatitis: The Smaller the Sample, the Milder the Disease," *Journal of Hepatology* 39, no. 2 (2003): 239–244.

10. M. E. Tublin, R. Blair, J. Martin, S. Malik, K. Ruppert, and A. Demetris, "Prospective Study of the Impact of Liver Biopsy Core Size on Specimen Adequacy and Procedural Complications," *AJR. American Journal of Roentgenology* 210, no. 1 (2018): 183–188.

11. R. J. Farrell, P. F. Smiddy, R. M. Pilkington, et al., "Guided Versus Blind Liver Biopsy for Chronic Hepatitis C: Clinical Benefits and Costs," *Journal of Hepatology* 30, no. 4 (1999): 580–587.

12. G. Tian, D. Kong, T. Jiang, and L. Li, "Complications After Percutaneous Ultrasound-Guided Liver Biopsy: A Systematic Review and Meta-Analysis of a Population of More Than 12,000 Patients From 51 Cohort Studies," *Journal of Ultrasound in Medicine* 39, no. 7 (2020): 1355–1365.

13. L. B. Seeff, G. T. Everson, T. R. Morgan, et al., "Complication Rate of Percutaneous Liver Biopsies Among Persons With Advanced Chronic Liver Disease in the HALT-C Trial," *Clinical Gastroenterology and Hepatology* 8, no. 10 (2010): 877–883.

14. B. A. Davison, S. A. Harrison, G. Cotter, et al., "Suboptimal Reliability of Liver Biopsy Evaluation Has Implications for Randomized Clinical Trials," *Journal of Hepatology* 73, no. 6 (2020): 1322–1332.

15. E. M. Brunt, A. D. Clouston, Z. Goodman, et al., "Complexity of Ballooned Hepatocyte Feature Recognition: Defining a Training Atlas for Artificial Intelligence-Based Imaging in NAFLD," *Journal of Hepatology* 76, no. 5 (2022): 1030–1041.

16. S. Dhingra, J. D. Mahadik, Y. Tarabishy, S. B. May, and J. M. Vierling, "Prevalence and Clinical Significance of Portal Inflammation, Portal Plasma Cells, Interface Hepatitis and Biliary Injury in Liver Biopsies From Patients With Non-Alcoholic Steatohepatitis," *Pathology* 54, no. 6 (2022): 686–693.

17. S. A. Harrison and J. Dubourg, "Liver Biopsy Evaluation in MASH Drug Development: Think Thrice, Act Wise," *Journal of Hepatology* 81, no. 5 (2024): 886–894.

18. L. Castera, M. E. Rinella, and E. A. Tsochatzis, "Noninvasive Assessment of Liver Fibrosis," *New England Journal of Medicine* 393, no. 17 (2025): 1715–1729.

19. I. N. Kazi, L. Kuo, and E. Tsai, "Noninvasive Methods for Assessing Liver Fibrosis and Steatosis," *Gastroenterology & Hepatology (New York, N.Y.)* 20, no. 1 (2024): 21–29.

20. R. K. Sterling, A. Duarte-Rojo, K. Patel, et al., "AASLD Practice Guideline on Imaging-Based Noninvasive Liver Disease Assessment of Hepatic Fibrosis and Steatosis," *Hepatology* 81, no. 2 (2025): 672–724.

21. R. K. Sterling, K. Patel, A. Duarte-Rojo, et al., "AASLD Practice Guideline on Blood-Based Noninvasive Liver Disease Assessment of Hepatic Fibrosis and Steatosis," *Hepatology* 81, no. 1 (2025): 321–357.

22. E. Truong, L. Alnimer, J. A. Gornbein, et al., "Agile 3+ and 4 Scores Accurately Predict Major Adverse Liver Outcomes, Liver Transplant, Progression of MELD Score, the Development of Hepatocellular Carcinoma, and Death in NAFLD," *Digestive Diseases and Sciences* 70, no. 7 (2025): 2487–2495.

23. Y. Zhang, H. W. Lee, H. Lin, E. Tsochatzis, S. Petta, and E. Bugianesi, "Head-to-Head Comparison Between Vibration-Controlled Transient Elastography and Histology in Predicting Liver-Related Events due to Metabolic Dysfunction-Associated Steatotic Liver Disease," *Hepatology (Baltimore, Md.)* 84 (2025): 175–190, https://doi.org/10.1097/HEP.0000000000001658.

24. J. Fichez, T. Mouillot, L. Vonghia, et al., "Non-Invasive Tests for Fibrotic MASH for Reducing Screen Failure in Therapeutic Trials," *JHEP Reports* 7, no. 4 (2025): 101351.

25. V. Ratziu, M. Hompesch, M. Petitjean, et al., "Artificial Intelligence-Assisted Digital Pathology for Non-Alcoholic Steatohepatitis: Current Status and Future Directions," *Journal of Hepatology* 80, no. 2 (2024): 335–351.

26. P. Sanyal, D. Biswas, and S. Mitra, "Artificial Intelligence in Liver Pathology: Precision Histology for Accurate Diagnoses," *Journal of Clinical and Experimental Hepatology* 15, no. 6 (2025): 103145.

27. A. Taylor-Weiner, H. Pokkalla, L. Han, et al., "A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH," *Hepatology* 74, no. 1 (2021): 133–147.

28. R. Loomba, A. J. Sanyal, K. V. Kowdley, et al., "Randomized, Controlled Trial of the FGF21 Analogue Pegozafermin in NASH," *New England Journal of Medicine* 389, no. 11 (2023): 998–1008.

29. S. A. Harrison, P. Bedossa, C. D. Guy, et al., "A Phase 3, Randomized, Controlled Trial of Resmetirom in NASH With Liver Fibrosis," *New England Journal of Medicine* 390, no. 6 (2024): 497–509.

30. R. Loomba, M. L. Hartman, E. J. Lawitz, et al., "Tirzepatide for Metabolic Dysfunction-Associated Steatohepatitis With Liver Fibrosis," *New England Journal of Medicine* 391, no. 4 (2024): 299–310.

31. R. Loomba, P. Bedossa, K. Grimmer, et al., "Denifanstat for the Treatment of Metabolic Dysfunction-Associated Steatohepatitis: A Multicentre, Double-Blind, Randomised, Placebo-Controlled, Phase 2b Trial," *Lancet Gastroenterology & Hepatology* 9, no. 12 (2024): 1090–1100.

32. A. J. Sanyal, P. Bedossa, M. Fraessdorf, et al., "A Phase 2 Randomized Trial of Survodutide in MASH and Fibrosis," *New England Journal of Medicine* 391, no. 4 (2024): 311–319.

33. K. Cusi, B. Orsak, F. Bril, et al., "Long-Term Pioglitazone Treatment for Patients With Nonalcoholic Steatohepatitis and Prediabetes or Type 2 Diabetes Mellitus: A Randomized Trial," *Annals of Internal Medicine* 165, no. 5 (2016): 305–315.

34. S. A. Harrison, J. P. Frias, G. Neff, et al., "Safety and Efficacy of Once-Weekly Efruxifermin Versus Placebo in Non-Alcoholic Steatohepatitis (HARMONY): A Multicentre, Randomised, Double-Blind, Placebo-Controlled, Phase 2b Trial," *Lancet Gastroenterology & Hepatology* 8, no. 12 (2023): 1080–1093.

35. B. A. Neuschwander-Tetri, R. Loomba, A. J. Sanyal, et al., "Farnesoid X Nuclear Receptor Ligand Obeticholic Acid for Non-Cirrhotic, Non-Alcoholic Steatohepatitis (FLINT): A Multicentre, Randomised, Placebo-Controlled Trial," *Lancet* 385, no. 9972 (2015): 956–965.

36. S. M. Francque, P. Bedossa, V. Ratziu, et al., "A Randomized, Controlled Trial of the Pan-PPAR Agonist Lanifibranor in NASH," *New England Journal of Medicine* 385, no. 17 (2021): 1547–1558.

37. A. J. Sanyal, P. N. Newsome, I. Kliers, et al., "Phase 3 Trial of Semaglutide in Metabolic Dysfunction-Associated Steatohepatitis," *New England Journal of Medicine* 392, no. 21 (2025): 2089–2099.

38. A. J. Sanyal, S. S. Shankar, K. P. Yates, et al., "Diagnostic Performance of Circulating Biomarkers for Non-Alcoholic Steatohepatitis," *Nature Medicine* 29, no. 10 (2023): 2656–2664.

39. FDA, "Letter of Intent Determination Letter to the Foundation for the National Institutes of Health Biomarkers Consortium Non-Invasive Biomarkers of MetaBolic Liver DiseasE (NIMBLE)," 2021, https://www.fda.gov/media/147794/download.

40. F. A. Anania, R. Hager, K. Higgins, G. A. Makar, J. Siegel, and T. T. Tran, "Non-Invasive Tests: Establishing Efficacy for Metabolic Dysfunction-Associated Steatohepatitis Beyond the Biopsy-Current Perspectives From the Division of Hepatology and Nutrition, US Food and Drug Administration," *Hepatology* 84 (2025): 321–327, https://doi.org/10.1097/HEP.0000000000001509.

41. FDA, "Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment," 2018, www.fda.gov/media/119044/download?attachment.

42. F. Idalsoaga, L. A. Diaz, S. Alghsoon, et al., "Placebo Rates in Metabolic Dysfunction-Associated Steatohepatitis Clinical Trials: A Systematic Review and Meta-Analysis," *Clinical Gastroenterology and Hepatology* 24, no. 5 (2025): 1258–1270.

43. C. H. Ng, J. Xiao, W. H. Lim, et al., "Placebo Effect on Progression and Regression in NASH: Evidence From a Meta-Analysis," *Hepatology* 75, no. 6 (2022): 1647–1661.

44. R. Loomba, R. Taub, D. Labriola, N. Alkhouri, and M. Noureddin, "Analysis of MAESTRO-NASH, Resmetirom Relative to Placebo Patients, on Primary and Secondary Liver Biopsy Endpoints Based on Aligned Biopsy Endpoints and Statistical Methods: From MASH Clinical Trials," *Hepatology* 82 (2025): S1855–S1856.

45. A. Shah, L. MacConell, A. Liberman, A. M. Di Bisceglie, and D. Shapiro, "Challenges in Histological Endpoints for MASH Therapies: An Exercise in Statistical Modelling," *Alimentary Pharmacology & Therapeutics* 61, no. 9 (2025): 1489–1499.

46. M. E. Rinella, B. A. Neuschwander-Tetri, M. S. Siddiqui, et al., "AASLD Practice Guidance on the Clinical Assessment and Management of Nonalcoholic Fatty Liver Disease," *Hepatology* 77, no. 5 (2023): 1797–1835.

47. M. B. Bansal, H. Patton, T. R. Morgan, R. M. Carr, J. A. Dranoff, and A. M. Allen, "Semaglutide Therapy for Metabolic Dysfunction-Associated Steatohepatitis: November 2025 Updates to AASLD Practice Guidance," *Hepatology* 83, no. 5 (2026): 1326–1340.