

Review Article

Best Practices in Liver Biopsy Histologic Assessment for Nonalcoholic Steatohepatitis Clinical Trials: Expert Opinion

Claudia M. Filozof ¹, **Carolyn Lackner**,² **Manuel Romero-Gómez**,³ **Joanne C. Imperial**,⁴ **Robert McGee**,⁵ **Lara Dimick-Santos**,⁶ **Oscar Cummings**,⁷ **Cynthia Behling**,⁸ **Troy Johnson**,⁹ and **Arun Sanyal**¹⁰

¹Labcorp Drug Development, Tel Aviv, Israel

²Quantitative Pathomorphology, Institute of Pathology Medical University of Graz, Austria

³UCM Digestive Diseases, Virgen del Rocío University Hospital, Liver Group at Institute of Biomedicine of Seville (IBIS), University of Seville, Sevilla, Spain

⁴Labcorp Drug Development, CA, USA

⁵Labcorp Drug Development, Indianapolis, IND, USA

⁶Labcorp Drug Development, DC, USA

⁷Lawrence M. Roth Professor of Pathology, Indianapolis, IND, USA

⁸Pacific Rim Pathology, San Diego CA, USA

⁹Labcorp Drug Development, Princeton, NJ, USA

¹⁰Division of Gastroenterology, Hepatology & Nutrition, And Education Core Center for Clinical and Translational Research At Virginia Commonwealth University in Richmond, VA, USA

Correspondence should be addressed to Claudia M. Filozof; claudia.filozof@labcorp.com

Received 29 April 2022; Revised 23 June 2022; Accepted 1 July 2022; Published 19 July 2022

Academic Editor: Stephen D. H. Malnick

Copyright © 2022 Claudia M. Filozof et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. In most clinical trials focusing on precirrhotic nonalcoholic steatohepatitis (NASH), a liver biopsy is required for confirmation of diagnosis, staging fibrosis, and grading steatohepatitis activity. Reliance on the biopsy, both as a requisite for study entry, as well as for a primary endpoint in clinical trials, poses several challenges that need to be overcome: patient reluctance to undergo the procedure; potential sampling error; concern regarding the handling, processing and shipping of the biopsy of the biopsy material to the central reader(s); and the degree of pathologists' intra- and interobserver variability in biopsy interpretation. **Aims.** To provide recommendations for improving the liver biopsy process in order to maximize the accuracy of its histological interpretation in NASH clinical trials. **Methods and Results.** These recommendations were created by an expert panel of participants from the United States and European Union who met multiple times and reached alignment through review of available data and their individual clinical experiences. The recommendations include the methodology for biopsy procedure, central lab and pathology processing of the specimen, and recommendations to minimize the intra- and intersubject variability. Finally, we are discussing digital pathology technology and machine learning applications as important additions to enhance liver biopsy interpretation. **Conclusions.** Liver biopsy poses multiple challenges in clinical trials in NASH, and there is a need to standardize the processes to maximize accuracy and minimize variability. Many questions remained unanswered due to limited available data. New evolving modalities may help in the future, but generation of robust data is warranted.

1. Introduction

Nonalcoholic fatty liver disease (NAFLD) is one of the most common forms of chronic liver disease [1]. Nonalcoholic steatohepatitis (NASH), the more advanced stage of NAFLD, is characterized by the presence of hepatic steatosis, lobular inflammation, and hepatocellular ballooning. NASH is a driver of liver fibrosis that can lead to cirrhosis, hepatocellular carcinoma, and death [2]. The worldwide prevalence of NAFLD ranges between 22% and 55%, highest in patients with type 2 diabetes mellitus (T2DM) [3, 4]. The overall prevalence of NASH ranges from 1.5% to 6.5%, and estimated global prevalence (10 studies) of NASH in patients with T2DM is 37.3% [4].

There are presently no approved therapies for the treatment of NASH in the United States (US) or European Union (EU) [5]. Due to its high prevalence, associated morbidity, growing burden of end-stage liver disease, and limited availability of livers for organ transplantation, identifying therapies that will slow the progress, halt, or reverse NASH is clearly an unmet medical need [6].

Currently, accepted surrogate endpoints for accelerated (US) or conditional (European Medicines Agency) approval of a therapeutic agent for an indication in precirrhotic NASH are based on improvements in liver histology (e.g., histological resolution of NASH with no worsening of fibrosis and/or at least a 1-point improvement in fibrosis with no worsening of NASH) [6]. Additionally, demonstration of long-term clinical benefit is required as part of the full approval process [6]. Clinical outcomes acceptable to Regulatory Agencies include improvement in composite endpoint of histological “progression to cirrhosis,” hepatic decompensation events, progression in model end-stage liver disease (from ≤ 12 to ≥ 15), liver transplantation, and all-cause mortality [1, 7].

There are nearly 200 compounds or agents in various stages of development for the treatment of NASH. However, many of them have failed to demonstrate an improvement in the surrogate histological endpoints [8]. Discordance in liver biopsy interpretation around the diagnosis, staging, and grading of NASH poses significant challenges [9]. Liver biopsy is an invasive and costly procedure with potential for complications, and its diagnostic accuracy depends on obtaining an adequate tissue specimen. Although definitive grading and staging systems for the diagnosis of NASH have been nearly universally adopted, few trials have implemented a comprehensive, standardized approach to the acquisition, processing, and interpretation of liver biopsies. Increased awareness and education about the need for standardizing the biopsy process is an important first step in assessment of the histologic endpoints. There is a high unmet medical need to establish best practices to generate an accurate and reproducible specimen for assessment of liver biopsies in clinical trials of NASH.

2. Materials and Methods

A panel of experts (4 pathologists, 3 hepatologists, a board-certified physician specializing in nutrition and metabolism,

a gastrointestinal surgeon, and a biostatistician) met to focus on the challenges using liver histology endpoints in clinical trials, strategies for reducing intra and inter-reader variability, and methods to improve specimens received by the pathologist. The selection of experts was based on 1, extensive experience in clinical trials in NASH; 2, representation of main specialty areas involved in these trials; and 3, the European and USA representation.

The available data from the literature references provided along with each member's experience were considered during the meetings. Consensus was achieved if >90% of members voted in favor of a recommendation. Based on the type of evidence, each recommendation was graded using a framework (Table 1).

3. Panel Recommendations for Optimization of Histological Interpretation of Liver Biopsies for Clinical Trials in NASH

3.1. Biopsy Procedure

- (i) The liver biopsy should be performed by an experienced clinician trained per his/her local regulations. Percutaneous biopsy is the preferred method. The procedure should be performed under radiologic guidance (e.g., ultrasound), if possible [10] (Class 2a, Level C). The tissue samples obtained percutaneously should be collected with a 16-gauge needle (Class 2b, Level C) as there is no definitive data to suggest that a 16-gauge needle carries a higher risk of complications (e.g., hemorrhage) than an 18-gauge needle [11]. The liver biopsy specimen should be 2.5 cm but no less than 1.5 cm in length (Class 2a, Level C) [11]
- (ii) If a transjugular liver biopsy is selected due to evidence of coagulopathy or other medical issues, liver tissue should be obtained using an automated Trucut-type transjugular liver biopsy needle system. Although the use of a 16-gauge needle is preferred, if a smaller (18- or 19-gauge needle) is used, 3 or more passes are recommended to collect sufficient tissue (3.0 cm or more) for analysis (Class 2b, Level C) [12]
- (iii) Due to the potential intersubject variability, the baseline and post-treatment biopsy should be performed by the same operator, ideally in a similar location within the liver (Class 2b, Level C), using similar technique (e.g., needle type and size, percutaneous vs transjugular approach).
- (iv) The biopsy specimen should be obtained from the right lobe whenever possible because increased fibrous septae present near the capsule of the liver may lead to over reading of fibrosis stage. Since the left lobe is thinner, it is difficult to obtain a deep tissue specimen with a percutaneous method (Class 2B, Level C)

TABLE 1: Grading framework.

| | Definition |
|----------------------------|--|
| <i>Category grade</i> | |
| Class I | There is agreement of the proposed procedure is beneficial, useful, and effective |
| Class II | Conflicting evidence and/or divergence of opinion about the usefulness/efficacy of the procedure |
| Class IIa | Weight of evidence/opinion is in favor of usefulness |
| Class IIb | Usefulness is less well established by evidence/opinion |
| Class III | Evidence and/or general agreement that a procedure is not useful/effective, and in some cases, it may be harmful |
| <i>Quality of evidence</i> | |
| Level A | Data derived from multiple randomized, clinical trials or meta-analysis |
| Level B | Data derived from a single randomized, clinical trials or meta-analysis or nonrandomized studies |
| Level C | Based on opinion of experts, case-studies, or standard of care |

- (v) Suction biopsy devices should be avoided as they can fragment tissue, which is more likely to occur in a severely fibrotic liver [12]

able formalin exposure time, including the shipping/transport time (Class 2b, Level C)

3.2. Central Lab Processing of the Tissue

- (i) An original uncut archival formalin-fixed paraffin-embedded (FFPE) tissue block is the preferred method (in the context of a clinical trial) to ship to a central processing laboratory. A previously cut FFPE is the second best choice. Unstained tissue section(s) placed on positively charged glass slides are the third best choice, and wet tissue specimens are the fourth choice. Wet tissue requires careful handling to prevent fragmentation of tissue during shipment. Locally stained slides used for local diagnosis are the least desirable for submission into a clinical trial, due to possible variability in tissue section thickness and hematoxylin and eosin (H&E) staining compared to sections produced and stained by a central laboratory. Wet tissue biopsy samples should be immersed in 10% neutral buffered formalin (NBF) immediately after collection to limit cold ischemic time [11]. The preferred fixative is 10% NBF in the US and is widely accepted and available on a global scale, although may not be used in some countries. We recommend standardizing the fixative across a global clinical trial with 10% NBF. Alternative fixatives (e.g., 20% buffered or unbuffered formalin, 16% buffered formalin, Bouin's fixative) could potentially induce histologic artifacts, which may impact the subsequent central pathology review (Class 2a, Level C)
- (ii) The time spent in tissue fixative may alter the histologic staining of the biopsy or affect future downstream genomic testing. The preferred fixation time is no less than 6 hours, with a maximum of 72 hours. For a wet biopsy to be shipped to a central laboratory, steps must be taken to ensure that the fixation time should not exceed 72 hours. These may include transfer to a 70% ethyl alcohol following formalin fixation to prevent exceeding the maximum allow-

3.3. Pathology Laboratory Processing of the Specimen

- (i) Ideally, the paraffin block should be sectioned at a central processing laboratory and the slides numbered as they are cut from the block. However, especially for baseline liver biopsy, the local laboratory may have a mandate to read the specimen, and obtaining samples for a clinical trial is secondary to their primary mandate. In this instance, the local lab may prepare the core specimen and either send unstained slides (at minimum 6-8 slides) to the central lab or send the remaining specimen embedded in paraffin (FFPE block) to the central lab, the later method being preferred. In general, multiple sections are taken by the local lab (e.g., 20) and at minimum 6 different stains are performed to assess for other liver diseases and perform a thorough evaluation. This complete evaluation may not be necessary for an EOT specimen, especially if the duration between samples is short. Communication by the site investigators with the local pathologist is essential in obtaining good quality specimens for the study
- (ii) For clinical trials, this complete assessment is generally unnecessary as most other liver diseases have been ruled out by exclusion criteria. In general, if the participant has a previous thorough pathologic evaluation, other staining is not necessary for NASH trials, unless requested by the reading pathologist. Two tissue sections closest to the biopsy core (center of the sample) should be used for staining (Class 2b, Level C). The first slide stained with H&E is used to identify and grade lobular inflammation, ballooning, and steatosis. The second slide stained with a Masson trichrome or Picrosirius red is used to identify and stage fibrosis. The use of H&E trichrome stained slides is based on the National Institutes of Health-sponsored Nonalcoholic Steatohepatitis Clinical Research Network

(NASH CRN) recommendations [13]. The stained-glass slides or whole slide digital images of the stained-glass slides (discussed below) should be reviewed by the central pathologist(s) not only for specimen adequacy (size and number of portal tracks), but also for staining quality and artifacts. Moreover, backup slides should be available for staining and review if the pathologist notes artifact or staining problems with the initial slides or images

- (iii) It has recently been proposed to adopt a more comprehensive, structured reporting style—known as synoptic reporting—for use in clinical trials. This would include several stains [14]. However, agreement on what constitutes the minimum dataset and which features should be included still needs to be defined and the final outcome accepted by health authorities
- (iv) Most clinical trials have used only one slide of each stain for evaluation for baseline and end-of-treatment (EOT) liver biopsy. However, it may be reasonable to review more than one section from the specimen and select one that is representative for readings. Currently, it is unknown if review of more than 1 slide from a specimen will increase accuracy of grading and staging in the context of clinical trials. Additionally, the potential subjectivity of selecting “the most representative one” may lead to bias in the interpretation of data. Further data are warranted for evaluation
- (v) In most trials, pre- and post-liver biopsies are read separately. Some trials have added the baseline biopsy with the EOT biopsy for final reading, but not labeled the order of the slides. Blinding may be compromised with this method as several years may have passed which may be apparent to the reader. However, it is unclear if baseline and EOT specimens should be grouped together for readings or read separately. There is no data to support one method over the other, and further studies comparing these two methods may be considered.

3.4. Histopathological Interpretation

- (i) The minimal criteria for the histopathologic diagnosis of NASH mandates the pathologists’ overall global interpretation of steatohepatitis. This overall diagnosis considers many lesions that can be seen with different diseases (e.g., Mallory-Denk Bodies), location of inflammation, among others [13]. The morphological pattern of NASH is complex. Some of the typical changes considered in semi-quantitative grading and staging for clinical trials of adult NASH are summarized below:
 - (1) The increased expression of lipogenesis genes and reduced expression of genes involved in β -oxidation of fatty acids in centrilobular as com-

pared to periportal hepatocytes is, among other factors, responsible for the centrilobular accentuation of types of hepatocellular injury, including macrovesicular steatosis and ballooning. Hepatocellular ballooning is characterized by swelling and rounding as well as rarification of the cytoplasm; the latter change is also referred to as cytoplasmic clarification. Hepatocellular injury is associated with mild inflammation and accumulations of pericellular collagen, a type of fibrosis characteristic for fatty liver disease termed pericellular fibrosis. The inflammatory infiltrates consist mainly of mononuclear cells and eventually admixed neutrophils. Although not accepted by all pathologists, macrovesicular steatosis, hepatocellular ballooning, and lobular inflammation are proposed as the minimum criteria for the histological diagnosis of NASH

- (2) Ongoing liver injury and hepatocellular ballooning-associated sonic hedgehog signaling contribute to the expansion of pericellular fibrosis from centrilobular to periportal portions of the hepatic lobules eventually linking central veins and portal tracts by fibrous septa. Disease progression is also associated with portal inflammation and ductular reaction-triggered periportal fibrosis. Lobular- and portal-based fibrogenesis contribute to the destruction of the lobular architecture, the formation of parenchymal nodules surrounded by fibrous septa characterizing the cirrhosis stage. In advanced fibrosis and cirrhosis, the centrilobular location as well as features of liver injury, steatosis, and hepatocellular ballooning may no longer be present
 - (ii) The most broadly used grading and staging scoring system is the one adapted by the NASH CRN group in 2005 [13]. The NASH CRN system describes disease activity (grade) by the NAFLD activity score (NAS), whereas fibrosis (stage) is defined by the CRN staging system [15]
 - (iii) Another more recently developed grading and staging system is the steatosis, activity, and fibrosis score (SAF). In contrast to the NAS which is the sum of semi-quantitative scores for steatosis hepatocellular ballooning and lobular inflammation, these items are graded separately in the SAF system. Only the prognostic relevant parameters ballooning and inflammation but not steatosis which does not influence prognosis are considered for grading of disease activity. The activity score (sum of ballooning and lobular inflammation scores) and stage can be used to define mild ($A < 3$ and/or $F < 3$) and substantial (A and/or $F \geq 2$) severity of NAFLD with higher reported interobserver agreement [16] and may be considered as an alternative for patient stratification in clinical trials. The clinical utility of the SAF has also been demonstrated [17]

TABLE 2: Intra and interobserver agreement of the histologic interpretation of nonalcoholic steatohepatitis.

| | Reference | Kappa coefficient | |
|----------------------|------------------|-------------------|-------------------|
| | | Interobserver | Intraobserver |
| Steatosis | Younossi et al. | 0.64 | 0.64 ^a |
| | Kleiner et al. | 0.79 | 0.83 |
| | Fukusato et al. | 0.54 | n.a. ^b |
| | Gawrieh et al. | 0.74 ^c | 0.75 ^c |
| | Bedossa et al. | 0.61 ^c | n.a. |
| | Davison et al. | 0.61 ^d | 0.75 ^e |
| | Sanyal AJ et al. | 0.89 | 0.81 |
| Ballooning | Younossi et al. | 0.50 | 0.51 ^a |
| | Kleiner et al. | 0.56 | 0.66 |
| | Fukusato et al. | 0.14 | n.a. |
| | Gawrieh et al. | 0.18 ^c | 0.56 ^c |
| | Bedossa et al. | 0.80 ^c | n.a. |
| | Davison et al. | 0.52 ^d | 0.66 ^e |
| | Sanyal AJ et al. | 0.62 | 0.75 |
| Lobular inflammation | Younossi et al. | 0.33 | 0.62 ^a |
| | Kleiner et al. | 0.45 | 0.60 |
| | Fukusato et al. | 0.10 | n.a. |
| | Gawrieh et al. | 0.20 ^c | 0.48 ^c |
| | Bedossa et al. | 0.72 ^c | n.a. |
| | Davison et al. | 0.33 | 0.44 ^e |
| | Sanyal AJ et al. | 0.60 | 0.61 |
| Fibrosis stage | Younossi et al. | 0.60 | 0.73 ^a |
| | Kleiner et al. | 0.85 | 0.84 |
| | Fukusato et al. | 0.55 | n.a. |
| | Gawrieh et al. | 0.56 ^c | 0.75 ^c |
| | Bedossa et al. | 0.84 ^c | n.a. |
| | Davison et al. | 0.48 ^d | 0.78 ^d |
| | Sanyal AJ et al. | 0.82 | 0.75 |

a = average, 2 observers; b = not analyzed; c = kappa value after the pathologists were trained by a tutorial; d = average, 3 observers; e = average, 2 observers and 2 different set of samples.

(iv) For precirrhotic NASH marketing authorization trials, patients with biopsy-confirmed NASH with NAFLD activity score (NAS) of ≥ 4 , with at least 1 point in each one of the components (steatosis, ballooned hepatocytes, and inflammation), and a fibrosis stage of F2-F3, have been the target population requested by Regulatory Agencies [6]. Patients with concomitant liver diseases (e.g., PSC, PBC, sarcoidosis, alcoholic use disorder, and autoimmune hepatitis) are excluded. This population, with more advanced liver disease, was selected secondary to the practical need to show clinical outcomes in a reasonable period of time for clinical trials

3.5. Observer Variability in NASH Trials

(i) Approximately 65% to 73% of subjects who underwent biopsy screened for clinical trials may not meet these eligibility criteria, contributing to the

high screen failure rates seen in NASH trials [18]. The intra- and interobserver variability around the histologic assessment of the biopsy contributes to uncertainty in the interpretation of the biopsy endpoint [19]. The intra- and inter-rater agreement, kappa (κ) score (an acceptable score to assess reliability for qualitative or categorical variables), is usually acceptable in global interpretation of steatohepatitis (yes/no) or cirrhosis (yes/no). However, the reported κ scores for some of the key features of NASH are low (mainly in inflammation and ballooning scores). Several studies involving histologic interpretation of NASH have identified concordance discrepancies between/among pathologists as defined by the κ score [20] (see Table 2).

(ii) Interobserver variability is particularly high for the identification of ballooned cells. While there are well defined histologic criteria describing a ballooned

hepatocyte, there are discrepancies in the definition in the early stages of formation. A recent paper concludes, “The substantial divergence in hepatocyte ballooning identified amongst expert hepatopathologists suggests that ballooning is a spectrum, too subjective for its presence or complete absence to be unequivocally determined as a trial endpoint [21]”

- (iii) Reasons for low interobserver concordance could also be related to technical problems. In particular, inadequate biopsy length and poor quality of the histology, like inappropriate thickness or inadequate staining as well as fragmentation or folding of sections, making it difficult for pathologists to agree on scorings. Another, maybe more important cause, is that the definitions of the scoring categories of the NAS and the SAF offer a range of possible interpretations leading to variable application of rules for the semiquantitative assessment which depend on opinions of individual pathologists. The problem is aggravated by variable histological definitions of key features of NAFLD in the literature. Therefore, efforts should be undertaken to standardize definitions for histological lesions and the rules for application of the categorical assessments. The utility of standardized definitions and tutorials has been investigated in studies, some of which reported markedly better interobserver agreement as compared to studies without these measures [15, 16, 22–25]
- (iv) Immunohistochemistry might be helpful for objective classification of ballooned cells. However, there are no studies correlating the different types of ballooned hepatocyte (i.e., grades 1 and 2 or classical vs. non-classical ballooned cells) and immunohistochemical (IH) staining patterns with antibodies against k8/18. Finally, there are no data on the clinical/prognostic utility of immunohistochemical hepatocellular ballooning in the literature. Therefore, currently no recommendations to use IH can be made that are based on published results

It has recently been suggested that a concordance atlas may be used to train AI assistive technologies to reproducibly quantify ballooned hepatocytes that standardize assessment of therapeutic efficacy. This atlas may serve as a reference standard for ongoing work to refine how ballooning is classified by both pathologists and AI [21].

The progression of fibrosis from stage to stage is not a continuum of connective tissue deposition but rather describes location of the connective tissue deposition (F2) plus architectural alterations (F3); on the other hand, some F3 biopsies are nearly F4 (cirrhosis), where some F2 biopsies may have only scarce zone 1 perisinusoidal fibrosis and focal periportal hepatocyte trapping.

Bedossa et al. reported that using the SAF score, together with the fatty liver inhibition of progression (FLIP) algorithm, resulted in an increase in biopsy interpretation concordance when 2 groups of blinded pathologists, initially

categorizing liver biopsies based on their own experience, reinterpreted them using the SAF score and FLIP algorithm. Kappa scores increased from moderate (defined as $\kappa = 0.54$) to substantial ($\kappa = 0.66$) in Group 1 and from fair ($\kappa = 0.35$) to substantial ($\kappa = 0.61$) in Group 2 [16], suggesting that the application of this algorithm based on SAF score could decrease interobserver variability (Table 2).

3.6. Use of Single or Multiple Pathologists for Biopsy Interpretation. In an attempt to decrease uncertainty around the pathologic endpoints, some sponsors request 2 pathologists to read each slide and compare results and a third to be available if there is any disagreement. Some allow the 2 initial pathologists to meet to discuss and agree on interpretation and involve a third one only if they cannot come to an agreement on interpretation.

In a recent study including baseline and 18-month slides for 100 subjects, Sanyal et al. reported κ values between 3 board-certified hepatopathologists comparable to the NASH CRN metrics. Two panels, each with 3 pathologists with identical NASH histology training, read digitized slides. Consensus score for each parameter (fibrosis, inflammation, ballooning, and steatosis) was defined as agreement by ≥ 2 pathologists (mode) within a panel. If mode was not achieved, the slide was flagged for a joint panel read with all 3 pathologists. Within each panel, agreement between 2 of 3 readers (mode) was reached in $\sim 90\%$ of slides. It was concluded that consensus score rates of $\geq 95\%$ based on the mode and median, provide a method for rapid and accurate reading of slides [26].

Use of the 2 or 3 reader approach has some limitations: (1) logistical and operational challenges in a global trial; (2) limited number of pathologists with expertise and/or experience in NASH interpretation; (3) extended turnaround time (TAT) in shipping glass slides to multiple readers; and (4) associated shipping costs and possible slide damage/loss if digital images are not used for the histologic evaluation. Additionally, data to support that the second (or third) pathologist approach improves accuracy in biopsy interpretation is limited.

Any pathologist involved in interpreting histopathology for clinical trials should have an acceptable intra-rater concordance ($\kappa \geq 0.6$) for the diagnosis of NASH and the components of the NAS (Class 2b, Level C). If so, a single pathologist could interpret all trial slides and inter-rater concordance would not be an issue. As it is difficult for a single pathologist to be available for the entire duration of a trial (illness, accident, and family events), a back-up pathologist must be considered. Given the variability in reading, it is critical to assess the selected pathologist intrarater concordance before the study starts (Class 2b, Level C).

If multiple pathologists are involved in interpretation of pathology, and the protocol states that each pathologist will read the slides, the recently proposed approach of reaching consensus score rates of $\geq 95\%$ based on the mode and median might be of help in maximizing consensus. In small studies, 2 or 3 pathologists might set aside a reading time (and provide a professional “recorder” for the results) and review the slides simultaneously on a screen or a broadcasted view of the

TABLE 3: Artificial intelligence based and other new methodologies for quantitative assessment of features of NASH and liver fibrosis.

| Brand name | Available literature |
|--------------------------|---|
| Histoindex | |
| <i>Brief description</i> | (i) Uses SHG/TPEF imaging-based tool to provide an automated, quantitative assessment of histological features pertinent to NASH (fibrosis and components of the NAS). (ii) The generated data quantifies fibrosis, steatosis, ballooning, and inflammation. It provides measurements of disease progression and regression in NASH |
| <i>Advantages</i> | (i) Can stage samples as small as 0.5-1.0 cm (ii) Stain-free imaging may reduce staining-related variation in interpretation (iii) Reproducible qFibrosis; prelim outcome data for HBV/NASH (iv) Stain-free imaging enables co-localization for fibrosis, steatosis, ballooning, and inflammation, which are all obtained on the same slide, which was not possible using conventional methods using multiple staining from consecutive slides. (v) AI-based algorithm is developed based with pathologists and is knowledge based, interpretation of the reading is logical for clinical and pathological assessment |
| <i>Limitations</i> | (i) Slow and expensive: (scanning takes ~2 hours per slide). (ii) Cannot detect both faint and dense collagen simultaneously (iii) Performance best for assessment of degree of steatosis and fibrosis (strong correlation with pathologists' scores for qFibrosis ($r = 0.776$) and qsteatosis ($r = 0.802$) compared to severe inflammation and higher ballooning grades) |
| Biocellvia | |
| <i>Brief description</i> | (i) Uses multiparametric image analysis and computerized analysis of high-resolution digitized whole histological sections. (ii) Combines specific stains and immunohistochemistry with advanced algorithms to quantify key morphometric parameters: Ssteatosis, collagen fibers, inflammation, ballooning, ductular reaction (iii) Whole section analyzed simultaneously |
| <i>Advantages</i> | (i) Short turnaround time (<1 h for 100 slides). (ii) Quantified histology has been validated in preclinical models of liver and lung fibrosis (iii) With superior performance compared with traditional scoring with respect to accuracy, reliability, reproducibility, and speed (iv) Purported to eliminate variability in interpretation and provide better insight into a compound's efficacy (v) Whole section analysis (vi) Provide zonal distribution of fibrosis (perisinusoidal, vascular, and septal) in human |
| <i>Limitations</i> | (i) No quantification of portal fibrosis (ii) Need to check correlation for inflammation in human |
| PathAI | |
| <i>Brief description</i> | (i) PathAI has developed a ML-based approach to liver histology assessment, characterizing disease severity and heterogeneity, and quantifying treatment response in NASH. (ii) Deep convolutional neural networks are leveraged for identifying and characterizing NASH histologic features. |
| <i>Advantages</i> | (i) PathAI ML scores and quantitative features are generated on digitized H&E and tTrichrome slides evaluated by central readers, facilitating integration of ML approach into existing NASH clinical trial workflows. (ii) ML-derived continuous quantitative histologic features capture subtle histologic changes not detectable through manual scoring alone; agreement between ML-derived and manual consensus scores for NASH is stronger than pathologist inter-rater agreement |
| <i>Limitations</i> | (i) The ML models are trained using inputs from NASH pathologists, who demonstrate high inter- and intrarater variability in scoring of NASH biopsies. (ii) PathAI ML algorithms can be sensitive to variability in H&E or tTrichrome stain quality |

TABLE 3: Continued.

| Brand name | Available literature |
|--------------------------|---|
| Fibronest | |
| <i>Brief description</i> | (i) Translational quantitative image analysis for the quantification of fibrosis and associated histological features of NASH (ii) FibroNest-pPredict uses AI to link digital pathology images and outcomes/biomarkers and establish image-based predictive models. |
| <i>Advantages</i> | (i) Quantifies same slide/image as used by pathologists to generate automated continuous scores and augmented pathology images to assess fibrosis severity and disease activity across multiple fibrotic conditions (liver, lung, kidney, skin, muscle, and heart). (ii) Turnaround time approximately 2 weeks. (iii) Availability of preclinical and translational supportive data |
| <i>Limitations</i> | (i) Young company (ii) Recent involvement in pPhase 2-3 trials (iii) Prospective data is pending |

Abbreviations: AI = artificial intelligence; HBV = hepatitis B virus; H&E = hematoxylin and eosin; ML = machine learning; NAS = NAFLD activity score; NASH = nonalcoholic steatohepatitis; qFibrosis = quantitative assessment of liver fibrosis; SHG = second harmonic generation; TPEF = two-photon excitation fluorescence.

microscope. However, this strategy may not be feasible in large trials enrolling hundreds of patients in multiple countries [26]

3.7. Digital Imaging. Whole slide digital imaging (WSDI) incorporates the acquisition of digital images from stained tissue sections and the visualization, analysis, interpretation, transfer, and storage of the resulting data. The digital images are acquired through an opto-electronic mechanism that maps the physical tissue information to a digital file and allows pathologists to remotely review and interpret them [27]. Digital transformation of anatomic pathology services is occurring worldwide, and there are several published experiences [28]. A recent study found comparable results between histological interpretations between the Philips IntelliSite Pathology Solution images and glass slides [29]. Another recent review article of 38 validation studies reported an overall diagnostic concordance between digital pathology and glass slides between 63% and 100%, with a weighted mean of 92.4% [29]. A 25-studies meta-analysis examining 10,410 samples cited an overall concordance of 98.3% (95% confidence interval 97.4 to 98.9). However, most of these studies were done in oncology indications and no study has yet evaluated the overall concordance between digital pathology and glass slides in NASH [29]. The Regulatory Agencies have accepted several digital imaging (DI) applications for use in diagnosis of NASH in clinical use [27, 30–32]; however, at this time, the FDA differentiates between “Diagnosis of NASH” and staging and grading (NAS) by the NASH-CRN scoring systems. FDA would like to examine the performance between optical microscope and reading WSI image. Therefore, from regulatory perspective, FDA is asking for data to validate these two systems to give similar readings for the scoring systems. In addition, submission of some representative images should be sent to the College of American Pathologists for proficiency testing of the potential digital scanner.

Potential advantages of WSDI

- (i) Pathologists can review the same scanned slides simultaneously
- (ii) Reduces time delay due to shipping
- (iii) Decreases breakage or slide loss
- (iv) Shortens TAT for pathology report
- (v) Serves as a surrogate for original stained slides and can be archived
- (vi) Enables images to be used for quantitative digital analysis and biomarker development

The panel also noted that there are important considerations if WSDI is used in clinical trials and that this technology is not universally accepted by all pathologists:

- (i) High quality sections should be scanned at high magnification (40× preferable to 20×)
- (ii) Tissue section thickness may interfere with the ability to detect architectural subtleties
- (iii) Glass slides are typically viewed in 3D, whereas digital images are in 2D; soft tissue section distortions or artifacts on the slide cannot be adjusted with focusing (e.g., identification of ballooned hepatocytes may be more difficult without depth of field)

3.8. Overcoming the Challenges of Semi-quantitative Scoring Systems. Fibrosis stage is the main predictor of overall morbidity and mortality in patients with chronic liver disease [32]. The use of semi-quantitative scoring systems has several limitations. For instance, fibrosis progression does not occur in a linear fashion [33]. Additionally, all scoring systems are based on histological changes in untreated individuals, but they do not account well for changes after successful therapy. Additionally, CRN staging implies that fibrosis progresses from perisinusoidal areas (F1), to

perisinusoidal plus portal fibrosis (F2), to bridging fibrosis (F3). Recent data suggest that does not follow the exact sequential steps in opposite way. Indeed, fibrosis regression was mainly due to regression in the perisinusoidal areas and more common than the reduction in septa parameters [34].

New methodologies under development for qualification and quantification of liver fibrosis may improve accuracy in assessing fibrosis remission. These include collagen proportionate area, which provides a percentage assessment of fibrosis on a continuous scale, but is limited by absence of architectural input. Another methodology is dual-photon microscopy-based quantitation of fibrosis-related parameters, which may be able to better define the dynamics of fibrogenesis and fibrosis resolution. Calculation of detailed variables of collagen fibers may be used to establish algorithm-based quantitative fibrosis scores (e.g., qFibrosis, q-FPs). Artificial intelligence and second harmonic generation-derived algorithms are being explored to further develop qFibrosis scoring methods. The inclusion of these methodologies as exploratory objectives in clinical trials can aid in the generation of the required data. However, at this point, more data and validation of the data will need to be performed and presented to the Regulatory Agencies for clearance before they can be used for evaluation of clinical impact.

Artificial intelligence (AI) or machine learning (ML) techniques are being developed to assist pathologists in reading slides. These AI methods are being set up to provide quantitative digital analysis of the slides and support the pathologist in her/his review and interpretation of the liver biopsy. The AI models currently under development will score slides, but the human pathologist will be required to review the output and identify other factors that an AI/ML model may miss, for example, other superimposed liver disease or cases that do not meet the criteria for diagnosis of NASH on global interpretation by the pathologist (Table 3) [35–45].

4. Conclusions

The unmet therapeutic need to treat or cure NASH points to a need to maximize efforts to improve liver biopsy interpretation for diagnosis and assessment of treatment effect on steatohepatitis and fibrosis in NASH clinical trials. Hence, it is critical to standardize all the steps in the process (from obtaining the tissue specimen to processing and the final histopathological assessment) so that they are performed in a consistent and uniform manner. The design of a trial (including single or >1 pathologist) should consider the phase, number of patients, and duration of the trial. If >1 pathologist is used, intra and inter-reader agreement might be improved by a harmonization step before the study starts to train the pathologists on the criteria for the histological interpretation of key features of NAFLD and liver fibrosis for the study. The use of AI/ML to assist pathologists in the identification of early balloon hepatocytes and/or the use of these new methodologies to minimize intra and inter-subject variability may help in the future but generation of more data is warranted.

Data Availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Conflicts of Interest

Dr. Claudia Filozof, Dr. Joanne C. Imperial, Dr. Lara Dimick-Santos, and Troy Johnson were employed by Labcorp Drug Development: no relevant COI. Dr. Carolin Lackner was a consultant for Dr. Falk Pharma, Galmed, Novo Nordisk, Medpace, Inventiva, and Biocellvia. Dr. Manuel Romero Gomez was a consultant for Abbvie, Alpha Sigma, Allergan, Astra Zeneca, Axcella, BMS, Boehringer-Ingelheim, Gilead, Intercept, Inventia, Kaleido, MSD, Novo Nordisk, Pfizer, Prosciento, Rubió, Siemens, Shionogi, Sobi, and Zydus. Research Grants: Gilead, Intercept, Siemens. Dr. Robert McGee was employed in Labcorp when this paper was written but is now employed in Medpace. Dr. Oscar Cummings, MD has no relevant COI. Dr. Cynthia Behling, MD has no relevant COI. Dr. Sanyal is the president of Sanyal Bio. He has stock options in Genfit, Galmed, Exhalenz, Durect, Tiziana, Algernon, and Indalo. He has served as a consultant to Intercept, Gilead, Bristol Myers Squibb, Novartis, Pfizer, Lilly, Novo Nordisk, Astra Zeneca, Medimmune, Merck, Allergan, Albireo, Boehringer Ingelheim, Celgene, NGM, Glympse, Conatus, Genentech, Tern, Takeda, Hemoshear, Immuron, Surrozen, Poxel, Path AI, Second Genome, Zydus, Chiasma, Surrozen, Poxel, Blade, Pliant, Liposcience, Cymabay, Salix, Ferring, and Teva. His institution has received grants from Intercept, Gilead, Novartis, Merck, Astra Zeneca, Mallinckrodt, Pfizer, Lilly, Salix, and Bristol Myers Squibb.

Authors' Contributions

All authors had equal responsibilities in planning, collecting, drafting, and approval of the consensus document.

Funding

Labcorp provided honorarium fee to the external KOLs.

Acknowledgments

The authors would like to thank Dr. Dakshayani SV, BDS and Mrs. Neelam Joglekar, M.Sc. from Labcorp Scientific Services & Solutions Pvt. Ltd., India, for the medical writing support.

References

- [1] M. E. Rinella, F. Tacke, A. J. Sanyal, Q. M. Anstee, and participants of the AASLD/EASL Workshop, "Report on the AASLD/EASL joint workshop on clinical trial endpoints in NAFLD," *Journal of Hepatology*, vol. 71, no. 4, pp. 823–833, 2019.
- [2] N. Chalasani, Z. Younossi, J. E. Lavine et al., "The diagnosis and management of nonalcoholic fatty liver disease: practice

- guidance from the American Association for the Study of Liver Diseases,” *Hepatology*, vol. 67, no. 1, pp. 328–357, 2018.
- [3] M. Pedrosa, M. Balp, N. Janssens et al., “PGI4- Global prevalence of nonalcoholic steatohepatitis (NASH): findings from a targeted literature review,” *Value in Health*, vol. 21, p. S82, 2018.
 - [4] Z. M. Younossi, P. Golabi, L. de Avila et al., “The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis,” *Journal of Hepatology*, vol. 71, no. 4, pp. 793–801, 2019.
 - [5] V. A. Piazzolla and A. Mangia, “Noninvasive diagnosis of NAFLD and NASH,” *Cell*, vol. 9, no. 4, p. 1005, 2020.
 - [6] Food and Drug Administration, “Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: developing drugs for treatment guidance for industry,” 2018.: <https://www.fda.gov/media/119044/download>.
 - [7] European Medical Agency, “Reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH),” 2018. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-requirements-development-medicinal-products-chronic-non-infectious-liver_en.pdf.
 - [8] Y. Sumida, T. Okanoue, A. Nakajima, and Japan Study Group of NAFLD (JSG-NAFLD), “Phase 3 drug pipelines in the treatment of non-alcoholic steatohepatitis,” *Hepatology Research*, vol. 49, no. 11, pp. 1256–1262, 2019.
 - [9] A. C. Sheka, O. Adeyi, J. Thompson, B. Hameed, P. A. Crawford, and S. Ikramuddin, “Nonalcoholic steatohepatitis: a review,” *Journal of the American Medical Association*, vol. 323, no. 12, pp. 1175–1183, 2020.
 - [10] D. C. Rockey, S. H. Caldwell, Z. D. Goodman, R. C. Nelson, and A. D. Smith, “Liver biopsy,” *Hepatology*, vol. 49, no. 3, pp. 1017–1044, 2009.
 - [11] A. J. Sanyal, S. L. Friedman, A. J. McCullough, L. Dimick-Santos, American Association for the Study of Liver Diseases, and United States Food and Drug Administration, “Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration Joint Workshop,” *Hepatology*, vol. 61, no. 4, pp. 1392–1405, 2015.
 - [12] J. Stift, G. Semmler, K. Wöran et al., “Comparison of the diagnostic quality of aspiration and core-biopsy needles for transjugular liver biopsy,” *Digestive and Liver Disease*, vol. 52, no. 12, pp. 1473–1479, 2020.
 - [13] P. Puri and A. J. Sanyal, “Nonalcoholic fatty liver disease: definitions, risk factors, and workup,” *Clinics in Liver Disease*, vol. 1, no. 4, pp. 99–103, 2012.
 - [14] C. R. Langford, M. H. Goldinger, D. Treanor et al., “Improved pathology reporting in NAFLD/NASH for clinical trials,” *Journal of clinical pathology*, vol. 75, no. 2, pp. 73–75, 2022.
 - [15] D. E. Kleiner, E. M. Brunt, M. Van Natta et al., “Design and validation of a histological scoring system for nonalcoholic fatty liver disease,” *Hepatology*, vol. 41, no. 6, pp. 1313–1321, 2005.
 - [16] Bedossa P, FLIP Pathology Consortium, “Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease,” *Hepatology*, vol. 60, no. 2, pp. 565–575, 2014.
 - [17] F. Nascimbeni, P. Bedossa, L. Fedchuk et al., “Clinical validation of the FLIP algorithm and the SAF score in patients with non-alcoholic fatty liver disease,” *Journal of Hepatology*, vol. 72, no. 5, pp. 828–838, 2020.
 - [18] A. Dennis, S. Mouchti, M. Kelly et al., “A composite biomarker using multiparametric magnetic resonance imaging and blood analytes accurately identifies patients with non-alcoholic steatohepatitis and significant fibrosis,” *Scientific Reports*, vol. 10, no. 1, p. 15308, 2020.
 - [19] P. J. Scheuer, “Liver biopsy size matters in chronic hepatitis: bigger is better,” *Hepatology*, vol. 38, no. 6, pp. 1356–1358, 2003.
 - [20] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochem Medica*, vol. 22, pp. 276–282, 2012.
 - [21] 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*, vol. 76, no. 5, pp. 1030–1041, 2022.
 - [22] 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*, vol. 73, no. 6, pp. 1322–1332, 2020.
 - [23] Z. M. Younossi, T. Gramlich, Y. C. Liu et al., “Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations,” *Modern Pathology*, vol. 11, no. 6, pp. 560–565, 1998.
 - [24] T. Fukusato, J. Fukushima, J. Shiga et al., “Interobserver variation in the histopathological assessment of nonalcoholic steatohepatitis,” *Hepatology Research*, vol. 33, no. 2, pp. 122–127, 2005.
 - [25] S. Gawrieh, D. M. Knoedler, K. Saeian, J. R. Wallace, and R. A. Komorowski, “Effects of interventions on intra- and interobserver agreement on interpretation of nonalcoholic fatty liver disease histology,” *Annals of Diagnostic Pathology*, vol. 15, no. 1, pp. 19–24, 2011.
 - [26] A. J. Sanyal, R. Loomba, Q. M. Anstee et al., “Minimizing Variability and Increasing Concordance for NASH Histological Scoring in NASH Clinical Trials,” *In HEPATOLOGY*, vol. 74, pp. 968A–969A, 2021.
 - [27] Food and Drug Administration (FDA), “Technical performance assessment of digital pathology whole slide imaging devices-guidance for industry and,” *Food and Drug Administration Staff*, pp. 1–27, 2016.
 - [28] M. G. Hanna, V. E. Reuter, O. Ardon et al., “Validation of a digital pathology system including remote review during the COVID-19 pandemic,” *Modern Pathology*, vol. 33, no. 11, pp. 2115–2127, 2020.
 - [29] A. S. Azam, I. M. Miligy, P. K.-U. Kimani et al., “Diagnostic concordance and discordance in digital pathology: a systematic review and meta-analysis,” *Journal of Clinical Pathology*, vol. 74, no. 7, pp. 448–455, 2021.
 - [30] US Food & Drug Administration, “FDA allows marketing of first whole slide imaging system for digital pathology,” 2020. <https://www.fda.gov/news-events/press-announcements/fda-allows-marketing-first-whole-slide-imaging-system-digital-pathology>.
 - [31] Imaging Technology News, “Leica Biosystems receives FDA clearance for Aperio AT2 DX Digital Pathology System 2019,” 2021, <http://www.itnonline.com/content/leica-biosystems-receives-fda-clearance-aperio-at2-dx-digital-pathology-system>.

- [32] M. Garcia-Rojo, D. De Mena, P. Muriel-Cueto, L. Atienza-Cuevas, M. Dominguez-Gomez, and G. Bueno, "New European union regulations related to whole slide image scanners and image analysis software," *Journal of pathology informatics*, vol. 10, no. 1, p. 2, 2019.
- [33] P. Angulo, D. E. Kleiner, S. Dam-Larsen et al., "Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease," *Gastroenterology*, vol. 149, no. 2, pp. 389–397.e10, 2015.
- [34] N. V. Naoumov, D. Brees, J. Loeffler et al., "Digital pathology with artificial intelligence analyses provides greater insights into treatment-induced fibrosis regression in NASH," *Journal of Hepatology*, 2022.
- [35] Y. Masugi, T. Abe, H. Tsujikawa et al., "Quantitative assessment of liver fibrosis reveals a nonlinear association with fibrosis stage in nonalcoholic fatty liver disease," *Hepatology communications*, vol. 2, no. 1, pp. 58–68, 2018.
- [36] 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*, vol. 74, no. 1, pp. 133–147, 2021.
- [37] S. Astbury, J. I. Grove, D. A. Dorward, I. N. Guha, J. A. Fallowfield, and T. J. Kendall, "Reliable computational quantification of liver fibrosis is compromised by inherent staining variation," *The Journal of Pathology. Clinical Research*, vol. 7, no. 5, pp. 471–481, 2021.
- [38] F. Liu, G. B.-B. Goh, D. Tiniakos et al., "qFIBS: an automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis," *Hepatology*, vol. 71, no. 6, pp. 1953–1966, 2020.
- [39] W.-Q. Leow, P. Bedossa, F. Liu et al., "An improved qFibrosis algorithm for precise screening and enrollment into non-alcoholic steatohepatitis (NASH) clinical trials," *Diagnostics*, vol. 10, no. 9, p. 643, 2020.
- [40] B. Wang, Y. Sun, J. Zhou et al., "SHG/TPEF-based image technology improves liver fibrosis assessment of minimally sized needle biopsies," *Hepatology International*, vol. 13, no. 4, pp. 501–509, 2019.
- [41] M. De Rudder, C. Bouzin, M. Nachit et al., "Automated computerized image analysis for the user-independent evaluation of disease severity in preclinical models of NAFLD/NASH," *Laboratory investigation*, vol. 100, no. 1, pp. 147–160, 2020.
- [42] M. Albadrani, R. K. Seth, S. Sarkar et al., "Exogenous PP2A inhibitor exacerbates the progression of nonalcoholic fatty liver disease via NOX2-dependent activation of miR21," *American Journal of Physiology. Gastrointestinal and Liver Physiology*, vol. 317, no. 4, pp. G408–G428, 2019.
- [43] O. Carrasco-Zevallos, A. Taylor-Weiner, H. Pokkalla et al., "OS-1611. AI-based histologic measurement of NASH (AIM-NASH): a drug development tool for assessing clinical trial endpoints," *The International Liver Congress™ EASL-European Association for the Study of the Liver2021*, https://www.natap.org/2021/EASL/EASL_82.htm.
- [44] H. Lara, Z. Li, E. Abels et al., "Quantitative image analysis for tissue biomarker use: a white paper from the digital pathology association," *Applied Immunohistochemistry & Molecular Morphology*, vol. 29, no. 7, pp. 479–493, 2021.
- [45] E. Goacher, R. Randell, B. Williams, and D. Treanor, "The diagnostic concordance of whole slide imaging and light microscopy: a systematic review," *Archives of Pathology & Laboratory Medicine*, vol. 141, no. 1, pp. 151–161, 2017.