Minimizing Variability and Increasing Concordance for NASH Histological Scoring in NASH Clinical Trials

Arun J. Sanyal[1]; Rohit Loomba[2]; Quentin M. Anstee[3]; Vlad Ratziu[4]; Amrik Shah[5]; Macky Natha[6]; Deepa Rajagopalan[7]; Nirav Shelat[6]; Mani Subramanian[6]; Katy Wack[8]; Oscar Carrasco-Zevallos[8]; Murray Resnick[8]; Kris Kowdley[9]; Mary E. Rinella[10]; Stephen A. Harrison[11]; Zobair Younossi[12]

[1]Virginia Commonwealth University, Richmond, Virginia, USA; [2]University of California, San Diego, La Jolla, CA, USA; [3]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; [4]Sorbonne Université, Paris, France; [5]Karma Statistics, LLC, Skillman, NJ, USA; [6]Intercept Pharmaceuticals, Inc., San Diego, CA, USA; [7]Intercept Pharmaceuticals, Inc., New York, NY, USA; [8]PathAI, Boston, MA, USA; [9]Liver Institute Northwest, Seattle, Washington, USA; [10]Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [11]Pinnacle Clinical Research, San Antonio, TX, USA; [12]Betty and Guy Beatty Center for Integrated Research, Inova Health System, Falls Church, VA, USA

Introduction

- Liver histology is the reference standard for predicting therapeutic benefit in clinical trials with patients with nonalcoholic steatohepatitis (NASH)[1-3]
- Histological assessments are limited by sampling variability and subjectivity of interpretation, even among individual expert pathologists, leading to inadequate intra- and inter-reader concordance[2-5]
- The United States Food and Drug Administration (FDA) recommends liver histology assessments for phase 3 clinical trials[1,3]
- The NASH Clinical Research Network (NASH CRN) approach whereby multiple expert pathologists meet in-person to review samples and reach consensus is the current standard for histology assessments[6,7]
- Convening a histopathology committee achieves high concordance, but can be logistically challenging and subject to bias (eg, impact of dominant voices within the committee)
- A recent FDA-issued regulatory perspective proposes using at least 2 pathologists trained in evaluating liver biopsy, with involvement of a 3rd pathologist for discordant readings, as a potential approach to ensure that histological endpoints are reliable and consistent[2,3]

Objective

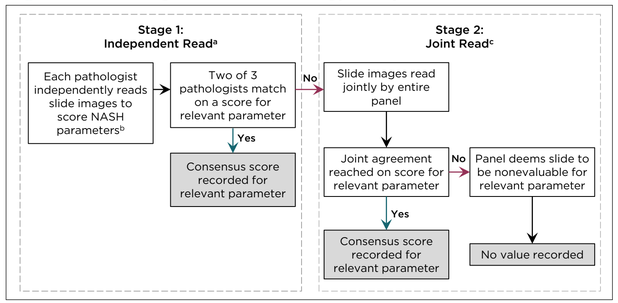
- The primary aims of this analysis were to assess concordance on NASH histological scoring between 3 independent pathologists within each of 2 separate panels and to estimate concordance between the 2 panels for comparison with published NASH CRN concordance estimates

Methods

STUDY DESIGN

- Six board-certified, NASH-trained pathologists who underwent proficiency testing for NASH CRN scoring were allocated to 2 separate panels: Panel A (n=3) or Panel B (n=3)
- Digitized slides taken at baseline and at 18 months from 100 patients with NASH in the ongoing phase 3 REGENERATE study were evaluated (**Figure 1**)
  - In Stage 1, each of 3 pathologists from a panel independently read 4 slides per patient (H&E + Trichrome at baseline and Month 18; 400 slides total) to score fibrosis stage, inflammation, ballooning, and steatosis
  - Slides for which all 3 pathologists in a panel were discordant in Stage 1 were marked for a Stage 2 joint read by all 3 pathologists in that panel

Figure 1. Flow Diagram for Methodology Study



[a]Each reader was blinded to other readers' scores.
[b]Scoring was performed for fibrosis (range: 0–4), inflammation (range: 0–2), ballooning (range: 0–2), and steatosis (range: 0–3).[6]
[c]Panel was blinded to scores from Stage 1 read.

ASSESSMENTS AND ANALYSIS

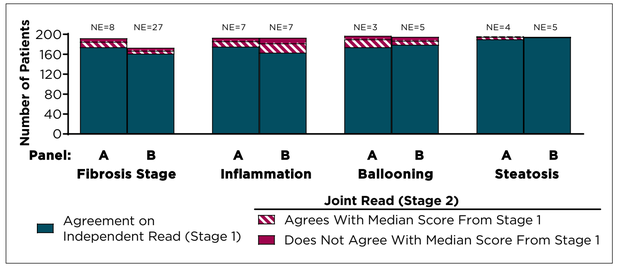
- In Stage 1, if 2 of the 3 pathologists within a panel reported the same score (mode) for a given parameter on a slide, this was chosen as the consensus score
  - If the pathologists were discordant, but a median value existed, the median score was recorded for the parameter; however, if ≥1 pathologist scored the image as nonevaluable, then no median score was recorded
- In Stage 2, a new score was determined based on joint read of each discordant parameter from Stage 1
  - If the panel deemed the slide as nonevaluable, no consensus score was entered for the discordant parameter(s)
- Pairwise kappa scores were determined to assess concordance between readers within each panel (intrapanel agreement) and between the 2 separate panels (interpanel agreement)
  - Kappa scores obtained in the current analysis were compared with previously published values[6-9]

Results

CONCORDANCE RATES

- Overall, high rates of concordance were achieved in Stage 1 for all parameters (**Figure 2**)
  - Highest concordance occurred with steatosis (97%–99%), followed by fibrosis (91%–93%)
  - Agreement rates for ballooning and inflammation were 88%–92% and 84%–91%, respectively

Figure 2. Concordance Rates[a] for Each Parameter by Panel



NE, nonevaluable.
[a]Defined as agreement between ≥2 readers within a panel. The denominator for slides with agreement on independent read is based on the number of evaluable slides.

INTRAPANEL AGREEMENT

- Calculated pairwise kappas from Stage 1 were highest for steatosis and lowest for inflammation (**Table 1**)
  - Values were aligned with those from the initial NASH CRN scoring system development and validation study[6]

Table 1. Concordance[a] and Pairwise Kappas for Each Parameter Within Each Panel and Historical NASH CRN Comparison

Variable	Panel A[b,c] (N=100)	Panel B[b,c] (N=100)	Kleiner 2005[6d] (N=32)
Fibrosis	0.75 (0.61–0.75)	0.71 (0.63–0.71)	0.85
Lobular inflammation	0.61 (0.23–0.61)	0.57 (0.38–0.57)	0.60
Ballooning	0.75 (0.25–0.75)	0.64 (0.44–0.64)	0.66
Steatosis	0.81 (0.69–0.81)	0.87 (0.79–0.87)	0.83

N, number of patients.
[a]Defined as agreement between ≥2 readers within a panel.
[b]Values represent highest intrapanel weighted Shrout-Fleiss kappa (lowest–highest).
[c]Ranges from pairwise kappas from pairs within panel.
[d]Values represent the average intrareader kappa.

INTERPANEL AGREEMENT

- Linear weighted kappa scores between panels A and B reveal concordance rates similar to previously published values from NASH CRN studies (**Table 2**)

Table 2. Comparison of Interpanel Kappa Score Results to Published Literature

	Shrout-Fleiss Weighted Kappa			Cicchetti-Allison Weighted Kappa		
Parameters	Panel A vs B (N=100)	Kleiner 2019[7a] (N=446)	Kleiner 2005[6a] (N=32)	Panel A vs B (N=100)	Davison 2020[8b] (N=339)	Newsome 2021[9b,c] (N=320)
Fibrosis	0.82	0.75	0.84	0.71	0.48	0.61–0.65
Lobular inflammation	0.60	0.46	0.45	0.46	0.33	0.38–0.39
Ballooning	0.62	0.54	0.56	0.51	0.52	0.41–0.61
Steatosis	0.89	0.77	0.79	0.83	0.61	0.63–0.76

N, number of patients.
[a]Average of pairwise kappas.
[b]Pairwise kappas.
[c]Range based on 2 values from baseline and Week 72 slides.
Results from the current analysis are based on nonmissing data.

AGREEMENT BETWEEN MEDIAN VALUE AND JOINT READ

- Overall, 50% to 83% of scores obtained jointly in Stage 2 matched the median value obtained in Stage 1 (**Table 3**)
  - Only 1 joint read was required for steatosis, the result of which did not match the median value from Stage 1

Table 3. Percentage of Slides Reaching Consensus Following Stage 2 Joint Panel Read That Agreed With Stage 1 Median Score

	Panel A	Panel B
Fibrosis		
Consensus reached in Stage 2, n	18	12
Panel consensus agreed with Stage 1 median score, n (%)	11 (61%)	6 (50%)
Inflammation		
Consensus reached in Stage 2, n	18	30
Panel consensus agreed with Stage 1 median score, n (%)	11 (61%)	19 (63%)
Ballooning		
Consensus reached in Stage 2, n	23	16
Panel consensus agreed with Stage 1 median score, n (%)	16 (70%)	8 (50%)
Steatosis		
Consensus reached in Stage 2, n	6	1
Panel consensus agreed with Stage 1 median score, n (%)	5 (83%)	0

Denominators for percentage calculations are based on the number of slides that achieved consensus in the joint panel read.
N, number of slides.

Conclusions

- Independent scoring of histological parameters by a panel of 3 board-certified hepato-pathologists produces high concordance and may reduce bias
- Concordance rates between 2 separate panels are comparable to NASH CRN metrics and underscore panel interchangeability
- The method's consensus rates and kappa values support its accuracy, reproducibility, and potential to reduce uncertainty around treatment effect estimates in NASH phase 3 trials

References

1. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: developing drugs for treatment. Draft guidance for industry. Rockville, MD: US Food and Drug Administration; 2018. **2.** Matsubayashi T. Drug development for nonalcoholic steatohepatitis (NASH) with fibrosis: a regulatory perspective. US Food and Drug Administration. 2021. [webinar]. **3.** Anania FA, et al. *Hepatology*. 2021;73(5):2023-2027. **4.** Ratziu V, et al. *Gastroenterology*. 2005;128:1898–1906. **5.** Vuppalanchi R, et al. *Clin Gastroenterol Hepatol*. 2009;7:481–486. **6.** Kleiner DE, et al. *Hepatology*. 2005;41:1313-1321. **7.** Kleiner DE, et al. *JAMA Netw Open*. 2019;2(10):e1912565. **8.** Davison BA, et al. *J Hepatol*. 2020;73(6):1322-1332. **9.** Newsome PN et al. *N Engl J Med*. 2021;384(12):1113-1124.

Funding Statement

This analysis was sponsored by Intercept Pharmaceuticals, Inc. Medical writing assistance was provided by Peloton Advantage, LLC, an OPEN Health company, and was funded by Intercept Pharmaceuticals, Inc.

Disclosures

Arun J. Sanyal is President of Sanyal Biotechnology; he reports stock options in Sanyal Bio, Exhalenz, Akarna, GenFit, HemoShear, Durect, Indalo, Tiziana, Rivus; paid consultant to 89Bio, Albireo, Amgen, Ardelyx, Boehringer Ingelheim, Bird Rock, Bristol-Myers, Conatus, Covance, Echosens-Sandhill, ENYO, Genentech, General Electric, Genfit, Gilead, HemoShear, HistoIndex, Inventiva, Janssen, Lilly, Madrigal, Mallinckrodt, Merck, NGM Bio, Nimbus, NorthSea, Novartis, Novo Nordisk, Owl, PathAI, Perspectum, Pfizer, Poxel, ProScienta, Regeneron, Rivus, Roche, Salix, Sanofi, Second Genome, Servier, Siemens, Takeda, Terns, Tiziana, Zydus; unpaid consultant to Immuron, Intercept Pharmaceuticals, Inc., Galectin, Sequana, Fractyl, Durect, Indalo, Allergan, Chemomab, Afimmune, Teva, BASF, AMRA, Perspectum, Biocellvia; advisory board member for Immuron; royalties from Elsevier and UptoDate; grant support from Gilead, Mallinckrodt, Salix, Novartis, Galectin, Bristol-Myers, Sequana; receipt of equipment from Echosens-Sandhill; ongoing research collaboration with Echosens-Sandhill, Owl, Second Genome, Siemens; his institution has received grant support from Conatus, Gilead, Mallinckrodt, Boehringer Ingelheim, Novartis, Bristol-Myers, Merck, Lilly, Novo Nordisk, Fractyl, Madrigal, Inventiva, Covance; his institution has received fees from AstraZeneca. Rohit Loomba reports research grants from Intercept Pharmaceuticals, Inc., Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, Galectin Therapeutics, Galmed Pharmaceuticals, Genfit, Gilead, Inventiva, Janssen, Madrigal Pharmaceuticals, Merck, NGM Biopharmaceuticals, Pfizer, Siemens; consulting fees from Alnylam/Regeneron, Amgen, Arrowhead Pharmaceuticals, AstraZeneca, Bristol-Myers Squibb, CohBar, Eli Lilly, Galmed, Gilead, Glympse Bio, Inipharm, Intercept, Ionis, Janssen Inc., Madrigal, Metacrine, Inc., NGM Biopharmaceuticals, Novartis, Novo Nordisk, Pfizer, Sagimet, 89Bio, Viking Therapeutics; advisory committee participation for Intercept Pharmaceuticals, Inc.; co-founder of Liponexus, Inc. Quentin M. Anstee reports support from Intercept Pharmaceuticals, Inc. in relation to the present study; research funds as Coordinator of the EU IMI-2 LITMUS consortium, which is funded by the EU Horizon 2020 programme and EFPIA. This multistakeholder consortium includes industry partners, in addition, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd; royalties or licenses from Elsevier Ltd; consulting fees from 89Bio, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, Bristol-Myers Squibb, BNN Cardio, Cirius, CymaBay, EcoRI, E3Bio, Eli Lilly & Company Ltd., Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Madrigal, Medimmune, Medpace, Metacrine, NGMBio, NorthSea Therapeutics, Novartis, Novo Nordisk A/S, PathAI, Pfizer Ltd., Poxel, ProScienta, Raptor Pharma, Roche, Servier, Terns, The Medicines Company, Viking Therapeutics; honoraria from Abbott Laboratories, Allergan/Tobira, Bristol-Myers Squibb, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, Medscape; member of advisory board for Medpace (NorthSea Therapeutics). Vlad Ratziu reports consulting fees from Galmed, Genfit, Madrigal, NGM, Bristol-Myers Squibb, Boehringer Ingelheim, Theratechnologies, Terns; advisory committee participation for Intercept Pharmaceuticals, Inc. Amrik Shah is an employee of and owns stock options in Intercept Pharmaceuticals, Inc. Macky Natha has no conflicts to disclose. Deepa Rajagopalan was an employee of and owned stock options in Intercept Pharmaceuticals, Inc. at the time of the study. Nirav Shelat has no conflicts to disclose. Mani Subramanian has received consulting fees and owns stock options in Intercept Pharmaceuticals, Inc. Katy Wack is an employee of PathAI. Oscar Carrasco-Zevallos has no conflicts to disclose. Murray Resnick is an employee of and owns stock options in PathAI. Kris Kowdley reports advisory committee or review panel participation for Gilead, Intercept Pharmaceuticals, Inc., Inipharm, Genfit, Enanta, NGM BioPharma, consulting for Corcept Therapeutics, Gilead, Intercept Pharmaceuticals, Inc., Calliditas, HighTide, Madrigal, research support from Intercept Pharmaceuticals, Inc., Terns, Metacrine, Gilead, NGM Bio, Madrigal, Viking, Enanta, HighTide, Pfizer, and Hanmi, speaking and teaching activities for AbbVie, Gilead, Intercept Pharmaceuticals, Inc. Mary E. Rinella was a consultant over the past 36 months for Alnylam, Amgen, AMRA, Bristol-Myers Squibb, Boehringer Ingelheim, Centara, Coherus, Enanta, Galecto, Intercept Pharmaceuticals, Inc., Madrigal, NGM Biopharmaceuticals, Novo Nordisk, Pfizer, Fractyl, Gelesis, Siemens, Thetis, Terns, Rivus, 3vbio (Sagimet), 89Bio, and Novartis. She currently has no active consulting contracts. Stephen A. Harrison is scientific advisor or consultant for Akero, Alentis, Altimmune, Arrowhead, Axcella, Cirius, CymaBay, Echosens, Fibronostics, Forest Labs, Galectin, Genfit, Gilead, Hepagene, Hepion, HistoIndex, Intercept, Madrigal, Medpace, Metacrine, NGM Bio, NorthSea, Novo Nordisk, PathAI, Poxel, Sagimet, Terns, Viking, 89Bio. He owns stock options in Akero, Cirius, Galectin, Genfit, Hepion, HistoIndex, PathAI, Metacrine, NGM Bio, and NorthSea; received grant/research support from Akero, Axcella, Bristol-Myers Squibb, Cirius, CiVi Biopharma, Conatus, CymaBay, Enyo, Galectin, Genentech, Genfit, Gilead, Hepion, Hightide, Intercept, Madrigal, Metacrine, NGM Bio, Novartis, Novo Nordisk, Northsea, Pfizer, Sagimet, Viking. Zobair Younossi reports consulting fees and research grants from Gilead Sciences, Intercept Pharmaceuticals, Inc., Bristol-Myers Squibb, Novo Nordisk, Viking, Terns, Siemens, Quest, AbbVie, Madrigal, Merck, and Novartis.

Corresponding Author

Arun J. Sanyal (arun.sanyal@vcuhealth.org)

ClinicalTrials.gov: NCT02548351

Presented at AASLD: The Liver Meeting 2021, American Association for the Study of Liver Diseases; November 12–15, 2021