Received: 28 June 2023 | Accepted: 3 December 2023

DOI: 10.1097/HEP.0000000000000723

ORIGINAL ARTICLE

OPEN

Artificial intelligence scoring of liver biopsies in a phase II trial of semaglutide in nonalcoholic steatohepatitis

Vlad Ratziu[1] | Sven Francque[2,3,4] | Cynthia A. Behling[5] | Vanja Cejvanovic[6] | Helena Cortez-Pinto[7] | Janani S. Iyer[8] | Niels Krarup[6] | Quang Le[8] | Anne-Sophie Sejling[6] | Dina Tiniakos[9,10] | Stephen A. Harrison[11]

[1]Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Hôpital Pitié Salpêtrière, Institute of Cardiometabolism and Nutrition (ICAN), Paris, France

[2]Antwerp University Hospital, Antwerp, Belgium

[3]InflaMed Centre of Excellence, Laboratory for Experimental Medicine and Paediatrics, Translational Sciences in Inflammation and Immunology, Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium

[4]European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Antwerp, Belgium

[5]Pacific Rim Pathology, San Diego, California, USA

[6]Novo Nordisk A/S, Søborg, Denmark

[7]Clínica Universitária de Gastrenterologia, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

[8]PathAI Inc., Boston, Massachusetts, USA

[9]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[10]Department of Pathology, Aretaieion Hospital, National and Kapodistrian University of Athens, Athens, Greece

[11]Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Correspondence
Stephen A. Harrison, Pinnacle Clinical Research, 5109 Medical Dr., Suite 200, San Antonio, TX 78229, USA.
Email: stephenharrison87@gmail.com

Abstract

Background and Aims: Artificial intelligence–powered digital pathology offers the potential to quantify histological findings in a reproducible way. This analysis compares the evaluation of histological features of NASH between pathologists and a machine-learning (ML) pathology model.

Approach and Results: This post hoc analysis included data from a subset of patients (n = 251) with biopsy-confirmed NASH and fibrosis stage F1–F3 from a 72-week randomized placebo-controlled trial of once-daily subcutaneous semaglutide 0.1, 0.2, or 0.4 mg (NCT02970942). Biopsies at baseline and week 72 were read by 2 pathologists. Digitized biopsy slides were evaluated by PathAI's NASH ML models to quantify changes in fibrosis, steatosis, inflammation, and hepatocyte ballooning using categorical assessments and continuous scores. Pathologist and ML-derived categorical assessments detected a significantly greater percentage of patients achieving the primary endpoint of NASH resolution without worsening of

Abbreviations: AI, artificial intelligence; CRN, Clinical Research Network; FAS, full analysis set; ML, machine learning; NAS, nonalcoholic fatty liver disease activity score.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepjournal.com.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.
Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

fibrosis with semaglutide 0.4 mg versus placebo (pathologist 58.5% vs. 22.0%, $p < 0.0001$; ML 36.9% vs. 11.9%; $p = 0.0015$). Both methods detected a higher but nonsignificant percentage of patients on semaglutide 0.4 mg versus placebo achieving the secondary endpoint of liver fibrosis improvement without NASH worsening. ML continuous scores detected significant treatment-induced responses in histological features, including a quantitative reduction in fibrosis with semaglutide 0.4 mg versus placebo ($p = 0.0099$) that could not be detected using pathologist or ML categorical assessment.

Conclusions: ML categorical assessments reproduced pathologists' results of histological improvement with semaglutide for steatosis and disease activity. ML-based continuous scores demonstrated an antifibrotic effect not measured by conventional histopathology.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

INTRODUCTION

Accurate assessment of NASH histopathological features is essential for determining prognostic risk, making disease management decisions, and measuring response to therapeutic agents.[1,2] Pathologists' review of liver biopsy specimens is currently the reference standard method for the diagnosis and staging of NASH that is accepted by regulatory agencies.[3] It is therefore used to assess inclusion in NASH therapeutic trials, while histological changes are considered a reasonable surrogate for clinical treatment benefit.[4] However, liver histology assessment by pathologists comes with caveats related to clinical experience and expertise, definitions of histopathological features and their interpretation by individuals or groups of pathologists, and limitations of applying a categorical assessment system to continuous variables.[1,2,5] Variation in the interpretation and assignment of categorical assessment can lead to intraobserver or interobserver variability, which can affect the assessment of NASH resolution, fibrosis regression, and other relevant features according to current guidelines.[1,3,6–8] Moreover, the limited dynamic range of the current semiquantitative classifications may not detect subtle changes in histological features. Thus, there is an unmet need for additional methods for objectively evaluating histological changes.

Artificial intelligence (AI) digital pathology tools have shown promise in evaluating NASH liver biopsy samples. They have the potential to support the pathologists' assessment by providing a quantitative supplement to the qualitative and semiquantitative pathology evaluation. They allow the quantification of histological features as continuous measures—thus enabling the measurement of granular levels of disease progression or improvement that are not detectable through categorical assessment.[9,10]

Machine learning (ML) is an AI application in which computer systems can learn and adapt from experience without explicit programming. It uses computer algorithms to analyze and infer from patterns in data while gradually improving in accuracy. The potential utility of ML for liver histology has been explored in several studies in patients with NASH.[10–13]

The current analysis evaluated liver biopsy samples from a randomized, double-blind, phase II trial that investigated the effect of semaglutide, a glucagon-like peptide-1 receptor agonist, on histological resolution of NASH in patients with biopsy-confirmed NASH and fibrosis.[14] The aim of this post hoc analysis was to compare the evaluation of key histological features of NASH by 2 methods: the traditional independent evaluation by expert liver pathologists and the ML-derived pathology models. The level of agreement between the 2 methods was assessed. The specific objective was to investigate if ML-derived quantitative assessments can uncover histological changes otherwise not detected by conventional histological assessment.

METHODS

Trial design

Details of this randomized, double-blind, placebo-controlled, parallel-group trial (NCT02970942) have been reported.[14] Briefly, 320 patients aged 18–75 years (20–75 years in Japan) with biopsy-confirmed NASH, a fibrosis stage of 1–3, and nonalcoholic fatty liver disease activity score (NAS) ≥ 4 with stages/grades ≥ 1 for each subcomponent (steatosis, hepatocyte ballooning, and lobular inflammation) were randomized to receive once-daily s.c. semaglutide 0.1, 0.2, or 0.4 mg, or placebo for 72 weeks.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

Liver biopsies were obtained up to 21 weeks before screening or during the screening period (ie, baseline) and at week 72 of the study.

Processing of liver biopsies

Liver biopsy tissue sections on glass slides were stained using hematoxylin and eosin and Masson's trichrome stains. This was primarily done at the central laboratory, but some slides were stained locally. Digital scans of glass slides were generated using an Aperio Digital Pathology Slide Scanner (Leica Biosystems) at ×40 (4 slides were scanned at ×20). Digital scanning was implemented in the protocol after the trial started; therefore, only a subset of 251 patients had digitized slides at baseline.

Central pathologist evaluation

As part of the standard operating procedures for the trial, stained glass histopathology slides from all biopsies at baseline and week 72 were assessed by 2 expert liver pathologists. Biopsy slides were sent for review throughout the study as they were performed; pathologists were unaware of the treatment assignments, patients' characteristics, biopsy time point (baseline or week 72), and the other pathologist's scores. A consensus score was reached based on a re-review of scores, sections, and discussion, and this score was used for data analysis in the trial and the present post hoc analysis. NASH diagnosis was confirmed and the NAS features were scored according to the NASH Clinical Research Network (CRN) scoring criteria[15]: steatosis (0–3), lobular inflammation (0–2), and hepatocyte ballooning (0–2), with the sum of scores yielding the NAS. Fibrosis was staged F0–F4.

PathAI's NASH ML models

Details surrounding PathAI's ML model development for convolutional neural networks, graph neural networks, and end-to-end models have been published.[10,16] Further details on model development and the methodology of this analysis can be found in the Supplemental Methods, http://links.lww.com/HEP/I172.

Data presented in this paper were analyzed during September 2020 using the most up-to-date version of the PathAI models available at that time.

Phase II and post hoc exploratory endpoints

Analyses involving ML readouts were performed as post hoc analyses and included the primary (NASH resolution without worsening of fibrosis) and confirmatory secondary (improvement in liver fibrosis of at least 1 stage with no worsening in steatohepatitis) endpoints. The composite primary and confirmatory secondary endpoints were assessed in terms of the percentage of responders. The individual histological components were assessed in terms of percentage of responders and with ranked assessment (categorized as worsening, improvement, or no change). Further endpoints assessed by the ML models were percent change (absolute change) from baseline to week 72 in the proportionate area of lobular inflammation, steatosis, hepatocyte ballooning, liver collagen, and portal inflammation; change from baseline to week 72 in ML-derived continuous fibrosis score (0–4); and change from baseline to week 72 in ML-derived categorical NASH CRN fibrosis stage and grades of steatosis, lobular inflammation, and hepatocyte ballooning. In addition, the interagreement variability between the ML model and pathologist assessment (consensus grade/stage) of liver biopsies for histological parameters was assessed.

Statistical methods

This post hoc analysis was performed on a subset of the main study, consisting of all randomized patients with F1–F3 fibrosis who had a digital liver biopsy slide at baseline [full analysis set (FAS) for AI]; as slide digitization was introduced after the trial had been initiated, digitized baseline slides were not available for all randomized patients. Changes in continuous measures were analyzed by ANCOVA with missing values imputed from the placebo group. Binary endpoints were assessed by Cochran–Mantel–Haenszel tests with missing outcomes imputed as nonresponse. Interagreement variability was evaluated based on descriptive statistics and weighted kappa statistics for patients having the same results based on ML and pathologist evaluations for each of the histopathology components (steatosis grade, lobular inflammation grade, hepatocyte ballooning grade, and fibrosis stage) overall and by treatment group.

Evaluations were made at baseline and for the change from baseline to week 72 outcome (improvement, no change, worsening). The kappa statistics were calculated using Cicchetti–Allison weights that applied penalties depending on the magnitude of the difference between results. Agreement levels based on kappa statistics were interpreted as reported by Landis and Koch.[17]

Ethics

Written informed consent was obtained from each patient included in the study. The study protocol conforms to the ethical guidelines of the 1975

Declaration of Helsinki as reflected in a priori approval by each institution's human research committee. Further information on ethical approval for the study can be found in the Supplemental Materials, http://links.lww.com/HEP/I173.

RESULTS

Patient disposition and biopsy availability

This post hoc analysis included a subset of patients with digitized biopsy slides available. The FAS for AI included a total of 251 patients who had baseline digital slides of liver biopsies. For 3 patients all with available hematoxylin and eosin–stained digitized slides at week 72, the ML model was not able to assess the NASH CRN scores for the steatosis grade, lobular inflammation grade, or hepatocyte ballooning grade at week 72. In addition, for 1 patient with available trichrome-stained digitalized slides at week 72, the ML model was not able to assess the liver collagen proportionate area or the ML-based fibrosis stage. Four slides were excluded from the analysis as they were deemed of insufficient quality by pathologist review. No quality checks other than what is built into the ML model were performed.

The full disposition of patients' liver biopsies and digitized slides is shown in Figure 1.

Demographics and baseline characteristics

The mean age of the FAS for AI was 54.5 years, 61.0% were women, and the mean baseline body weight was 99.0 kg. The mean body mass index was 36.0 kg/m^2. Approximately 64.1% had type 2 diabetes at baseline [mean glycated hemoglobin (HbA$_{1c}$) 7.4%]. The FAS for AI was considered generally representative of the overall trial population (Table 1).

By central pathologist evaluation, 44.6% of the FAS for AI had fibrosis stage 3, 23.9% had stage 2, and 31.5% had stage 1 (Table 1). Patients with stage 0 or 4 were excluded as outlined in the Methods section.

ML-derived categorical assessment of the FAS for AI determined that 3 patients had baseline fibrosis stage 0 and 11 had stage 4 (Supplemental Table S1A, http://links.lww.com/HEP/I172). Some patients had ML-derived categorical NASH CRN scores of 0 for steatosis grade (n = 1, Supplemental Table S2A, http://links.lww.com/HEP/I172), lobular inflammation grade (n = 10, Supplemental Table S3A, http://links.lww.com/HEP/I172), or hepatocyte ballooning grade (n = 10, Supplemental Table S4A, http://links.lww.com/HEP/I172).

Endpoints assessed by central pathologist versus ML

As calculated from the consensus pathologist scores, NASH resolution without fibrosis worsening was achieved by significantly more patients receiving semaglutide 0.4 mg (58.5%) compared with placebo (22.0%; $p = 0.0001$) (Figure 2A). There was also a significant difference for semaglutide 0.1 mg versus placebo (46.3% vs. 22.0%, $p = 0.0097$), but not semaglutide 0.2 mg versus placebo (33.3% vs. 22.0%, $p = 0.1561$).

The proportion of patients achieving NASH resolution without worsening of fibrosis was lower when determined by the ML-derived categorical assessment than by consensus pathologist score (no statistical comparison between methods was performed); nevertheless, ML-derived categorical assessment still found a significant difference between patients treated with semaglutide 0.4 mg (36.9%) and those who received placebo (11.9%; $p = 0.0015$) (Figure 2B). ML-derived categorical assessment also detected a dose-dependent treatment response, with a greater percentage of patients achieving NASH resolution without worsening of fibrosis in the semaglutide 0.2 mg treatment arm relative to the semaglutide 0.1 mg treatment arm.

Fibrosis stage improvement without NASH worsening, as assessed by the consensus pathologist scores,

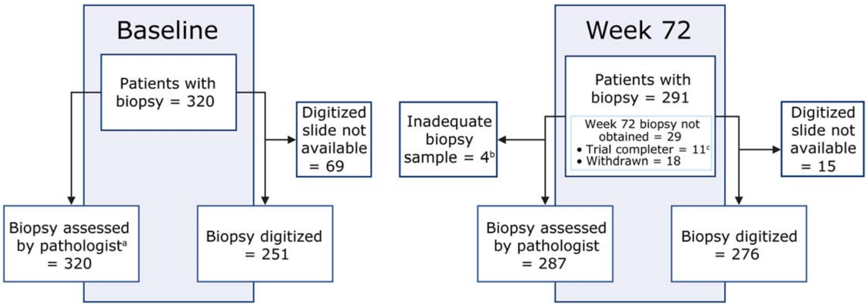

Baseline
Patients with biopsy = 320
Digitized slide not available = 69
Biopsy assessed by pathologist[a] = 320
Biopsy digitized = 251
Week 72
Patients with biopsy = 291
Week 72 biopsy not obtained = 29
• Trial completer = 11[c]
• Withdrawn = 18
Inadequate biopsy sample = 4[b]
Digitized slide not available = 15
Biopsy assessed by pathologist = 287
Biopsy digitized = 276

FIGURE 1 Disposition of patients' liver biopsies and digitized slides. The number of patients with digitized slides at both baseline and week 72 was 221. Only people with a baseline biopsy were included. Missing responses were imputed as nonresponse. [a]For baseline biopsies, a substitute pathologist evaluated 8 (2.5%) of the slides. [b]Four patients had a biopsy that was not centrally read as it was deemed to be of inadequate technical quality by the central pathologist.[c]Trial completers with baseline biopsy data, but no biopsy was performed at week 72.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

TABLE 1 Key demographic and baseline characteristics

Mean ± SD or n (%)	FAS for AI (n = 251)	Full trial population (N = 320)
Age (y)	54.5 ± 10.6	55.0 ± 10.6
Female	153 (61.0)	194 (60.6)
Body mass index (kg/m^2)	36.0 ± 6.3	35.8 ± 6.4
Body weight (kg)	99.0 ± 20.9	98.4 ± 21.7
Type 2 diabetes	161 (64.1)	199 (62.2)
HbA$_{1c}$[a] (%)	7.4 ± 1.2	7.3 ± 1.2
Fibrosis stage[b]		
0	0 (0.0)	0 (0.0)
1	79 (31.5)	90 (28.1)
2	60 (23.9)	72 (22.5)
3	112 (44.6)	158 (49.4)
4	0 (0.0)	0 (0.0)
Hepatocyte ballooning[b]		
0	0 (0.0)	0 (0.0)
1	175 (69.7)	218 (68.1)
2	76 (30.3)	102 (31.9)
Steatosis[b]		
0	0 (0.0)	0 (0.0)
1	72 (28.7)	90 (28.1)
2	129 (51.4)	162 (50.6)
3	50 (19.9)	68 (21.3)
Lobular inflammation[b]		
0	0 (0.0)	0 (0.0)
1	108 (43.0)	135 (42.2)
2	136 (54.2)	174 (54.4)
3	7 (2.8)	11 (3.4)
Total NAS[b]		
4	108 (43.0)	133 (41.6)
5	95 (37.8)	114 (35.6)
6	37 (14.7)	59 (18.4)
7	9 (3.6)	12 (3.8)
8	2 (0.8)	2 (0.6)

[a] Patients with type 2 diabetes.

[b] Central pathologist evaluation.

Abbreviations: AI, artificial intelligence; FAS, full analysis set; HbA$_{1c}$, glycated hemoglobin; NAS, nonalcoholic fatty liver disease activity score.

was achieved by 43.1% of patients receiving semaglutide 0.4 mg compared with 28.8% of patients receiving placebo ($p = 0.1913$) (Figure 3A). There was no significant difference between the semaglutide 0.1 or 0.2 mg groups and placebo for this endpoint. When measured by ML-derived categorical assessment, 32.3% of patients receiving semaglutide 0.4 mg and 20.3% receiving placebo achieved liver fibrosis improvement without NASH worsening ($p = 0.2036$) (Figure 3B); there were no significant differences between the semaglutide 0.1 or 0.2 mg treatment arms and placebo.

Ranked assessment of individual endpoints by central pathologists versus ML

Ranked assessment (either improvement, worsening, or no change) of changes in individual histological components from baseline to week 72 were assessed by central pathologists and ML-derived categorical assessment. Overall, results obtained via ML were consistent with those obtained via pathologist assessment. For both methods, a numerically higher proportion of patients achieved an improvement in the fibrosis stage with semaglutide 0.4 mg compared with placebo (Figure 4).

Both pathologist and ML-derived categorical assessments detected a higher proportion of patients with improvement in steatosis grades and lobular inflammation grade and a lower proportion with worsening of steatosis grades and lobular inflammation grade with all semaglutide doses compared with placebo (Supplemental Figure S1, http://links.lww.com/HEP/I172 and Supplemental Figure S2, http://links.lww.com/HEP/I172).

While both pathologist and ML-derived categorical assessments detected a higher proportion of patients with an improvement in hepatocyte ballooning grade for semaglutide versus placebo, a treatment benefit favoring semaglutide versus placebo for worsening in hepatocyte ballooning grade was detected by ML-derived categorical assessment, but not pathologist assessment (Supplemental Figure S3, http://links.lww.com/HEP/I172).

Change in histological component as measured by ML-derived continuous scores

ML-derived continuous scores (ranging from 0 to 4) detected significant treatment-induced responses in histological features including fibrosis, steatosis, lobular inflammation, portal inflammation, and hepatocellular ballooning.

ML-derived continuous fibrosis scores detected a statistically significant treatment-induced reduction in fibrosis in patients who received semaglutide 0.4 mg versus those who received placebo ($p = 0.0099$) (Figure 5).

The ML-derived mean fibrosis proportionate area (analogous to the collagen proportionate area) was numerically reduced in the semaglutide treatment arms in a dose-dependent manner and increased in the placebo treatment arm between baseline and week 72 but did not reach statistical significance (Figure 6).

Reductions in ML-derived mean steatosis proportionate area showed a dose–response relationship, with the greatest treatment difference in the semaglutide 0.4 mg

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

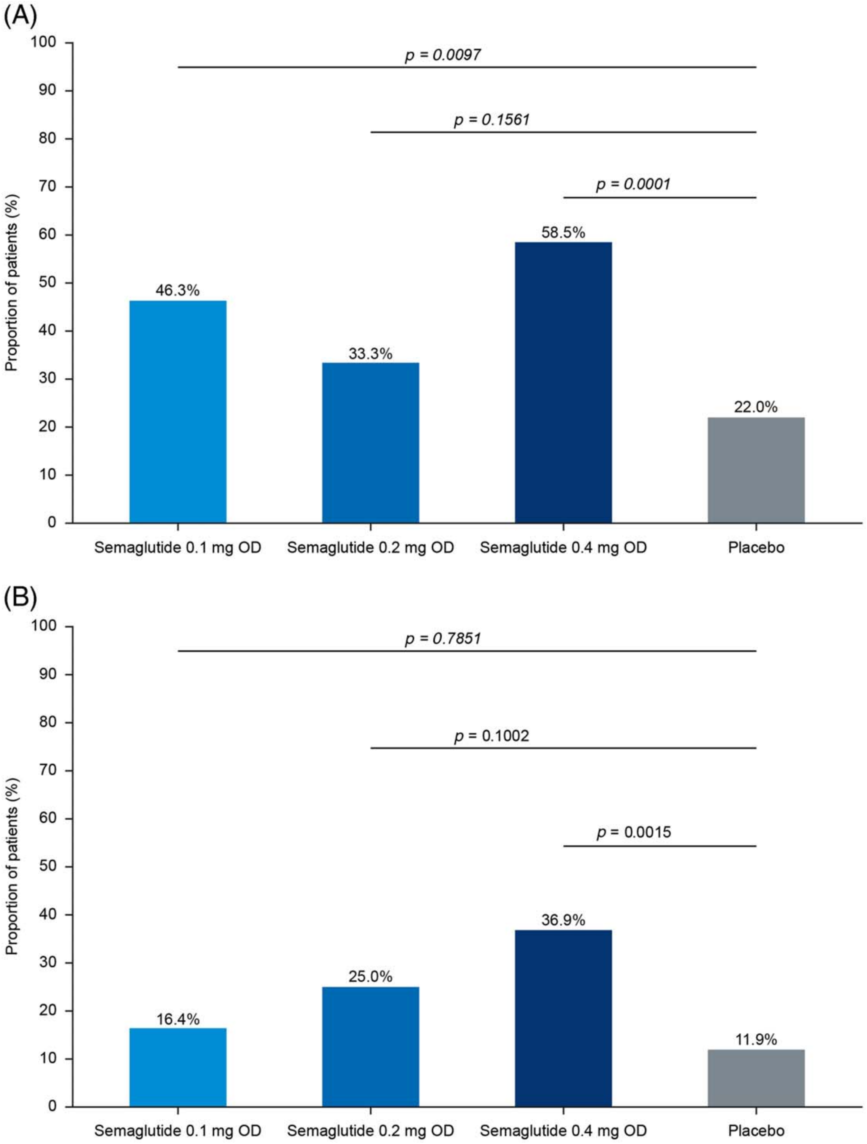


FIGURE 2 Proportions of patients in the FAS (AI) population with NASH resolution without fibrosis worsening (primary endpoint) at week 72 as assessed by (A) pathologist evaluation and (B) ML evaluation. *p* values are 2-sided and taken from a Cochran–Mantel–Haenszel test stratified by baseline diabetes status and baseline fibrosis stage. Patients with missing outcomes were imputed as nonresponders. Abbreviations: AI, artificial intelligence; FAS, full analysis set; ML, machine learning; OD, once daily.

group [−11.27% vs. placebo (−0.59%; $p < 0.0001$)] (Supplemental Figure S4, http://links.lww.com/HEP/I172).

Reductions in ML-derived mean lobular inflammation proportionate area were statistically significantly greater in all semaglutide treatment arms versus the placebo arm (Supplemental Figure S5, http://links.lww.com/HEP/I172).

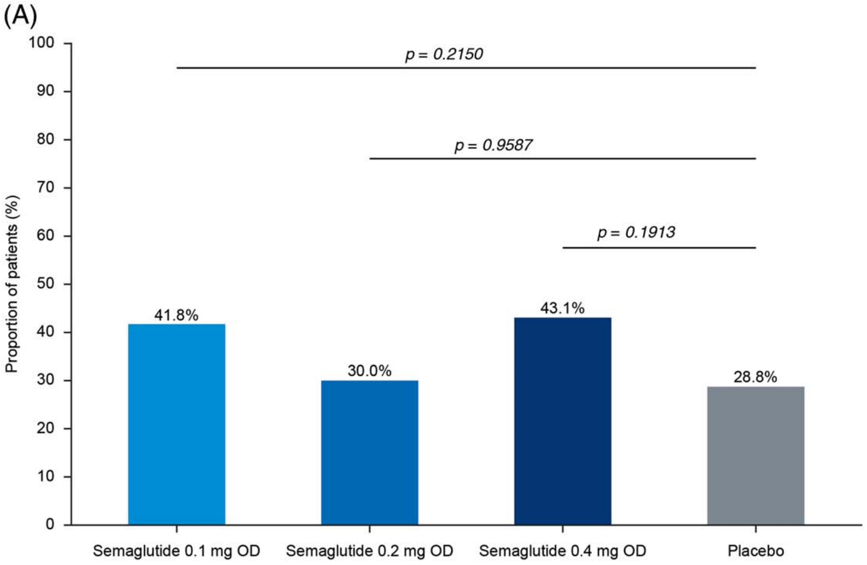
The ML-derived mean portal inflammation proportionate area was reduced in the semaglutide treatment arms and increased in the placebo arm between baseline and week 72 and was statistically significantly different between semaglutide 0.4 mg and placebo (Supplemental Figure S6, http://links.lww.com/HEP/I172).

Reductions in ML-derived mean hepatocyte ballooning proportionate area were statistically significantly greater in the semaglutide 0.2 and 0.4 mg treatment arms versus placebo, with the greatest treatment difference seen in the semaglutide 0.4 mg arm (−0.70 vs. −0.06 in the placebo arm) (Supplemental Figure S7, http://links.lww.com/HEP/I172).

Concordance between central pathologist and ML assessment

At baseline, the overall percentage agreement across the NASH CRN scores between the ML and consensus pathologist scores ranged from 49.8% to 71.8%. Weighted kappa statistics were 0.62 for steatosis and 0.28–0.36 for fibrosis, lobular inflammation, and hepatocyte ballooning (Table 2 and Supplemental Tables S1–S4, http://links.lww.com/HEP/I172). Where there

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024



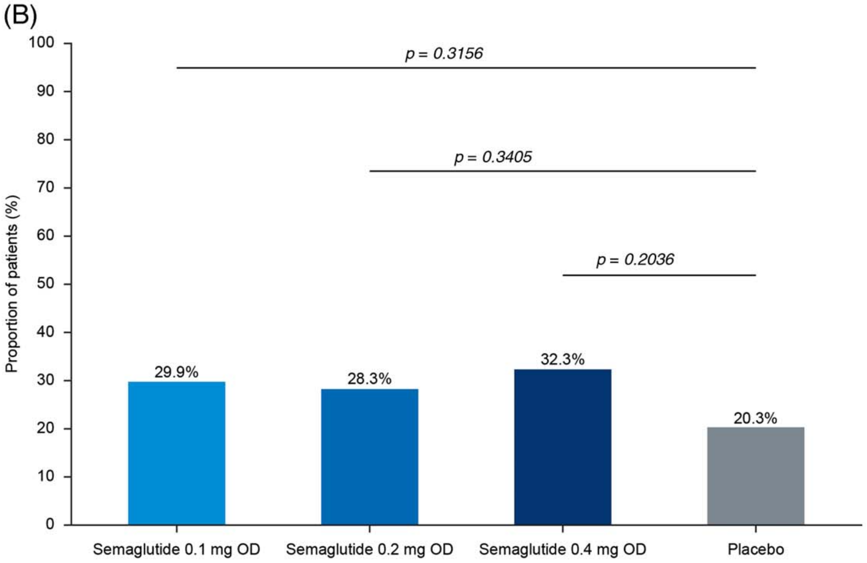


FIGURE 3 Proportions of patients in the FAS (AI) population with improvement in liver fibrosis of at least 1 stage with no worsening in steatohepatitis (confirmatory secondary endpoint) at week 72. (A) Pathologist evaluation and (B) ML evaluation. *p* values are 2-sided and taken from a Cochran–Mantel–Haenszel test stratified by baseline diabetes status and baseline fibrosis stage. Patients with missing outcomes were imputed as nonresponders. Abbreviations: AI, artificial intelligence; FAS, full analysis set; ML, machine learning; OD, once daily.

was disagreement between the ML and consensus pathology, the ML-derived categorical assessment tended to be higher for all histological measurements versus pathologist assessment.

The overall percentage agreement between the ML-derived and consensus pathologist scores for the change from baseline to week 72 across NASH CRN scores ranged from 49.0% to 68.0% (Supplemental Tables S1–S4, http://links.lww.com/HEP/I172). Weighted kappa statistics were 0.46 for steatosis and 0.22–0.39 for fibrosis, lobular inflammation, and hepatocyte ballooning (Table 2). For both fibrosis and steatosis, there was no apparent trend for either of the methods of assessment to consistently assess the change from baseline to week 72 (improvement, no change, worsening, missing) as discordant compared with the other method of assessment. For lobular inflammation, the ML model had more tendency to determine changes as “improvement” as opposed to “no change” when assessed by pathologists, while the ML model for hepatocyte ballooning tended to read a change as “no change” rather than “improvement” when assessed by pathologists.

Scatter plots were generated to evaluate the degree of correlation between ML-derived categorical versus continuous histological parameters in baseline slides (Supplemental Figure S8, http://links.lww.com/HEP/I172), and for evaluating the degree of correlation between the pathologist-derived categorical and ML-derived continuous histological parameters (Supplemental Figure S9, http://links.lww.com/HEP/I172). A positive linear trend was observed for all assessed parameters, indicating a moderate positive relationship

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

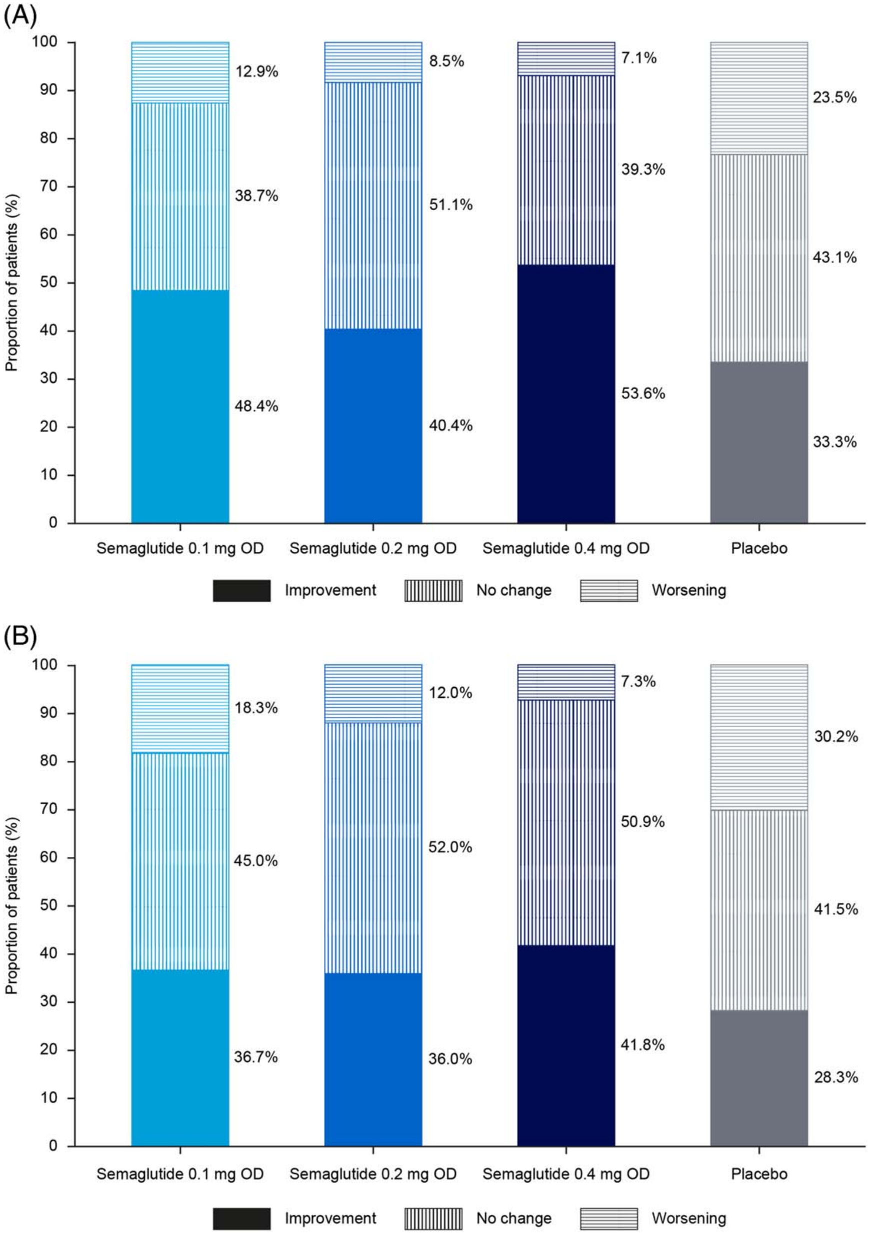


FIGURE 4 Change in NASH CRN fibrosis stage between baseline and week 72 in FAS (AI)* population. (A) Pathologist assessment and (B) ML assessment. *This figure does not include patients with missing biopsy data (n = 25 for pathologist assessment and n = 31 for ML assessment). Abbreviations: AI, artificial intelligence; CRN, Clinical Research Network; FAS, full analysis set; ML, machine learning; OD, once daily.

between baseline ML-derived categorical and continuous histological parameters, and between consensus pathologist-based categorical and ML-derived continuous histological parameters.

DISCUSSION

In a post hoc analysis of a double-blind, phase II trial of semaglutide versus placebo in patients with NASH,[14] we used digitized images of the biopsy slides to explore the ability of ML pathology models to detect histological changes and identify any potential benefit over conventional histological assessment, in the context of a clinical trial. Our main findings are 2-fold. Firstly, we were able to reproduce, using ML evaluations, the key histological findings observed through pathologist evaluation. In particular, we confirmed a significant treatment benefit of semaglutide 0.4 mg versus placebo for the primary endpoint (resolution of steatohepatitis without worsening of fibrosis), while benefits were also observed for endpoints related to fibrosis, steatosis, lobular inflammation, and hepatocyte ballooning when assessed categorically using both methods. Secondly, ML-derived continuous measurements detected significant treatment responses in fibrosis, steatosis, lobular inflammation, portal inflammation, and hepatocellular ballooning, identifying a significant

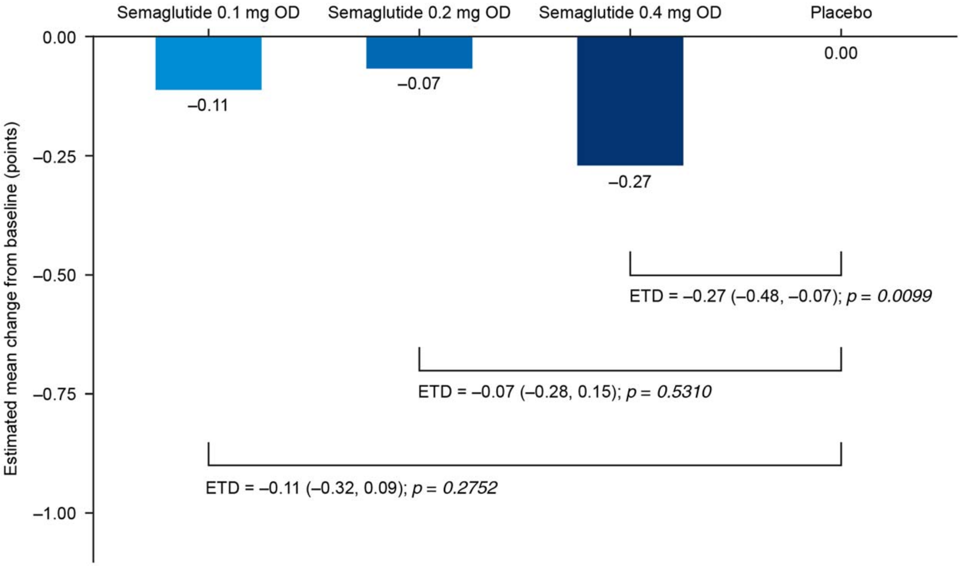


FIGURE 5 Change from baseline to week 72 in fibrosis assessed using ML-derived continuous fibrosis scoring in the FAS (AI) population. Mean changes from baseline were estimated from an ANCOVA with treatment, baseline diabetes status, baseline fibrosis stage, and diabetes-by-fibrosis interaction as factors, and baseline body weight and baseline value of the analyzed parameter as covariates. ETDs with 95% CIs and 2-sided *p* values are from the same analysis. Missing data were imputed from the observed data in the placebo group using the same ANCOVA model but without treatment as a factor. Abbreviations: AI, artificial intelligence; ETD, estimated treatment difference; FAS, full analysis set; ML, machine learning; OD, once daily.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

reduction in fibrosis for semaglutide 0.4 mg versus placebo that was not detected by categorical pathologists nor ML-derived categorical evaluation. These findings are important, as they could improve our ability to detect drug-induced histological changes while using automated methods that are less prone to contextual variability.

Pathologist assessment of liver biopsy samples is the current reference standard for the diagnosis and assessment of NASH severity in both clinical practice and clinical trials.[4] However, there are documented challenges regarding the reproducibility of pathologist assessment,[1,3,5–8] and there are still important limitations to overcome: for example, the application of a categorical evaluation system to continuous variables. With the advent of advanced computer vision technologies for digital pathology applications, several ML approaches for assessing disease severity from whole slide images of NASH biopsies have been developed.[10,11] While data from preliminary studies have shown the potential of ML-based approaches for detecting treatment responders and predicting disease progression and clinical outcomes in patients with NASH,[10,11,16] there is a clear need for further investigation.

In this post hoc analysis, the agreement between the ML-derived categorical assessment and the pathologist evaluations at baseline, assessed using weighted kappa statistics that accounted for the magnitude of disagreement between the methods, ranged from 0.28 to 0.62 across the 4 parameters (fibrosis, steatosis, lobular inflammation, and hepatocyte ballooning). The greatest level of agreement was recorded for steatosis (0.46–0.62), while agreement was below this range for the remaining parameters. This agreement was lower than has been observed between liver pathology experts, who had moderate-to-excellent agreement for the stage of fibrosis, steatosis, and hepatocyte ballooning, reflecting the consistency of assessment when performed by experienced pathologists.[1,3,15,18] The agreement in this analysis is also lower than the agreement measured for the same AI algorithms against pathologist consensus scoring in other clinical trial settings.[19] The higher variability between ML and pathologists in this study may be a result of multiple different factors, including the different analysis methods used by ML and pathologists to assess the histological features, and the quality of the slides impacting both ML and pathologist assessment. Generally, the ML-derived categorical assessment tended to assess the baseline fibrosis stage and the grades for steatosis, lobular inflammation, and hepatocyte ballooning higher than the pathologists' scores at baseline. The low concordance between pathologist and ML staging of fibrosis may be partially explained by the fact that reaching a pathologist consensus score can be more challenging when fibrosis is less severe compared with when it is more severe. Based on the study inclusion criteria of fibrosis stage 1–3 and steatosis, lobular inflammation, and hepatocyte ballooning grades ≥ 1,[14] a total of 14 enrolled patients would have failed screening by ML-derived categorical assessment of fibrosis stage, while the number of patients who would have failed screening based on ML-derived categorical

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

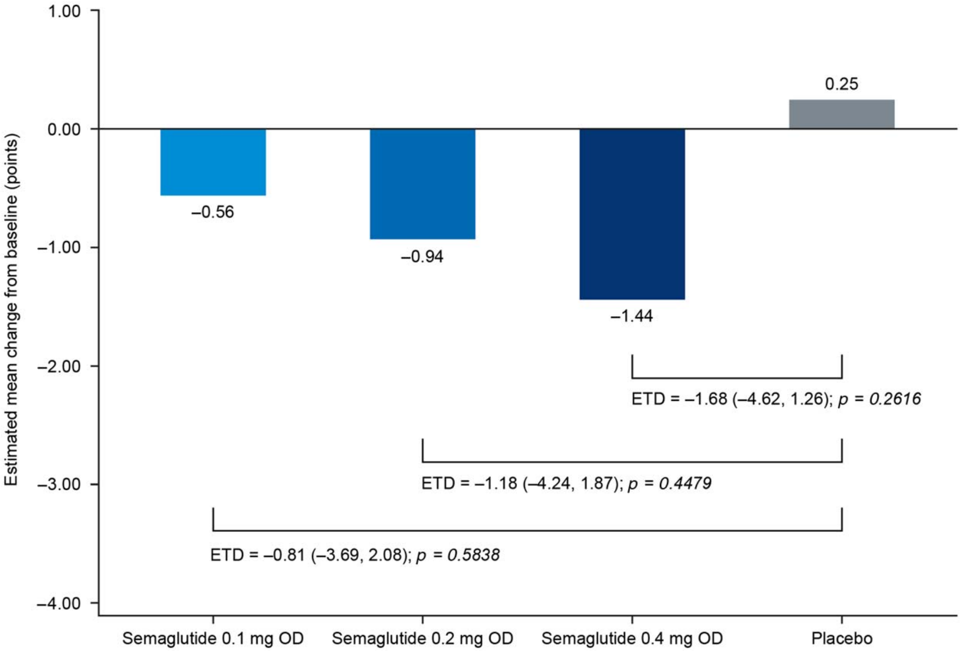


FIGURE 6 Change from baseline to week 72 in ML-based liver collagen proportionate area assessed by ML in the FAS (AI) population. Mean changes from baseline were estimated from an ANCOVA with treatment, baseline diabetes status, baseline fibrosis stage, and diabetes-by-fibrosis interaction as factors, and baseline body weight and baseline value of the analyzed parameter as covariates. ETDs with 95% CIs and 2-sided *p* values are from the same analysis. Missing data were imputed from the observed data in the placebo group using the same ANCOVA model but without treatment as a factor. Abbreviations: AI, artificial intelligence; ETD, estimated treatment difference; FAS, full analysis set; ML, machine learning; OD, once daily.

assessment for steatosis, lobular inflammation, and hepatocyte ballooning would have been 1, 10, and 10, respectively.

For the clinical trial endpoints measured over 72 weeks, results were broadly similar across both ML and pathologist evaluation; however, interesting differences were observed when comparing the 2 methods. For example, for the primary endpoint of NASH resolution without worsening of fibrosis, the percentage of patients achieving the endpoint was lower when assessed using ML evaluation versus pathologist evaluation across all treatment groups. The reason for this is unclear but may be partially explained by the potential for downgrading of NAS components when reaching a consensus score, or that the ML model has a stricter threshold for assigning a “0” for histological NASH features. The clinical significance of the observed change in ML-derived continuous fibrosis stage remains to be demonstrated. Studies have shown that fibrosis progression and regression in patients with NASH may take years to manifest in a way that can be captured by categorical evaluation.[20] Subsequently, it is possible that ML-derived continuous fibrosis staging may have the potential to identify improvement in the categorical stage earlier than changes that can be identified by pathologist review.[21] However, further investigation is required to analyze the reasons for differences between AI and pathologist review, address any evaluative bias, and examine how AI can best be implemented in a clinical setting.

The present findings demonstrate not only the critical importance of pathologists’ expertise and skill in interpreting and applying the NASH CRN guidelines but also the benefit of integrating ML approaches into existing NASH clinical trial biopsy assessment workflows. AI-assisted pathologist evaluation in NASH clinical trials has the potential to restrict the impact of intra-rater and inter-rater variability using conventional scoring schema while leveraging the medical knowledge and expertise of NASH pathologists to ensure that final decisions regarding disease severity are accurate and appropriately contextualized. Several AI-assisted pathology platforms have been proposed, ranging from showing central pathologists ML-derived heatmaps overlaying whole slide images that highlight (or “spotlight”) features that are relevant to staging disease, to providing ML-derived scores that central pathologists would either accept or reject.[10,12,22] Others have identified patterns of regression or progression of fibrous septa.[1,23] Further studies should be conducted to specifically characterize the ways in which AI tools can be used to support pathologists in NASH clinical trials, with a common goal of ensuring that physicians, regulatory agencies, and drug developers are confident that the eventual drugs that are prescribed to patients with NASH are safe and effective.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

TABLE 2 Agreement between consensus pathologist and ML evaluation of NASH CRN score at baseline and the change from baseline to week 72 (improvement, worsening, or no change)

	Baseline			Change from baseline to week 72[a]		
Score	N_{obs}	ML/pathologist agreement, N (%)	Weighted kappa[b]	N_{obs}	ML/pathologist agreement, N (%)	Weighted kappa[b]
Fibrosis						
Overall	249	124 (49.8)	0.3405	208	102 (49.0)	0.2190
Semaglutide 0.1 mg	67	34 (50.7)	0.3480	60	25 (41.7)	0.1534
Semaglutide 0.2 mg	60	33 (55.0)	0.3773	46	25 (54.3)	0.2148
Semaglutide 0.4 mg	63	29 (46.0)	0.3021	52	30 (57.7)	0.2877
Placebo	59	28 (47.5)	0.3241	50	22 (44.0)	0.1601
Steatosis						
Overall	248	178 (71.8)	0.6216	206	140 (68.0)	0.4555
Semaglutide 0.1 mg	65	52 (80.0)	0.7399	56	40 (71.4)	0.5315
Semaglutide 0.2 mg	60	40 (66.7)	0.5161	49	35 (71.4)	0.4029
Semaglutide 0.4 mg	65	47 (72.3)	0.6624	53	39 (73.6)	0.4685
Placebo	58	39 (67.2)	0.5269	48	26 (54.2)	0.2405
Lobular inflammation						
Overall	248	156 (62.9)	0.3585	207	112 (54.1)	0.2864
Semaglutide 0.1 mg	65	40 (61.5)	0.3424	56	31 (55.4)	0.2814
Semaglutide 0.2 mg	60	40 (66.7)	0.4045	49	25 (51.0)	0.2043
Semaglutide 0.4 mg	65	44 (67.7)	0.4740	53	32 (60.4)	0.2967
Placebo	58	32 (55.2)	0.2028	49	24 (49.0)	0.2176
Hepatocyte ballooning						
Overall	248	145 (58.5)	0.2792	206	133 (64.6)	0.3881
Semaglutide 0.1 mg	65	35 (53.8)	0.2602	56	36 (64.3)	0.3860
Semaglutide 0.2 mg	60	38 (63.3)	0.3373	49	33 (67.3)	0.4171
Semaglutide 0.4 mg	65	38 (58.5)	0.2490	53	38 (71.7)	0.1787
Placebo	58	34 (58.6)	0.2780	48	26 (54.2)	0.2718

[a] Definition of change (improvement, worsening, or no change): improvement defined as at least 1 stage/grade decrease from baseline to week 72 in corresponding histological parameter; worsening defined as at least 1 stage/grade increase from baseline to week 72 in corresponding histological parameter; no change defined as no change in stage/grade from baseline to week 72 in corresponding histological parameter.

[b] Calculation of kappa statistics was not based on missing data.

Abbreviations: CRN, Clinical Research Network; ML, machine learning; N_{obs}, number of observations.

This analysis has some limitations that should be considered when interpreting the results. Digital pathology images were not available for all patients; therefore, analyses in this paper were only performed on a subset of patients (251 patients) from the original trial. In terms of enrollment, the fact that the population was enrolled based on pathologist assessment means that the endpoint assessment may have been different than if they had been enrolled using ML. Finally, only patients with a fibrosis stage of 1–3 were included in the current post hoc analysis; therefore, further studies are needed to evaluate the application of the ML pathology model to stages F0 and F4.

In conclusion, in this post hoc analysis of data from a phase II study, we found a significant treatment benefit of semaglutide 0.4 mg versus placebo in patients with NASH and fibrosis, regardless of whether the results were based on pathologist or ML evaluation of liver biopsies. Overall, the results for the primary and secondary endpoints were similar across the 2 biopsy assessment methods; however, ML-derived continuous assessment detected a statistically significant reduction in fibrosis for semaglutide 0.4 mg versus placebo that could not be detected by categorical pathologists nor ML evaluation. These results demonstrate that ML evaluation can provide additional value in the interpretation of histological results compared with using only change in categorical NASH CRN scores, with important implications for aspects such as natural history, predicting treatment response, or monitoring disease progression in clinical trials. Further studies investigating the similarities, differences, and complementarity between pathologist and AI evaluations of liver biopsies, and the long-term correlations between AI evaluations and disease progression in patients with NASH are required.

DATA AVAILABILITY

Data sets will be shared with bona fide researchers who submit a research proposal approved by an independent

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

review board after research completion and approval of the product and product use in the EU and the United States. Information about data access request proposals can be found at novonordisk-trials.com.

AUTHOR CONTRIBUTIONS

Concept and design: Anne-Sophie Sejling. Acquisition, analysis, or interpretation of data: All authors. Drafting of the manuscript: All authors. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: Niels Krarup. Final approval of the version to be published: All authors. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved: All authors.

CONFLICTS OF INTEREST

Vlad Ratziu consults and received grants from Intercept. He consults for Enyo, Madrigal, NorthSea, Novo Nordisk, Poxel, and Sagimet. He received grants from Gilead. Sven Francque consults, is on the speakers' bureau, and received grants from GENFIT, Gilead, Inventiva, Janssen, and Merck. He consults and is on the speakers' bureau for AbbVie, Allergan, Bayer, Eisai, Intercept, Novo Nordisk, and Promethera. He consults and received grants from Astellas, Bristol Myers Squibb, and Roche. He consults for Actelion, Aelin, Aligos, AstraZeneca, Boehringer Ingelheim, Coherus, CSL Behring, Echosens, Enyo, Galapagos, Galmed, Genentech, Julius Clinical, Madrigal, Medimmune, NGM Bio, and Novartis. He received grants from Falk, Glympse Bio, and Pfizer. He is employed by the Research Foundation Flanders. Cynthia A. Behling consults for Hologic. She is on the speakers' bureau for Alimentiv and Pfizer. She is employed by Pacific Rim Pathology Laboratory. She has other interests with AcelaBio and Novo Nordisk. Vanja Cejvanovic is employed by and owns stock in Novo Nordisk. Helena Cortez-Pinto consults and is on the speakers' bureau for Eisai, Intercept, Orphalan, and Roche. She consults for Novo Nordisk. Janani S. Iyer is employed by PathAI. Niels Krarup is employed by Novo Nordisk. Quang Le is employed by and owns stock in PathAI. Anne-Sophie Sejling is employed by and owns stock in Novo Nordisk. Dina Tiniakos consults for Alimentiv, Allergan, Cirius, Clinnovate, CymaBay, ICON, Intercept, Inventiva, Ionis, Madrigal, Merck, and Verily Life Sciences. She received grants from HistoIndex. Stephen A. Harrison consults, advises, received grants, and owns stock in Akero, Cirius, Galectin, GENFIT, Hepion, Metacrine, NGM Bio, and NorthSea. He consults, advises, and received grants from Axcella, CiVi Biopharma, CymaBay, Enyo, Galmed, Gilead, HighTide, Intercept, Madrigal, Novartis, Novo Nordisk, Pfizer, Sagimet, and Viking. He consults, advises, and owns stock in ChronWell, Hepta Bio, HistoIndex, and Sonic Incytes. He consults and advises 89bio, Agomab, Alentis, Aligos, Alimentiv, Altimmune, Arrowhead, Blade, Bluejay, Boston Pharmaceuticals, Boxer Capital, BVF Partners, Can-Fite BioPharma, Chronic Liver Disease Foundation (CLDF), CohBar, Corcept, Echosens, Fibronostics, Foresite Labs (MetreaBiosciences), Fortress, Galecto, Gelesis, Glaxo Smith Kline, GNS Healthcare, GRI Bio, Hepagene, Humana, Indalo, Inipharm, Innovate, Ionis, Kowa, Medpace, Merck, MGGM, NeuroBo, Nutrasource, PathAI, Perspectum, Piper Sandler, Poxel, Prometic, Ridgeline, Silverback, Terns, Theratechnologies, and Zafgen. He received grants from Bristol Myers Squibb, Conatus, Genentech, Gilead, Immuron, and Second Genome.

ACKNOWLEDGMENTS

Medical writing support was provided by Andy Bond and Catherine Stanton of Titan, OPEN Health Communications, and funded by Novo Nordisk A/S, in accordance with Good Publication Practice 3 (GPP3) guidelines (www.ismpp.org/gpp3).

FUNDING INFORMATION

This study was funded by Novo Nordisk A/S, Søborg, Denmark.

ORCID

Vlad Ratziu https://orcid.org/0000–0002–6865–3791
Helena Cortez-Pinto https://orcid.org/0000–0002–8537–8744
Dina Tiniakos https://orcid.org/0000–0003–4657–7780
Stephen A. Harrison https://orcid.org/0000–0001–8285–2204

REFERENCES

1. Bedossa P. 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. 2014;60:565–75.
2. Kleiner DE, Bedossa P. Liver histology and clinical trials for nonalcoholic steatohepatitis—Perspectives from 2 pathologists. Gastroenterology. 2015;149:1305–8.
3. Gawrieh S, Knoedler DM, Saeian K, Wallace JR, Komorowski RA. Effects of interventions on intra- and interobserver agreement on interpretation of nonalcoholic fatty liver disease histology. Ann Diagn Pathol. 2011;15:19–24.
4. US Department of Health and Human Services, US Food and Drug Administration, Center for Drug Evaluation and Research. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: Developing drugs for treatment (draft guidance for industry). 2018. Accessed November 10, 2021. https://www.fda.gov/media/119044/download
5. Davison BA, Harrison SA, Cotter G, Alkhouri N, Sanyal A, Edwards C, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol. 2020;73:1322–32.
6. Brunt EM, Clouston AD, Goodman Z, Guy C, Kleiner DE, Lackner C, et al. Complexity of ballooned hepatocyte feature recognition: Defining a training atlas for artificial intelligence-based imaging in NAFLD. J Hepatol. 2022;76:1030–41.

Downloaded from http://journals.lww.com/hep by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 02/06/2024

7. Jung ES, Lee K, Yu E, Kang YK, Cho MY, Kim JM, et al. Interobserver agreement on pathologic features of liver biopsy tissue in patients with nonalcoholic fatty liver disease. J Pathol Transl Med. 2016;50:190–6.
8. Robinson M, James J, Thomas G, West N, Jones L, Lee J, et al. Quality assurance guidance for scoring and reporting for pathologists and laboratories undertaking clinical trial work. J Pathol Clin Res. 2019;5:91–9.
9. Liu F, Goh GBB, Tiniakos D, Wee A, Leow WQ, Zhao JM, et al. qFIBS: An automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis. Hepatology. 2020;71:1953–66.
10. Taylor-Weiner A, Pokkalla H, Han L, Jia C, Huss R, Chung C, et al. A machine learning approach enables quantitative measurement of liver histology and disease monitoring in NASH. Hepatology. 2021;74:133–47.
11. Bosch J, Chung C, Carrasco-Zevallos OM, Harrison SA, Abdelmalek MF, Shiffman ML, et al. A machine learning approach to liver histological evaluation predicts clinically significant portal hypertension in NASH cirrhosis. Hepatology. 2021;74:3146–60.
12. Prakash A, Elliott H, Montalto M, Beck A, Resnick M, Wapinski I, et al. A deep learning approach to analysis of MRCP images predicts clinical events and progression to cirrhosis in patients with primary sclerosing cholangitis. Poster presentation at the European Association for the Study of the Liver International Liver Congress (EASL ILC 2022), London, UK. 2022. Accessed November 29, 2022. https://easl.eu/wp-content/uploads/2021/06/EASL_2021_Version-3-new.pdf
13. Shevell DE, Brown E, Du S, et al. Comparison of manual vs machine learning approaches to liver biopsy scoring for NASH and fibrosis: A post hoc analysis of the FALCON 1 study. Hepatology. 2021;74:1415A.
14. Newsome PN, Buchholtz K, Cusi K, Linder M, Okanoue T, Ratziu V, et al. A placebo-controlled trial of subcutaneous semaglutide in nonalcoholic steatohepatitis. N Engl J Med. 2021;384:1113–24.
15. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
16. Wang JK, Pouryahya M, Leidal K, Pokkalla H, Juyal D, Shanis Z, et al. Liver biopsy graph neural networks for automated histologic scoring using the NASH CRN system. Poster presentation at The International Liver Congress (EASL 2021), Virtual. 2021. Accessed November 28, 2022. https://easl.eu/wp-content/uploads/2021/06/EASL_2021_Version-3-new.pdf
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
18. Pournik O, Alavian SM, Ghalichi L, Seifizarei B, Mehrnoush L, Aslani A, et al. Inter-observer and intra-observer agreement in pathological evaluation of non-alcoholic fatty liver disease suspected liver biopsies. Hepat Mon. 2014;14:e15167.
19. Harrison SA, Iyer JS, Bedossa P, Guy C, Biddle-Snead C, Hoffman S, et al. Retrospective AI-based measurement of NASH histology (AIM-NASH) analysis of biopsies from phase 2 study of resmetirom confirms significant treatment-induced changes in histologic features of non-alcoholic steatohepatitis. Poster presentation at the European Association for the Study of the Liver International Liver Congress (EASL ILC 2022), London, UK. 2022. Accessed September 27, 2022. https://www.postersessiononline.eu/173580348_eu/congresos/NAFLDsummit2022/aula/-P03_1_NAFLDsummit2022.pdf
20. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: A systematic review and meta-analysis of paired-biopsy studies. Clin Gastroenterol Hepatol. 2015;13:643–54.e1–9; quiz e39–e40.
21. Juyal D, Shukla C, Pokkalla H, Taylor A, Zevallos O, Resnick M, et al. Machine learning identifies histologic features associated with regression of cirrhosis in treatment for chronic hepatitis B. Poster presentation at the European Association for the Study of the Liver International Liver Congress (EASL ILC 2020), Virtual. 2020. Accessed November 28, 2022. https://easl.eu/wp-content/uploads/2020/12/digital-ilc-2020-abstract.pdf
22. Ratziu V, Francque S, Sanyal A. Breakthroughs in therapies for NASH and remaining challenges. J Hepatol. 2022;76:1263–78.
23. Naoumov NV, Brees D, Loeffler J, Chng E, Ren Y, Lopez P, et al. Digital pathology with artificial intelligence analyses provides greater insights into treatment-induced fibrosis regression in NASH. J Hepatol. 2022;77:1399–409.

How to cite this article: Ratziu V, Francque S, Behling CA, Cejvanovic V, Cortez-Pinto H, Iyer JS, et al. Artificial intelligence scoring of liver biopsies in a phase II trial of semaglutide in nonalcoholic steatohepatitis. Hepatology. 2024;■■:■■–■■. https://doi.org/10.1097/HEP.0000000000000723