Article type : Original

A Machine Learning Approach to Liver Histological Evaluation Predicts Clinically Significant Portal Hypertension in NASH Cirrhosis

Jaime Bosch,[1*] Chuhan Chung,[2*] Oscar M. Carrasco-Zevallos,[3] Stephen A. Harrison,[4] Manal F. Abdelmalek,[5] Mitchell L. Shiffman,[6] Don C. Rockey,[7] Zahil Shanis,[3] Dinkar Juyal,[3] Harsha Pokkalla,[3] Quang Huy Le,[3] Murray Resnick,[3] Michael Montalto,[3] Andrew H. Beck,[3] Ilan Wapinski,[3] Ling Han,[2] Catherine Jia,[2] Zachary Goodman,[8] Nezam Afdhal,[9] Robert P. Myers,[2] Arun J. Sanyal[10]

[1]Department of Biomedical Research, University of Bern, Switzerland, and University of Barcelona-IDIBAPS, Spain; [2]Gilead Sciences, Inc., Foster City, CA; [3]PathAI, Inc., Boston, MA; [4]Pinnacle Clinical Research, San Antonio, TX; [5]Duke University, Durham, NC; [6]Bon Secours Liver Institute of Richmond, VA; [7]Medical University of South Carolina, Charleston; [8]Inova Fairfax Hospital, Falls Church, VA; [9]Beth Israel Deaconess Medical Center and Harvard Medical School, Boston; [10]Virginia Commonwealth University, Richmond, VA

* Joint first authors.

Address reprint requests to: Dr. Jaime Bosch, M.D., Ph.D., F.R.C.P., F.A.A.S.L.D., Maurice Müller Haus (F 805), Inselspital, Murtenstrasse 35, 3010 Bern, Switzerland.
E-mail: jaime.bosch@dbmr.unibe.ch

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/HEP.32087

Funding: This analysis was supported by Gilead Sciences, Inc.

Conflicts of Interest: Jaime Bosch is on the advisory committees or consults for Actelion, Gilead, , Biovie, , Bristol-Myers Squibb, Blade, Lipocine and Surrozen; and receives grant and research support from Gilead. Chuhan Chung is an employee of and holds stock in Gilead Sciences. Oscar M. Carrasco-Zevallos is an employee of PathAI. Stephen A. Harrison consults, advises, and received grants from Axcella, Cirius, CymaBay, Galectin, Galmed, Genfit, Gilead, HighTide, Intercept, Madrigal, NGM, Novartis, Novo Nordisk, and Pfizer. He consults and advises 3V Bio, Akero, Albireo, Blade, Bristol-Myers Squibb, Civi, CLDF, Contravir, Consynance, Corcept, Echosens, HistoIndex, Innovate, IQVIA, Intercept, Lipocine, Medpace, Metacrine, Perspectum, PPD, Prometheus, Prometic, and Terns. He is on the speakers' bureau for Alexion. He received grants from Conatus, Immuron, Second Genome, and Tobira/Allergan.
Manal F. Abdelmalek receives grants from Gilead Sciences. Mitchell L. Shiffman advises, is on the speakers' bureau, and received grants from AbbVie, Bristol-Myers Squibb, Gilead, Intercept, and Merck; he advises and is on the speakers' bureau for Bayer, Dova, Salix, and Shionogi; he is on the speakers' bureau and received grants from Enanta; he consults for Optum Rx; he is on the speakers' bureau for Easai and Daiichi Sankyo; and he received grants from Conatus, CymaBay, Exalenz, Galectin, Genfit, Genkyotex, Immuron, NGM, Novartis, and Shire. Don C. Rockey receives grant and research support from Gilead, Actelion, Conatus, Galectin, Shire, and Cumberland. Zahil Shanis, Dinkar Juyal, Harsha Pokkalla, Quang Huy Le, Murray Resnick, Michael Montalto are employees of, and Andrew H. Beck and Ilan Wapinski are co-founders of PathAI. Ling Han and Catherine Jia are employees of and hold stock in Gilead Sciences. Zachary Goodman has received grants from Gilead, Intercept, Novartis, Bristol-Myers Squibb, and Allergan. Nezam Afdhal consults and advises Gilead, Merck, Echosens, Ligand, and Shionogi. He advises and owns stock in Trio and owns stock in Spring Bank. Robert P. Myers is an employee of and holds stock in Gilead Sciences. Arun J. Sanyal consults and received grants from Malinckrodt, Novartis, and Conatus. He is employed by Sanyal Bio. He consults for Pfizer, Nitto Denko, Hemoshear, Novo Nordisk, Terns, and Lilly. He received grants from Galectin, Malinckrodt, Gilead, Bristol-Myers Squibb, Salix, Merck, and Sequana. He received royalties from Elsevier and Uptodate. He owns stock in Akarna, Genfit, Indalo, Dureet, and Tiziana.

Author Contributions:

Study design: Jaime Bosch, Chuhan Chung, Robert P. Myers

Machine learning analysis: Oscar M. Carrasco-Zevallos, Zahil Shanis, Dinkar Juyal, Harsha Pokkalla, Quang Huy Le, Murray Resnick, Michael Montalto, Andrew H. Beck, Ilan Wapinski

Study investigators: Stephen A. Harrison, Manal F. Abdelmalek, Mitchell L. Shiffman, Don C. Rockey, Nezam Afdhal, Arun J. Sanyal

Centralized HVPG review: Jaime Bosch

Centralized pathology review: Zachary Goodman

Statistical analysis: Ling Han, Catherine Jia

Manuscript drafting: Jaime Bosch, Chuhan Chung, Robert P. Myers

All authors critically reviewed and approved the manuscript for submission.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under receiver operating characteristic curve; CNN, convolutional neural network; CRN, Clinical Research Network; CSPH, clinically significant portal hypertension; ELF, Enhanced Liver Fibrosis; FHVP, free hepatic vein pressure; FIB-4, fibrosis-4 index; GGT, gamma-glutamyl transferase; HbA1c, hemoglobin A1c; HR, hazard ratio; HVPG, hepatic venous pressure gradient; H&E, hematoxylin and eosin; INR, international normalized ratio of prothrombin time; MELD, Model for End-Stage Liver Disease; ML, machine learning; NAFLD, nonalcoholic fatty liver disease; NAS, NAFLD Activity Score; NASH, nonalcoholic steatohepatitis; PSR, picrosirius red; WHVP, wedged hepatic venous pressure

ABSTRACT

Background: The hepatic venous pressure gradient (HVPG) is the standard for estimating portal pressure but requires expertise for interpretation. We hypothesized that HVPG could be extrapolated from liver histology using a machine learning (ML) algorithm.

Methods: NASH patients with compensated cirrhosis from a phase 2b trial were included. HVPG and biopsies from baseline and weeks 48 and 96 were reviewed centrally, and biopsies evaluated with a convolutional neural network (PathAI; Boston, MA). Using trichrome-stained biopsies in the training set (n=130), an ML model was developed to recognize fibrosis patterns associated with HVPG and the resultant ML HVPG score was validated in a held-out test set (n=88). Associations between the ML HVPG score with measured HVPG and liver-related events, and performance of the ML HVPG score for clinically significant portal hypertension (CSPH, HVPG ≥10 mm Hg) were determined.

Results: The ML HVPG score was more strongly correlated with HVPG than hepatic collagen by morphometry (ρ=0.47 vs ρ=0.28; $p<0.001$). The ML HVPG score differentiated patients with normal (0-5 mmHg) and elevated HVPG (5.5-9.5 mmHg), and CSPH (median: 1.51 vs 1.93 vs 2.60; all $p<0.05$). The AUROCs (95%CI) of the ML HVPG score for CSPH were 0.85 (0.80,0.90) and 0.76 (0.68,85) in the training and test sets, respectively. Discrimination of the ML HVPG score for CSPH improved with addition of a ML parameter for nodularity, ELF, platelets, AST, and bilirubin (AUROC in test set: 0.85;95%CI 0.78,0.92). While baseline ML HVPG score was not prognostic, changes were predictive of clinical events (HR 2.13; 95%CI 1.26,3.59) and associated with hemodynamic response and fibrosis improvement.

Conclusions: A ML-model based on trichrome-stained liver biopsy slides can predict CSPH in NASH patients with cirrhosis.

Keywords: cirrhosis, portal hypertension, HVPG

INTRODUCTION

Clinical events of hepatic decompensation in patients with cirrhosis result mainly from portal hypertension. Measurement of the hepatic venous pressure gradient (HVPG) by hepatic vein catheterization is the reference standard for assessing the severity of sinusoidal portal hypertension, and a wealth of data has confirmed its utility for predicting clinical events in patients with cirrhosis of diverse etiologies.(1-3) Specifically, an HVPG of 10 mm Hg or greater is considered the threshold for clinically significant portal hypertension (CSPH), and predicts the development of esophageal varices, hepatic decompensation, and liver-related mortality. The therapeutic corollary to these findings is the importance of reducing HVPG to lessen the probability of adverse clinical outcomes.(4, 5)

Despite the utility of HVPG measurement for risk stratification, impediments to its proper use and interpretation are well described.(6, 7) The procedure is invasive and costly, and due to the need for interventional radiographic expertise and facilities, may be inadequately performed outside of experienced, academic centers. Moreover, multiple technical aspects, as well as deficiencies in the assessment of waveform tracings, may limit its application in everyday practice.(7) Finally, recent studies point to far greater variability in HVPG measurement in patients with cirrhosis due to nonalcoholic steatohepatitis (NASH) versus other etiologies such as alcohol or viral hepatitis.(8)

Given these limitations to HVPG measurement, alternative approaches to the evaluation of portal hypertension have been evaluated, including serum markers of fibrosis, elastographic measures of liver and spleen stiffness, imaging methods, and functional tests (e.g., cholate clearance).(9-12) Although correlations between these measures and HVPG have been reported, data in patients with cirrhosis due to NASH and clinical utilization for diagnosis of CSPH are limited.

An alternative approach is to utilize findings from liver biopsies to estimate the severity of portal hypertension. For example, Garcia-Tsao and colleagues found that the thickness of fibrotic septae and small nodule size were independently predictive of CSPH.(13, 14) These observations are consistent with the marked alterations in structural components found in cirrhosis, which is a major contributor to portal hypertension. However, a limitation of traditional histologic assessment is its subjectivity, including the grading of septal thickness using terms such as ‘thin’, ‘medium’, or ‘thick.’ Adoption of digital image analysis to quantify fibrosis and nodule size may circumvent this issue. Indeed, multiple

studies have shown associations between HVPG and collagen proportionate area quantified by morphometry.(14-16) More recently, digital image analysis based on machine learning (ML) algorithms has shown promise for improving the reliability of histological evaluation in NASH.(17-19) This approach has also enabled the identification of novel histologic features associated with clinical disease progression, including portal inflammation and indicators of fibrosis severity based on fibrosis patterns that are distinct from collagen proportionate area.(20)

In the current study, we hypothesized that HVPG could be extrapolated by a ML algorithm that used deep convolutional neural networks (CNN) for quantitative characterization of liver biopsy tissue. This ML approach was applied in a retrospective analysis of histological samples from a clinical trial in patients with NASH cirrhosis where serial measurements of HVPG were centrally reviewed and clinical outcomes prospectively recorded.(21) Our data suggest that a ML-based approach to assessment of the structural features that comprise the major component of portal hypertension may represent a practical solution to the assessment of HVPG in NASH cirrhosis.

METHODS

Study Population

This analysis used data from a large, randomized, placebo-controlled phase 2b study of simtuzumab in patients with compensated cirrhosis due to NASH (ClinicalTrials.gov NCT01672879). The primary results of the study are reported elsewhere, where the methods are fully described.(21) Briefly, the study enrolled 258 patients with histologically confirmed NASH and compensated cirrhosis (modified Ishak fibrosis stage 5-6). Patients were randomized in a 1:1:1 ratio to receive 200 mg of simtuzumab, 700 mg of simtuzumab, or placebo by intravenous infusion every 2 weeks for a planned treatment duration of 240 weeks. Randomization was stratified by the presence or absence of diabetes and CSPH. A pre-planned interim analysis conducted after all patients had completed at least 96 weeks of treatment found that there were no meaningful differences between either of the simtuzumab groups or the placebo group in any efficacy endpoint, including hepatic fibrosis and HVPG. Therefore, treatment groups were combined for the current analysis.

Patients in the study were deemed to have a clinical event if they had evidence of hepatic decompensation (ascites, newly diagnosed varices in a patient without prior varices, esophageal variceal bleeding, hepatic encephalopathy, or ≥2-point increase in Child–Pugh–Turcotte score), liver transplantation, qualification for transplantation (Model for End-stage Liver Disease [MELD] ≥15), or death, confirmed by an independent Hepatic Events Adjudication Committee.

In total, 218 subjects screened or enrolled in the study with available HVPG measurements, liver histology, and appropriate consent were included in the current analysis.

HVPG Measurement, Liver Histology, and Serum Markers

HVPG measurements were performed according to a standardized protocol during the screening period and at weeks 48 and 96 of the study. As previously described,(21) measurements of wedged (occluded) hepatic venous pressure (WHVP) and free hepatic vein pressure (FHVP) were made in triplicate. Permanent tracings for each measurement were obtained and the mean value was recorded for that visit. HVPG was calculated as the difference between the mean WHVP and mean FHVP. All tracings were evaluated centrally by a single reader (JB); the intra-class correlation coefficients (ICC) at each time point were 0.97 or greater.

Core liver biopsies obtained at screening and at weeks 48 and 96 and stained with hematoxylin and eosin (H&E) and Masson's trichrome, were read by a single expert hepatopathologist. Histologic assessments included the adequacy of the biopsy specimen, confirmation of the diagnosis, NAFLD Activity Score (NAS) grading, and fibrosis staged according to a modified Ishak classification. Hepatic collagen content was digitally quantified using picrosirius red (PSR)-stained biopsies which were scanned at 20x magnification using an Aperio ScanScope XT (Leica Biosystems, Vista, CA); results were expressed as the percentage of the entire biopsy section that was positive for PSR.(21)

Finally, serum markers including routine liver biochemistry and the noninvasive markers of fibrosis, APRI, FIB-4, and the Enhanced Liver Fibrosis (ELF) score (Siemens Healthcare, Erlangen, Germany), were measured.

Training and Testing a Machine Learning Model for Assessment of HVPG

Study subjects were randomly separated in a 3:2 ratio into training (n=130) and test (n=88) sets which were used for model training and performance evaluation, respectively. Using the PathAI research platform (PathAI; Boston, MA), digitized images of Masson's trichrome–stained biopsies (training set: n=320 slides; n=276 HVPG measurements) were used to develop an "end-to-end" (E2E), CNN model to recognize fibrosis patterns associated with HVPG measurements in six clinically relevant categories (0–5 mm Hg [normal], 5.5–9.5 mm Hg [elevated], and 10–11.5, 12–15.5, 16–19.5, and ≥20 mm Hg [all indicative of CSPH]). These ascending gradations of CSPH have been associated with increased risk of liver-related clinical events including mortality.(22) These HVPG categories (labelled from 0 to 5 in ascending order of HVPG) were the trichrome image-associated labels used for model training. After label assignment, we utilized an HVPG model training process as described previously for training E2E models to predict NASH CRN or Ishak fibrosis stage.(23) Briefly, pathologist-derived annotations of whole slide images of trichrome-stained biopsies were used to generate training sets of 531x531 pixel image patches on the order of 500,000 samples. These patches were used to train a deep CNN with stochastic minibatch gradient descent using the ADAM optimizer to produce pixel-level predictions of fibrosis, bile duct, blood vessel, and lumen. Models are comprised of 8-12 blocks of compound layers with a topology inspired by residual networks and inception networks with a softmax loss. Model training was monitored and hyperparameters were adjusted based on the performance of the model on pathologist annotations from the held-out test set, until convergence was achieved. HVPG model predictions were restricted to regions identified as fibrosis with the previously developed fibrosis segmentation model.(23) The HVPG model predicted an HVPG category (0-5) within fibrotic regions in the tissue directly from images of trichrome-stained images. Each pixel of fibrotic tissue within each image was associated with an HVPG category. To compute a slide-level ML HVPG score, the pixels predicted to be each category of HVPG within fibrosis were first summed and normalized by total area of fibrosis. Each fraction was then multiplied by the corresponding HVPG category label (0-5) and then summed to generate a slide-level, weighted average ML HVPG score. Heatmaps generated by the model depicted pixel-level associations of fibrosis tissue with the six HVPG categories (**Figure 1**). Model training was monitored, and hyperparameters adjusted based on the Spearman correlation of the ML HPVG score with true HVPG on the training set, until convergence was

achieved. Once model trained was completed, the algorithm was locked and deployed on the test set.

The ML HVPG score ranges from 0 to 5, consistent with the six HVPG categories used in its derivation. In a pilot study in which 29 biopsy images from the validation set were evaluated twice by the ML model, perfect agreement was observed between the first and second reading (**Supplemental Figure 1**).

Machine Learning Assessment of Other Histological Features

ML parameters from previously developed models for other histological features of NASH were also included in this study (23). Briefly, these models were trained using pathologist annotations to detect NASH features including steatosis, lobular inflammation, hepatocellular ballooning, and portal inflammation on images of H&E-stained biopsies, and fibrosis and bile ducts on images of trichrome-stained biopsies. Slide-level ML parameters were generated by computing proportionate areas of each histologic feature, including fibrosis patterns consistent with each NASH CRN fibrosis stage. In addition, we calculated the weighted mean of these predictions to generate a single, slide-level, continuous score referred to as the ML NASH CRN fibrosis score, that summarizes the underlying heterogeneity of fibrosis in the slide.

Two additional ML parameters were computed to quantify nodules on images of trichrome-stained biopsies (**Supplemental Figure 2**). In brief, fibrotic and non-fibrotic tissues were identified using a previously developed ML model (24), and connected component analysis was performed using morphological image analysis operations to identify nodules as non-contiguous regions of non-fibrotic tissue. ML nodularity was defined as the proportion of nodule perimeter surrounded by fibrosis, and ML normalized nodule perimeter was defined as the nodule perimeter surrounded by fibrosis, normalized by nodule area. Both parameters were determined for all nodules in a biopsy, and image-level averages of the parameters were computed.

Statistical Analyses

Associations between true (measured) HVPG with the ML HVPG score and other parameters were evaluated using Spearman correlations and Mann-Whitney tests. Performance of the ML HVPG score

and other parameters for predicting CSPH was determined using logistic regression models and areas under receiver operating characteristic curves (AUROCs). To evaluate the performance of combinations of parameters for predicting CSPH, stepwise multivariate models including ML histologic and routine clinical parameters (e.g., liver biochemistry, platelets, APRI, FIB-4, ELF) were trained in the training set ($p<0.20$ for model entry and $p<0.15$ for model retention) and the fitted models were then applied to the test set.

Associations between the ML HVPG score and measured HVPG, both at baseline and changes over time, with time to first liver-related clinical event (as defined above) were evaluated using log-rank tests and Cox proportional hazards regression. We implemented a last observation carried forward approach, including baseline, to impute missing post baseline values. Univariate models were used for baseline predictors, while all models for change from baseline adjusted for baseline values. Discrimination of the parameters at baseline for clinical events during follow-up was evaluated using c-statistics, which are analogous to AUROCs for survival data.

RESULTS

Study Population

The median age of the study cohort was 57 years (IQR 51, 61); most patients were female (62%), white (94%), and approximately two-thirds had diabetes (**Table 1**). The median HVPG at baseline was 12.0 mm Hg (IQR 8.5 mm Hg, 16.0 mm Hg), 69% of patients had CSPH, and 15% were taking non-selective beta blockers. Patients in the training and test sets had similar demographics, liver biochemistry, and fibrosis assessments; however, median serum albumin was lower in the test set (3.7 vs 3.9 g/dL; $p=0.018$). The median length of liver biopsy specimens at baseline was 2.3 cm (IQR 1.6, 3.3); biopsy length was similar between the training and test sets.

A Machine Learning-Based Model Categorizes Discrete Histologic Regions Associated with HVPG in NASH Cirrhosis

The ML model for HVPG generated a pixel-wise classification of fibrotic tissue and calculated the weighted mean of these predictions to generate the ML HVPG score. The resulting slide-level features capture the heterogeneity of fibrosis severity and patterns on the entire slide for the

generated HVPG predictions. Illustrative trichrome-stained biopsies with heatmaps representing the ML-based predictions of HVPG are provided in **Figure 1**. In a patient with a measured HVPG of 4.5 mm Hg, relatively sparse bands of fibrosis (hepatic collagen, 3.2%; ML NASH CRN fibrosis score, 3.06) were observed along with a ML HVPG score of 0.42 **(Figure 1A)**. In contrast, two other specimens (**Figures 1B and 1C**), demonstrated thick fibrous bands with markedly positive trichrome staining (hepatic collagen ~13%; ML NASH CRN fibrosis score ~3.6), yet the less prominently stained section (**Figure 1C**) recorded a higher ML HVPG score (4.03 vs 2.90), consistent with higher measured HVPG (22 vs 12.5 mm Hg in **Figure 1B**). These data demonstrate that the ML HVPG algorithm can detect fibrotic regions corresponding to higher true HVPG despite differences in staining intensity.

Correlations Between Measured HVPG with the ML HVPG Score and Other Parameters

The ML HVPG score was significantly correlated with true HVPG in both the training (ρ=0.67; p<0.001 [**Figure 2A**]) and test sets (ρ=0.47; p<0.001 [**Figure 2B, Table 2**]). In the test set, these correlations were similar between patients with (ρ=0.35; p=0.019) and without CSPH (ρ=0.42; p=0.0495). To determine whether the end-to-end, ML HVPG model was simply analogous to a continuous score of hepatic fibrosis, we assessed the association between hepatic collagen content by conventional morphometry and true HVPG. As shown in **Figure 2C**, hepatic collagen was less strongly correlated with true HVPG in the test set (ρ=0.28; p<0.001) than the ML HVPG score. In contrast, the ELF score demonstrated a higher correlation with true HVPG (ρ=0.60; p<0.001 [**Figure 2D**]). Correlations between true HVPG and ML parameters for nodularity (ρ=0.15; p=0.04) and normalized nodule perimeter (ρ=0.21; p<0.001) were modest, yet statistically significant.

Correlations between the ML HVPG score and true HVPG with other parameters in the test set are provided in **Table 2**. The ML HVPG score was moderately correlated with hepatic collagen by morphometry (ρ=0.43; p<0.0001), ELF (ρ=0.44; p<0.0001) and other noninvasive tests of fibrosis (NITs), platelets (ρ=-0.39; p<0.0001), albumin (ρ=-0.44; p<0.0001), INR (ρ=-0.39; p<0.0001), and MELD (ρ=0.35; p<0.0001). The ML HVPG score was also correlated with other ML histologic parameters including the ML NASH CRN fibrosis score (ρ=0.39; p<0.0001), and proportionate areas of F4 fibrosis (ρ=0.44; p<0.0001), portal inflammation (ρ=0.50; p<0.0001), bile ducts (ρ=0.59;

p<0.0001), and steatosis (ρ=-0.27; p=0.0001). In general, correlations between true HVPG and these parameters were similar, except for stronger correlations between the ML HVPG score and other ML features for fibrosis, portal inflammation, and bile duct area.

ML HVPG Score Discriminates Clinically Distinct Ranges of HVPG

The ability to discern CSPH from normal and mildly elevated portal pressures based on slide-level analysis alone would be expected to have prognostic value.(13, 14) The slide-level ML HVPG score differentiated patients with normal HVPG (0-5 mm Hg), mildly elevated HVPG (5.5-9.5 mm Hg), and CSPH **(**median ML HVPG scores: 1.51 vs 1.93 vs 2.60; **Figure 3A**). However, among patients with CSPH, the ML HVPG score could not discriminate between smaller HVPG categories (**Supplemental Figure 3**). Similar findings were observed for the ELF score (**Figure 3B**). Hepatic collagen content by morphometry was able to differentiate patients with CSPH from those with elevated HVPG (median hepatic collagen: 13.3 vs 10.0%; p<0.001), but not between those with normal and elevated HVPG (**Figure 3C**).

The discrimination of the ML HVPG score and other parameters, alone or in combination, for CSPH is outlined in **Table 3**. In the training and test sets, the AUROCs of the ML HVPG score for CSPH were 0.85 (95% CI 0.80, 0.90) and 0.76 (95% CI 0.68, 85), respectively. Performance of the ML HVPG score was not influenced by beta blocker therapy at baseline (AUROCs in patients on vs not on beta blockers: 0.66 (0.27, 1.00) vs 0.78 (0.69, 0.86) p=0.55). Corresponding AUROCs for the ELF score and hepatic collagen content in the test set were 0.78 (95% CI 0.71, 0.86 [p=0.42 vs ML HVPG score) and 0.65 (95% CI 0.56, 0.75 [p=0.054 vs ML HVPG score]), respectively.

Detection of CSPH with the ML HVPG score was improved by the addition of ELF (AUROC in test set: 0.82 [95% CI 0.75, 0.89]), APRI (AUROC: 0.81 [95% CI 0.74, 0.89]), FIB-4 (AUROC: 0.85 [95% CI 0.75, 0.81]), and a combination of ELF, platelets, AST, and bilirubin (AUROC 0.84 [95% CI 0.78, 0.91] [**Table 3**]). Addition of other ML histologic parameters including the ML NASH CRN fibrosis score and ML nodularity had minimal impact on diagnostic performance of the ML HVPG score (AUROC in test set: 0.76 [95% CI 0.68, 0.84]). A full model including optimal ML and clinical

parameters (ML HVPG score, ML nodularity, ELF, platelets, and AST) had an AUROC of 0.93 (95% CI 0.90, 0.97) in the training set and 0.85 (95% CI 0.78, 0.92) in the test set.

Associations Between Changes in ML HVPG Score and Improvements in Fibrosis and HVPG

Reductions in portal pressure may be observed in patients with NASH cirrhosis treated with hemodynamically active therapies such as beta blockers or due to fibrosis regression.(25) We therefore compared changes in the ML HVPG score according to the presence or absence of a >20% decrease in HVPG or Ishak fibrosis stage improvement from baseline to week 96 (**Figure 4**). Changes in the ML HVPG score and true HVPG between baseline and week 96 were correlated (ρ=0.24; p=0.012) and similar between patients with and without CSPH (ρ=0.19 vs 0.21). In patients with a >20% decrease in HVPG, the median (IQR) ML HVPG score declined more than among those who did not achieve a hemodynamic response (-0.8 [-1.3, -0.3] vs -0.4 [-0.7, 0.0]; p=0.013 [**Figure 4A]**). Differences in NITs such as ELF and FIB-4 at baseline and changes from baseline to week 96 were also seen in patients with versus without a hemodynamic response (**Supplemental Table 1)**. Similar findings were observed when patients were categorized according to a >10% decrease in HVPG (**Supplemental Figure 4**). Similarly, the ML HVPG score declined significantly in patients with versus without fibrosis improvement (-1.1 [-1.6, -0.5] vs -0.4 [-0.7, 0.0]; p<0.001 **[Figure 4B])**. These data indicate that the ML HVPG score is responsive to changes in true HVPG and fibrosis stage in patients with cirrhosis due to NASH.

Prognostic Utility of the ML HVPG Score for Liver-Related Clinical Events

During a median follow-up of 31.0 months (IQR 27.9, 35.0), 33/176 patients (19%) in both the training and test sets had liver-related clinical events (**Supplemental Table 2**). At baseline, true HVPG (hazard ratio [HR] per mm Hg: 1.181 [95% CI 1.107, 1.261]; p<0.001 [**Figure 5A**]), but not ML HVPG score (HR per unit: 1.349 [95% CI 0.836, 2.178]; p=0.22 [**Figure 5B**]), was associated with liver-related clinical events (c-statistics 0.705 vs 0.558; p=0.036). After adjustment for baseline values, increases from baseline in both true HVPG and ML HVPG score were associated with an increased risk of clinical events (**Figure 4C**).

A recent publication described underestimation of portal pressure with HVPG measurement in patients with cirrhosis due to NASH.(8) In the current study, among 32 patients who had liver-related clinical events during the trial, six had baseline HVPG below 10 mm Hg. Baseline ML HVPG scores in these patients were non-significantly lower than among those with clinical events and CSPH at baseline (p=0.084; **Supplemental Figure 5**).

DISCUSSION

The present study utilized a CNN model to learn the morphological signatures of fibrosis associated with HVPG using trichrome-stained liver biopsy slides from patients with cirrhosis due to NASH. Importantly, the ML-predicted HVPG score was able to discriminate CSPH from normal and elevated portal pressure, and changes in the score predicted the occurrence of liver-related clinical events. The performance of the model was not influenced by concomitant beta blocker therapy. Advantages of this ML-based approach over human assessment of HVPG include the lack of requirement for specialized interventional radiology procedures and facilities, or the human expertise to interpret HVPG tracings, if a liver biopsy is available.(7) Moreover, the ML HVPG score is an additional piece of clinical data that can be extracted from liver histology, which is frequently performed in patients with NASH, particularly in the context of clinical trials. Importantly, the ML HVPG score was derived and validated using trichrome-stained liver biopsy slides, which are collected in routine clinical practice, eliminating the need for special staining. Indeed, ML models generated using slides stained for picrosirius red and α-smooth muscle actin did not outperform the model using trichrome-stained slides (data not shown). Finally, unlike the interpretation of HVPG tracings and liver histology by humans, the ML HVPG score is 100% reproducible; measurements taken from the same histologic slide produce identical scores on repeat evaluation (**Supplemental Figure 1**). This provides a defined benchmark for evaluation of changes in the score that may occur with intervention.

Used alone, the performance of the ML HVPG score for discriminating CSPH was superior to hepatic collagen by morphometry (AUROCs in test set: 0.76 vs 0.65), and similar to that of the ELF score (AUROC 0.78). Platelet count, the non-proprietary fibrosis markers, APRI and FIB-4, and a combination of clinical parameters (ELF, platelets, APRI, and bilirubin) also had good diagnostic performance (AUROCs 0.75 to 0.83). These data highlight the potential of the ML HVPG score,

NITs, and routine clinical parameters to identify CSPH in NASH patients with cirrhosis. The superiority of the ML HVPG score over hepatic collagen quantified by conventional morphometry suggests that the ML model is capturing phenotypic features beyond just the quantity of fibrosis. Indeed, as shown in **Figure 1**, despite very similar metrics of fibrosis (hepatic collagen, ML NASH CRN fibrosis score) in the representative patients, the ML model can identify characteristics of the fibrotic matrix associated with increased portal pressure. However, due to the 'end-to-end' nature of the model, the specific features that contribute to this improved diagnostic performance cannot be elucidated.

The discrimination of ELF, FIB-4, and APRI for CSPH is in keeping with prior observations and supports the utility of non-biopsy derived serum markers that have been reported in multiple studies.(10, 16, 21, 26) Models combining the ML HVPG score with clinical parameters associated with portal hypertension (ELF, platelets, AST, bilirubin, APRI, FIB-4) improved upon the performance of the ML HVPG score for detecting CSPH (AUROCs 0.81 to 0.85 in the test set). A 'full' model, which included the ML HVPG score, a ML feature of nodularity, ELF, platelets, and AST had very good discrimination for CSPH (AUROC 0.85). These findings suggest that these alternative features may be capturing pathophysiologic components other than the fibrotic scar itself that are reflective of the 'dynamic component' of portal hypertension (e.g., inflammation and endothelial dysfunction).(3, 27) Indeed, the ELF score was significantly correlated with portal inflammation assessed by ML in this dataset (data not shown). Several additional ML histologic features including proportionate areas of bile ducts, lobular inflammation, and steatosis, had prognostic utility in isolation but not in multivariate analyses, perhaps due to the small sample size of the study. Inclusion of such features may help to identify patients in which HVPG underestimated their actual portal pressure.(8)

In light of data from a study that identified small nodule size as an independent predictor of CSPH in a cohort with predominantly alcoholic or viral cirrhosis,(13) we attempted to train the ML algorithm to recognize parenchymal nodularity (**Supplemental Figure 2**). These parameters were weakly correlated with HVPG (**Table 2**) and had limited diagnostic performance for CSPH alone (AUROCs ~0.60), perhaps due to the different stages of cirrhosis in the two studies. The majority of patients

(34/43) in the study that identified small nodule size as a predictor of CSPH had liver biopsies performed in the context of transjugular intrahepatic portosystemic shunt insertion, and therefore likely had more advanced cirrhosis than those in the present study. Nevertheless, the addition of ML nodularity to the ML HVPG score improved its discrimination for CSPH (**Table 3**). These findings suggest that refinement of these nodularity features and/or exploration of other novel parameters should be pursued in order to improve the performance of the ML models for assessment of HVPG.

Our study confirmed the relationship between the cirrhotic architecture, portal pressure, and the incidence of hepatic decompensation in NASH cirrhosis. Notably, unlike measured HVPG, the ML HVPG score at baseline was not predictive of clinical events. We speculate that this may stem from the inability of the ML algorithm to distinguish grades of portal pressure above the CSPH threshold (i.e., between HVPG categories above 10 mm Hg [**Supplemental Figure 3**]). Similar observations have been made for ELF and for liver stiffness by vibration-controlled transient elastography; specifically, weaker correlations with HVPG at higher portal pressures.(10, 28) In these patients, pathophysiologic contributors to portal hypertension other than fibrosis, such as splanchnic vasodilatation and alterations within the intrahepatic vasculature, may play a more prominent role, especially in decompensated patients (not included in this study).(3) With this in mind, we attempted to train the ML algorithm to recognize vascular features that differentiated different degrees of CSPH but were unsuccessful in this endeavor (data not shown).

Another approach to testing whether the ML approach recognizes changes to the vascular component that are not obvious to human observation is to examine associations between the ML HVPG score and hemodynamic responses to treatment. Unfortunately, only a small number of patients started beta blocker therapy during the trial (n=8 who provided consent for this analysis). However, 27 patients experienced a >20% decrease in HVPG between baseline and week 96, potentially due to lifestyle modification.(29) In these patients, the ML HVPG score declined more than among those who did not achieve a hemodynamic response (**Figure 4A**). Consistent with this observation, a statistically significant association was observed between changes in the ML HVPG score and measured HVPG (ρ=0.24; p=0.012). These data suggest that the ML HVPG score is responsive to changes in true HVPG and are supported by the statistically significant association between changes in the ML

HVPG score and the incidence of liver-related clinical events (**Figure 5C**). Finally, the ML HVPG score declined significantly in patients with, compared to those without, an improvement in fibrosis stage (**Figure 4B**). In totality, these data suggest that the ML HVPG score is responsive to changes in parameters that have prognostic relevance for patients with NASH cirrhosis. If validated, these observations support consideration of the ML HVPG score as a surrogate endpoint in clinical trials of therapies for NASH cirrhosis.

Our study has several strengths including the measurement of HVPG according to a standardized protocol, centralized review of HVPG measurements and liver histology, and prospective adjudication of liver-related clinical events in the context of a rigorously conducted clinical trial. However, several limitations of the study warrant discussion. First, the test set was small; hence, external validation in an additional dataset is necessary. Second, the number of patients with clinical events and duration of follow-up were limited. Therefore, firm conclusions regarding the prognostic utility of the ML HVPG score require evaluation in larger cohorts with longer follow-up. Third, the study was restricted to patients with compensated cirrhosis due to NASH. Validation of the ML HVPG score in patients with cirrhosis of alternate etiologies is necessary to confirm generalizability of our findings. In addition, liver fibrosis was not staged according to the Laennec fibrosis scoring system, which has been associated with the severity of portal hypertension and risk of clinical events.(30, 31) In addition, we did not collect data on liver stiffness in this trial, although it has demonstrated utility for predicting CSPH.(16) Finally, data regarding potentially confounding factors that may influence HVPG, including alcohol consumption, diet, and exercise, are not available. However, changes in body weight were not associated with changes in HVPG (data not shown).

In summary, a score based on a ML algorithm trained on liver biopsies alone is correlated with measured HVPG and can identify CSPH in patients with cirrhosis due to NASH. Continued improvements in ML-based assessments may further refine the accuracy of the ML HVPG score, identify novel histologic parameters associated with portal hypertension, and define features that predict response to interventions that reduce HVPG.

ACKNOWLEDGEMENTS

This manuscript is dedicated to the memory of Dr. Roberto J. Groszmann who made seminal contributions to the understanding of cirrhosis and portal hypertension and pioneered the technique of using a balloon catheter to measure the hepatic venous pressure gradient.

We thank the patients who participated in this study, their families, and all participating investigators. Sandra Chen of Gilead Sciences provided editorial assistance. This analysis was supported by Gilead Sciences.

REFERENCES

1. Groszmann RJ, Garcia-Tsao G, Bosch J, Grace ND, Burroughs AK, Planas R, et al. Beta-blockers to prevent gastroesophageal varices in patients with cirrhosis. N Engl J Med. 2005;353(21):2254-61.
2. Ripoll C, Groszmann R, Garcia-Tsao G, Grace N, Burroughs A, Planas R, et al. Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. Gastroenterology. 2007;133(2):481-8.
3. Bosch J, Groszmann RJ, Shah VH. Evolution in the understanding of the pathophysiological basis of portal hypertension: How changes in paradigm are leading to successful new treatments. J Hepatol. 2015;62(1 Suppl):S121-30.
4. Feu F, Garcia-Pagan JC, Bosch J, Luca A, Teres J, Escorsell A, et al. Relation between portal pressure response to pharmacotherapy and risk of recurrent variceal haemorrhage in patients with cirrhosis. Lancet. 1995;346(8982):1056-9.
5. Groszmann RJ, Bosch J, Grace ND, Conn HO, Garcia-Tsao G, Navasa M, et al. Hemodynamic events in a prospective randomized trial of propranolol versus placebo in the prevention of a first variceal hemorrhage. Gastroenterology. 1990;99(5):1401-7.
6. Silva-Junior G, Baiges A, Turon F, Torres F, Hernandez-Gea V, Bosch J, et al. The prognostic value of hepatic venous pressure gradient in patients with cirrhosis is highly dependent on the accuracy of the technique. Hepatology. 2015;62(5):1584-92.
7. Groszmann RJ, Wongcharatrawee S. The hepatic venous pressure gradient: anything worth doing should be done right. Hepatology. 2004;39(2):280-2.
8. Ferrusquía-Acosta J, Bassegoda O, Turco L, Reverter E, Pellone M, Bianchini M, et al. Agreement between wedged hepatic venous pressure and portal pressure in non-alcoholic steatohepatitis-related cirrhosis. Journal of hepatology. 2020 [In Press].
9. Abraldes JG, Bureau C, Stefanescu H, Augustin S, Ney M, Blasco H, et al. Noninvasive tools and risk of clinically significant portal hypertension and varices in compensated cirrhosis: The "Anticipate" study. Hepatology. 2016;64(6):2173-84.

10. Simbrunner B, Marculescu R, Scheiner B, Schwabl P, Bucsics T, Stadlmann A, et al. Non-invasive detection of portal hypertension by enhanced liver fibrosis score in patients with different aetiologies of advanced chronic liver disease. Liver Int. 2020;40(7):1713-24.
11. Shili-Masmoudi S, Wong GL, Hiriart JB, Liu K, Chermak F, Shu SS, et al. Liver stiffness measurement predicts long-term survival and complications in non-alcoholic fatty liver disease. Liver Int. 2020;40(3):581-9.
12. Berzigotti A, Seijo S, Arena U, Abraldes JG, Vizzutti F, Garcia-Pagan JC, et al. Elastography, spleen size, and platelet count identify portal hypertension in patients with compensated cirrhosis. Gastroenterology. 2013;144(1):102-11 e1.
13. Nagula S, Jain D, Groszmann RJ, Garcia-Tsao G. Histological-hemodynamic correlation in cirrhosis-a histological classification of the severity of cirrhosis. J Hepatol. 2006;44(1):111-7.
14. Sethasine S, Jain D, Groszmann RJ, Garcia-Tsao G. Quantitative histological-hemodynamic correlations in cirrhosis. Hepatology. 2012;55(4):1146-53.
15. Calvaruso V, Burroughs AK, Standish R, Manousou P, Grillo F, Leandro G, et al. Computer-assisted image analysis of liver collagen: relationship to Ishak scoring and hepatic venous pressure gradient. Hepatology. 2009;49(4):1236-44.
16. Sanyal AJ, Harrison SA, Ratziu V, Abdelmalek MF, Diehl AM, Caldwell S, et al. The Natural History of Advanced Fibrosis Due to Nonalcoholic Steatohepatitis: Data From the Simtuzumab Trials. Hepatology. 2019;70(6):1913-27.
17. Vanderbeck S, Bockhorst J, Kleiner D, Komorowski R, Chalasani N, Gawrieh S. Automatic quantification of lobular inflammation and hepatocyte ballooning in nonalcoholic fatty liver disease liver biopsies. Hum Pathol. 2015;46(5):767-75.
18. Liu F, Goh GB, 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(6):1953-66.
19. Forlano R, Mullish BH, Giannakeas N, Maurice JB, Angkathunyakul N, Lloyd J, et al. High-Throughput, Machine Learning-Based Quantification of Steatosis, Inflammation, Ballooning, and

Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol. 2020;18(9):2081-90 e9.

20. Loomba R, Noureddin M, Kowdley KV, Kohli A, Sheikh A, Neff G, et al. Combination Therapies Including Cilofexor and Firsocostat for Bridging Fibrosis and Cirrhosis Attributable to NASH. Hepatology (Baltimore, Md). 2021;73(2):625-43.

21. Harrison SA, Abdelmalek MF, Caldwell S, Shiffman ML, Diehl AM, Ghalib R, et al. Simtuzumab Is Ineffective for Patients With Bridging Fibrosis or Compensated Cirrhosis Caused by Nonalcoholic Steatohepatitis. Gastroenterology. 2018;155(4):1140-53.

22. Berzigotti A. Non-invasive evaluation of portal hypertension using ultrasound elastography. J Hepatol. 2017;67(2):399-411.

23. 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.

24. Pokkalla H, Pethia K, Taylor-Weiner A, Glass B, Elliott H, Han L, et al. Machine Learning Models Identify Novel Histologic Features Predictive of Clinical Disease Progression in Patients With Advanced Fibrosis Due to Nonalcoholic Steatohepatitis. Journal of hepatology. 2020;73(Suppl 1):S402.

25. Lassailly G, Caiazzo R, Ntandja-Wandji LC, Gnemmi V, Baud G, Verkindt H, et al. Bariatric Surgery Provides Long-term Resolution of Nonalcoholic Steatohepatitis and Regression of Fibrosis. Gastroenterology. 2020;159(4):1290-301 e5.

26. Anstee QM, Lawitz EJ, Alkhouri N, Wong VW, Romero-Gomez M, Okanoue T, et al. Noninvasive Tests Accurately Identify Advanced Fibrosis due to NASH: Baseline Data From the STELLAR Trials. Hepatology (Baltimore, Md). 2019;70(5):1521-30.

27. Gronbaek H, Sandahl TD, Mortensen C, Vilstrup H, Moller HJ, Moller S. Soluble CD163, a marker of Kupffer cell activation, is related to portal hypertension in patients with liver cirrhosis. Aliment Pharmacol Ther. 2012;36(2):173-80.

28. Vizzutti F, Arena U, Romanelli RG, Rega L, Foschi M, Colagrande S, et al. Liver stiffness measurement predicts severe portal hypertension in patients with HCV-related cirrhosis. Hepatology. 2007;45(5):1290-7.
29. Berzigotti A, Albillos A, Villanueva C, Genesca J, Ardevol A, Augustin S, et al. Effects of an intensive lifestyle intervention program on portal hypertension in patients with cirrhosis and obesity: The SportDiet study. Hepatology. 2017;65(4):1293-305.
30. Kim SU, Oh HJ, Wanless IR, Lee S, Han KH, Park YN. The Laennec staging system for histological sub-classification of cirrhosis is useful for stratification of prognosis in patients with liver cirrhosis. Journal of hepatology. 2012;57(3):556-63.
31. Kim MY, Cho MY, Baik SK, Park HJ, Jeon HK, Im CK, et al. Histological subclassification of cirrhosis using the Laennec fibrosis scoring system correlates with clinical stage and grade of portal hypertension. Journal of hepatology. 2011;55(5):1004-9.

Table 1. Baseline Demographics and Clinical Characteristics

	Training Set (N=130; 320 slides)	Test Set (N=88; 216 slides)	p-Value
Demographics and Concomitant Medications			
Age, y	56 (50, 60)	58 (53, 62)	0.11
White	121 (93)	83 (94)	0.79
Hispanic	21 (16)	14 (16)	1.00
Woman	76 (58)	60 (68)	0.16
BMI, kg/m^2	33.11 (29.91, 38.63)	35.02 (30.11, 39.17)	0.54
Diabetes	79 (67)	56 (75)	0.27
Beta blocker therapy	20 (15)	13 (15)	1.00
Liver Biochemistry			
ALT, U/L	36 (27, 50)	35 (24, 47)	0.42
AST, U/L	41 (31, 54)	41 (29, 55)	0.51
ALP, U/L	93 (72, 113)	100 (78, 128)	0.13
GGT, U/L	72 (39, 122)	87 (49, 136)	0.22
Albumin, mg/dL	3.9 (3.7, 4.2)	3.7 (3.5, 4.1)	**0.018**
Total bilirubin, mg/dL	0.7 (0.4, 1.1)	0.6 (0.5, 1.0)	0.54
Platelets, $x10^3/mL$	138 (97, 191)	151 (96, 192)	0.49
MELD	7 (6, 8)	7 (6, 8)	0.08
Fibrosis Parameters and HVPG			
F5	37/128 (29)	23/87 (26)	0.39
F6	77/128 (60)	57/87 (66)	
Hepatic collagen content, %	11.15 (7.80, 17.40)	14.25 (7.80, 20.10)	0.10
ML NASH CRN fibrosis score	3.42 (2.95, 3.63)	3.26 (3.07, 3.52)	0.51
ELF score	10.82 (9.91, 11.48)	10.65 (9.92, 11.52)	0.86

HVPG, mm Hg	12.0 (8.5, 16.0)	12.0 (8.5, 16.0)	0.99
0 - 5 mm Hg (normal)	1/105 (1)	2/70 (3)	0.70
5.5 – 9.5 mm Hg (elevated)	32/105 (30)	20/70 (29)	
≥10 mm Hg (CSPH)	72/105 (69)	48/70 (69)	

Data are median (IQR) or n (%).

Table 2. Correlations Between ML HVPG Score and True HVPG with Other Parameters in the Test Set

Parameters	Correlations with ML HVPG Score			Correlations with True HVPG		
	N	Correlation Coefficient (ρ)	p-Value	N	Correlation Coefficient (ρ)	p-Value
True HVPG	178	0.4698	<0.001	--	--	--
ML HVPG score*	--	--	--	178	0.4698	<0.001
Hepatic collagen content	168	0.4284	<0.001	174	0.2812	<0.001
ELF	175	0.4356	<0.001	183	0.5986	<0.001
FIB-4	188	0.4751	<0.001	179	0.5953	<0.001
ALT	193	0.0566	0.43	184	-0.1271	0.086
AST	190	0.2844	<0.001	180	0.1319	0.078
GGT	194	0.2688	<0.001	185	0.0736	0.32
Platelets	195	-0.3855	<0.001	186	-0.5331	<0.001
MELD	192	0.3511	<0.001	182	0.4828	<0.001
Total bilirubin	195	0.3157	<0.001	186	0.3409	<0.001
INR	194	0.3865	<0.001	184	0.4525	<0.001
Albumin	195	-0.4422	<0.001	186	-0.5241	<0.001
ML NASH CRN fibrosis score*	198	0.3940	<0.001	182	0.1589	0.032
ML proportionate area of F4*	198	0.4407	<0.001	182	0.2109	0.004
ML proportionate area of ≤F2*	198	-0.3149	<0.001	182	-0.1081	0.146
ML proportionate area of steatosis	196	-0.2747	<0.001	184	-0.3789	<0.001
ML proportionate area of hepatocellular ballooning	196	0.1223	0.088	184	-0.0469	0.53
ML proportionate area of lobular inflammation	196	-0.2775	<0.001	184	-0.2937	<0.001
ML proportionate area of portal inflammation	196	0.4964	<0.001	184	0.2808	<0.001
ML proportionate area of bile ducts*	198	0.5854	<0.001	182	0.4107	<0.001
ML nodularity*	194	0.2306	0.001	175	0.1874	0.013
ML normalized nodule perimeter*	194	0.2452	0.001	175	0.1466	0.053

* ML parameters from trichrome-stained liver biopsy slides. Other ML parameters from H&E-stained slides.

Table 3. Discrimination of ML HVPG Score and Other Parameters for CSPH

Predictors of CSPH	Training Set	Test Set					
	AUROC (95% CI)	AUROC (95% CI)	Optimal Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
ML Parameters							
ML HVPG score*	0.85 (0.80, 0.90)	0.76 (0.68, 0.85)	2.01	85 (77, 92)	62 (47, 75)	81 (73, 88)	68 (53, 81)
ML NASH CRN fibrosis score*	0.77 (0.70, 0.83)	0.60 (0.50, 0.69)	3.51	42 (32, 52)	79 (65, 89)	80 (66, 89)	41 (31, 51)
ML proportionate area of F4, %*	0.77 (0.71, 0.83)	0.64 (0.54, 0.73)	59.5	52 (42, 62)	71 (57, 83)	78 (66, 87)	43 (32, 54)
ML proportionate area of ≤F2, %*	0.74 (0.68, 0.81)	0.56 (0.46, 0.65)	8.6	41 (32, 51)	77 (63, 87)	78 (64, 88)	40 (30, 50)
ML proportionate area of steatosis, %	0.55 (0.47, 0.64)	0.72 (0.64, 0.81)	1.2	45 (35, 55)	88 (77, 96)	88 (77, 96)	45 (35, 55)
ML proportionate area of lobular inflammation, %	0.56 (0.48, 0.64)	0.73 (0.64, 0.82)	0.7	80 (71, 88)	60 (45, 73)	80 (71, 87)	61 (46, 74)
ML proportionate area of portal inflammation, %	0.70 (0.63, 0.77)	0.65 (0.56, 0.74)	15.1	37 (28, 47)	90 (79, 97)	88 (75, 96)	42 (33, 52)
ML proportionate area of hepatocellular ballooning, %	0.63 (0.55, 0.70)	0.46 (0.36, 0.55)	1.9	53 (43, 63)	52 (38, 66)	68 (57, 78)	36 (25, 48)
ML proportionate area of bile ducts, %*	0.73 (0.66, 0.80)	0.72 (0.63, 0.80)	3.4	44 (34, 54)	88 (77, 96)	88 (76, 96)	45 (35, 55)
ML normalized nodule perimeter, %*	0.62 (0.55, 0.70)	0.58 (0.48, 0.67)	0.50	57 (47, 67)	63 (49, 76)	75 (64, 84)	43 (32, 55)
ML nodularity, %*	0.60 (0.52, 0.68)	0.59 (0.49, 0.69)	0.03	68 (58, 77)	50 (36, 64)	73 (63, 81)	44 (31, 58)

Clinical Parameters							
Hepatic collagen content	0.63 (0.55, 0.70)	0.65 (0.56, 0.75)	11.9	57 (47, 67)	67 (53, 80)	77 (66, 86)	44 (33, 56)
ELF	0.85 (0.80, 0.90)	0.78 (0.71, 0.86)	10.73	68 (58, 77)	81 (67, 90)	87 (78, 94)	56 (44, 67)
Platelets	0.80 (0.74, 0.86)	0.82 (0.75, 0.90)	136	73 (63, 81)	81 (67, 90)	88 (79, 94)	60 (48, 72)
AST	0.74 (0.67, 0.80)	0.54 (0.44, 0.63)	31	71 (61, 79)	38 (25, 53)	69 (59, 78)	40 (26, 55)
ALT	0.51 (0.44, 0.59)	0.43 (0.33, 0.52)	105	2 (0, 7)	100 (93, 100)	100 (16, 100)	34 (27, 42)
APRI	0.85 (0.80, 0.90)	0.75 (0.67, 0.83)	0.70	82 (74, 89)	58 (43, 71)	79 (70, 87)	63 (47, 76)
FIB-4	0.86 (0.81, 0.91)	0.82 (0.75, 0.89)	2.90	72 (62, 80)	92 (81, 98)	95 (87, 99)	62 (51, 73)
INR	0.62 (0.55, 0.69)	0.74 (0.66, 0.81)	1.1	68 (58, 77)	73 (59, 84)	83 (73, 90)	54 (41, 65)
Bilirubin	0.67 (0.59, 0.74)	0.68 (0.60, 0.77)	0.7	60 (50, 69)	75 (61, 86)	82 (72, 90)	49 (37, 60)
Albumin	0.70 (0.63, 0.77)	0.79 (0.71, 0.86)	3.9	75 (66, 83)	67 (53, 80)	82 (73, 89)	58 (45, 71)
MELD	0.61 (0.53, 0.68)	0.73 (0.65, 0.80)	7	76 (67, 84)	62 (47, 75)	80 (70, 87)	57 (43, 70)
Combined Parameters							
ML HVPG score + ELF	0.90 (0.86, 0.94)	0.82 (0.75, 0.89)	0.5	80 (71, 88)	71 (57, 83)	85 (76, 91)	65 (51, 77)
ML HVPG score + APRI	0.91 (0.87, 0.95)	0.81 (0.74, 0.89)	0.5	73 (63, 81)	75 (61, 86)	85 (76, 92)	58 (46, 70)
ML HVPG score + FIB-4	0.91 (0.87, 0.95)	0.85 (0.78, 0.91)	0.5	78 (69, 86)	79 (65, 89)	88 (79, 94)	65 (52, 77)
Optimal clinical parameters (ELF, platelets, APRI, bilirubin)[†]	0.90 (0.86, 0.94)	0.83 (0.76, 0.90)	0.5	77 (68, 85)	75 (61, 86)	86 (77, 92)	63 (50, 75)

ML HVPG score + optimal clinical parameters (ELF, platelets, AST, bilirubin)†	0.93 (0.90, 0.96)	0.84 (0.78, 0.91)	0.5	80 (71, 88)	75 (61, 86)	86 (78, 93)	66 (53, 78)
Optimal ML histologic parameters (ML HVPG score, ML NASH CRN fibrosis score, ML nodularity)†	0.86 (0.81, 0.91)	0.76 (0.68, 0.84)	0.5	80 (71, 88)	56 (41, 70)	78 (69, 86)	59 (44, 73)
Full model including optimal ML histologic and clinical parameters† (ML HVPG score, ELF, platelets, AST, ML nodularity)	0.93 (0.90, 0.97)	0.85 (0.78, 0.92)	0.5	81 (72, 88)	77 (63, 87)	87 (79, 93)	68 (54, 79)

NPV, negative predictive value; PPV, positive predictive value.

***** ML parameters from trichrome-stained liver biopsy slides. Other ML parameters from H&E-stained slides.

† Parameters listed in order of decreasing statistical significance. Stepwise multivariate models included all ML histologic and/or clinical parameters listed in Table. Final models included variables with $p<0.25$ for model entry and $p<0.15$ for model retention. Models from training set applied to test set for validation. Sensitivity, specificity, PPV, and NPV for combined parameters are at model probability estimate of 0.5.

FIGURE LEGENDS

Figure 1. ML-based models for estimating HVPG on Masson's trichrome-stained biopsies. Images on left of trichrome-stained biopsies. Images on right represent heatmap overlay of fibrosis patterns associated with HVPG categories. The colors reflect fibrosis patterns consistent with each of the HVPG categories (0-5) shown in the legend.

Figure 2. Correlations between true HVPG and ML HVPG score, hepatic collagen content, and ELF. Data from the (A) training set and (B-D) test sets.

Figure 3. Associations between true HVPG and ML HVPG score, hepatic collagen content, and ELF. All data from the test set. Comparisons between groups based on Mann-Whitney test.

Figure 4. Associations between changes in the ML HVPG score between baseline and week 96 with (A) hemodynamic response and (B) fibrosis improvement. Data from the training and test sets. Comparisons between groups based on Mann-Whitney test.

Figure 5. Associations between ML HVPG score and true HVPG with liver-related clinical events. (A-B) Cutoffs for ML HVPG score and true HVPG based on maximal sum of sensitivity and specificity for discrimination of clinical events. (C) Models for change from baseline adjusted for baseline value. All analyses include training and test sets.

Figure 1. ML-based models for estimating HVPG on trichrome-stained biopsies

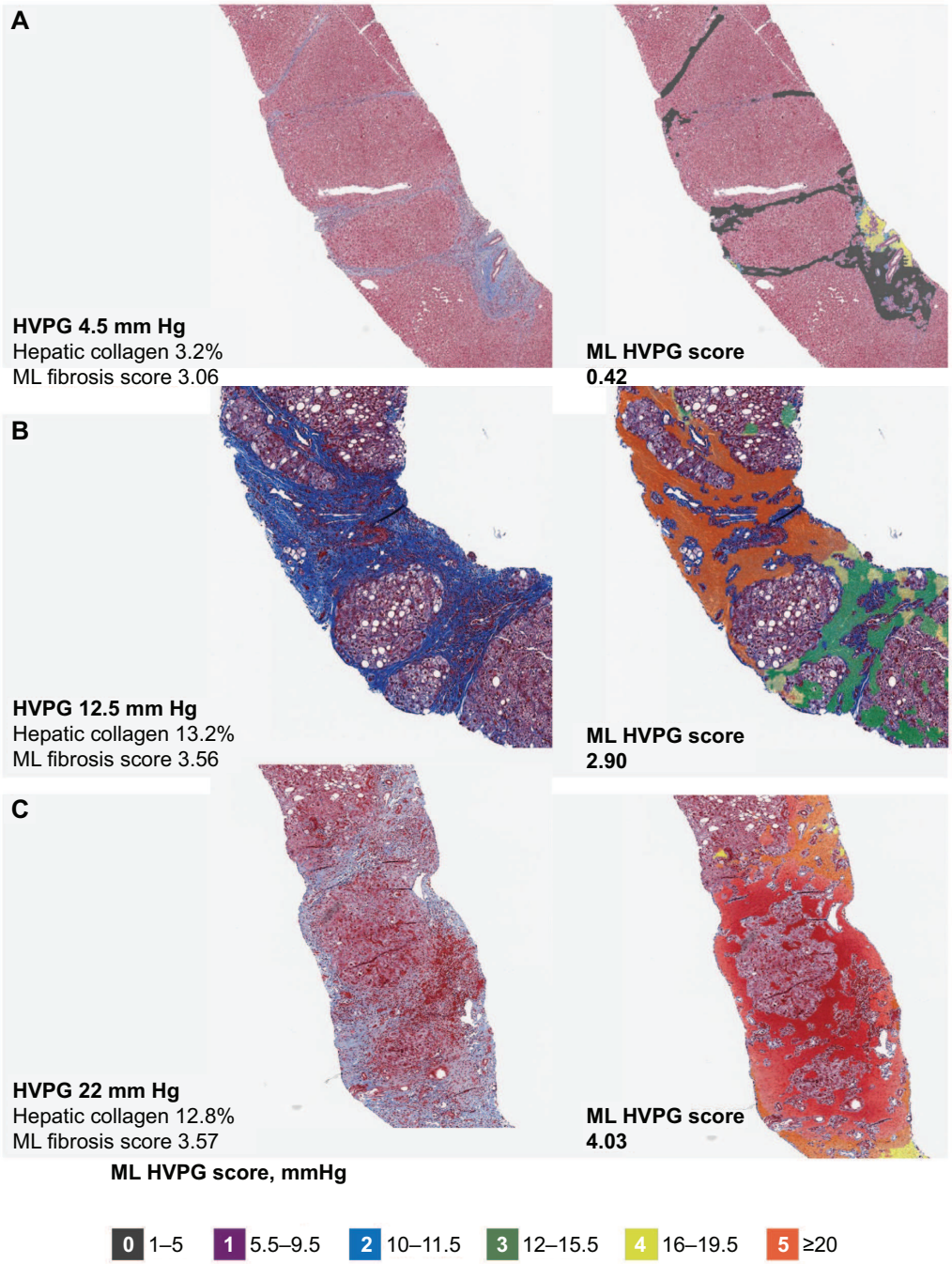


Figure 2. Correlations between true HVPG and ML HVPG score, hepatic collagen content, and ELF

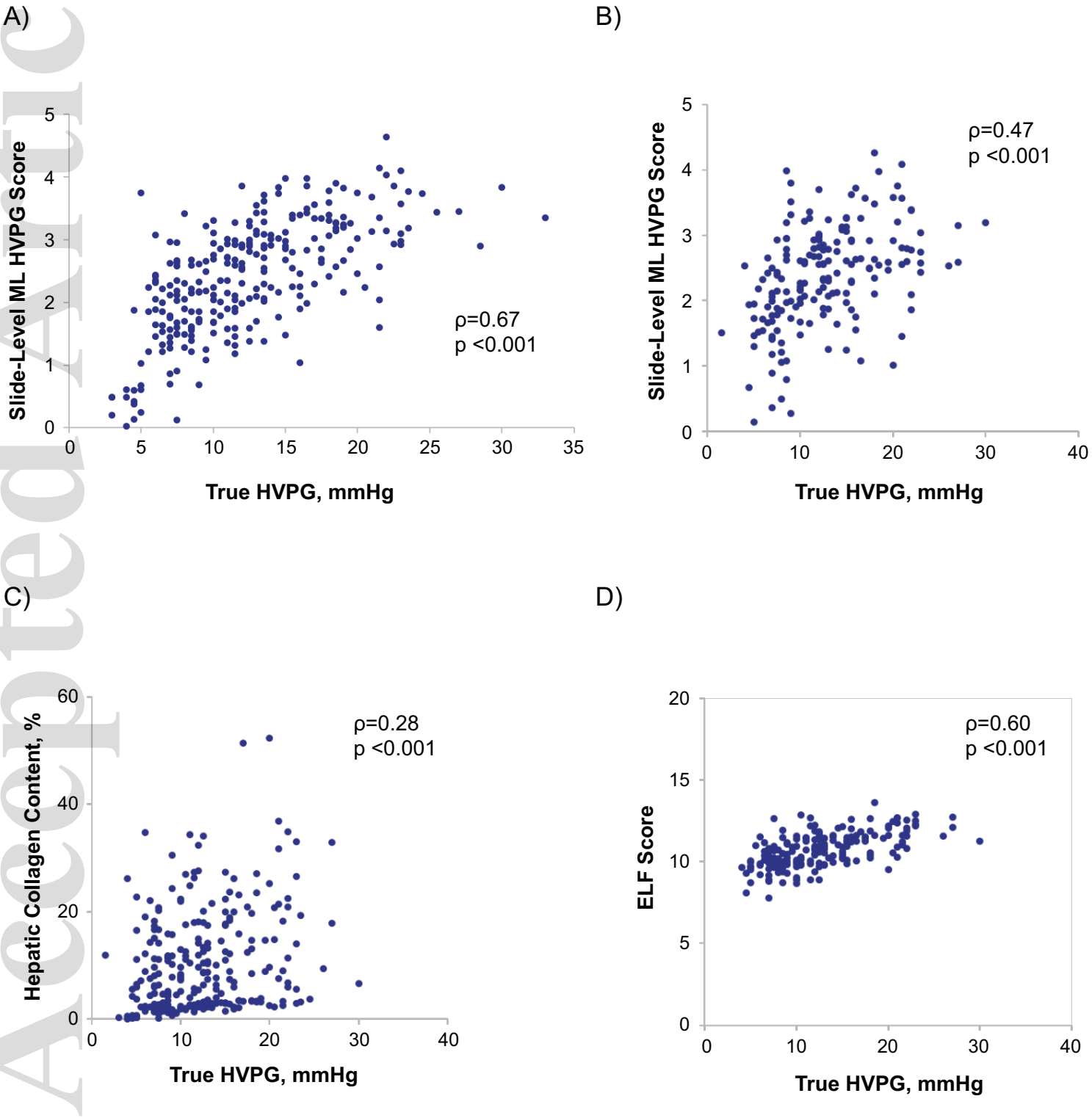
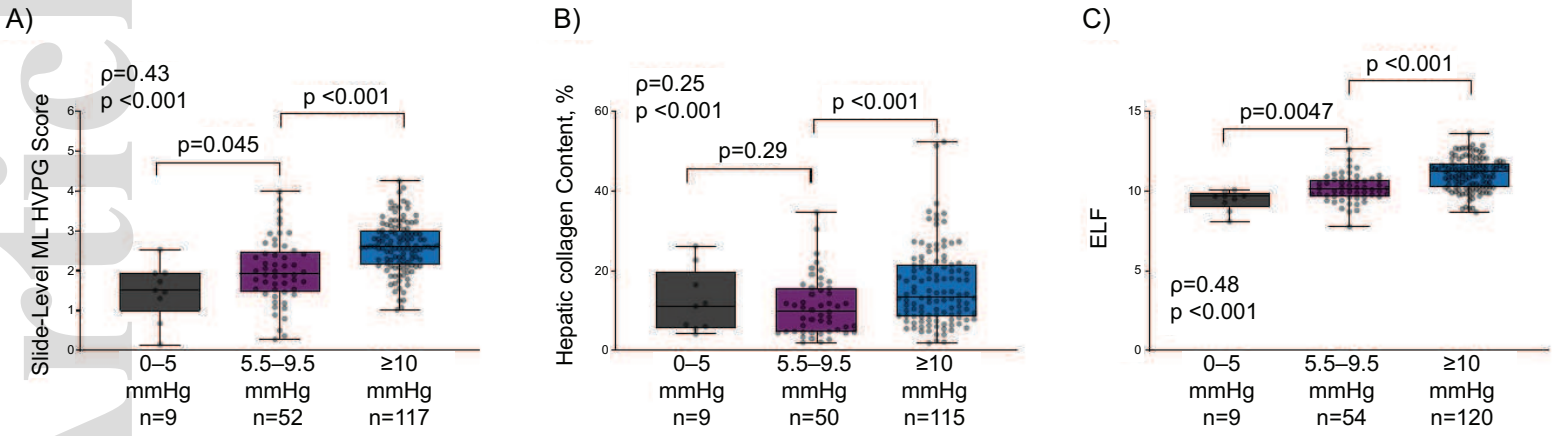


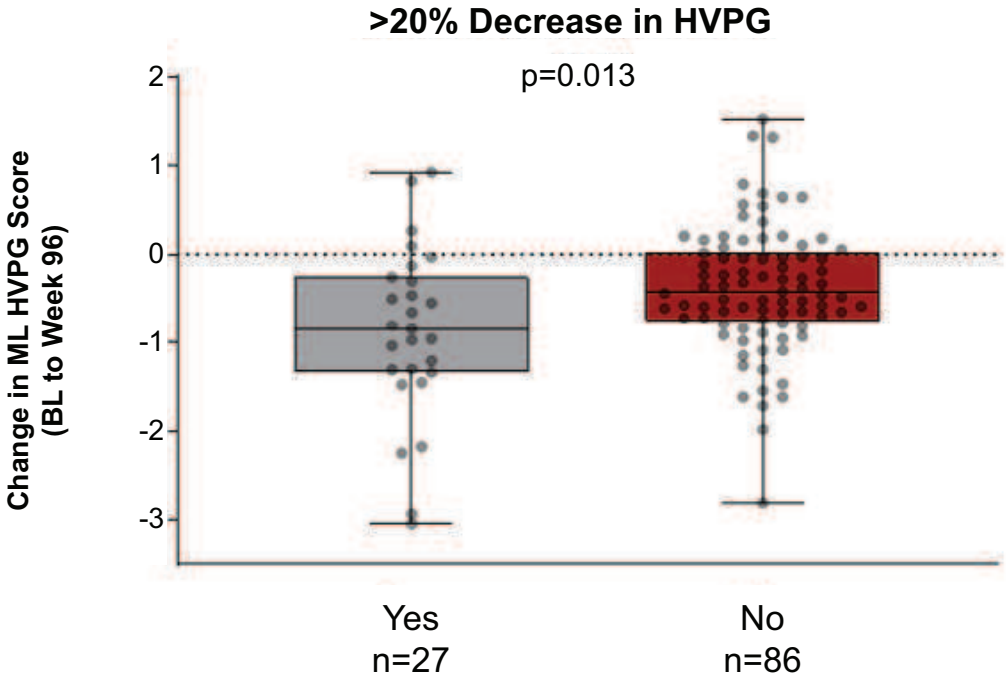
Figure 3. Associations between true HVPG and ML HVPG score, hepatic collagen content, and ELF



hep_32087_f3.eps

Figure 4. Associations between changes in the ML HVPG score between baseline and week 96 with hemodynamic response and fibrosis improvement.

A)



B)

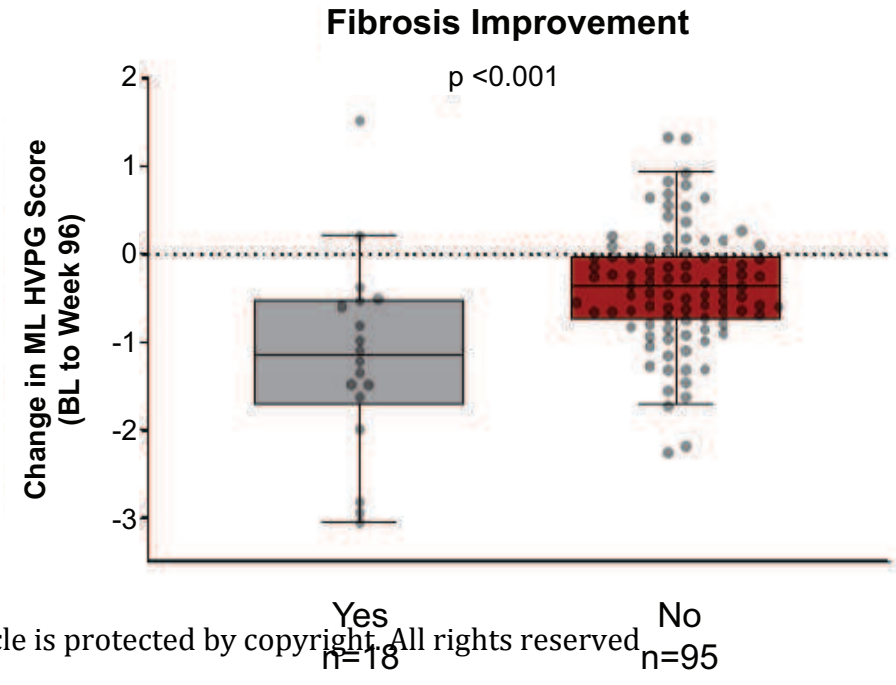
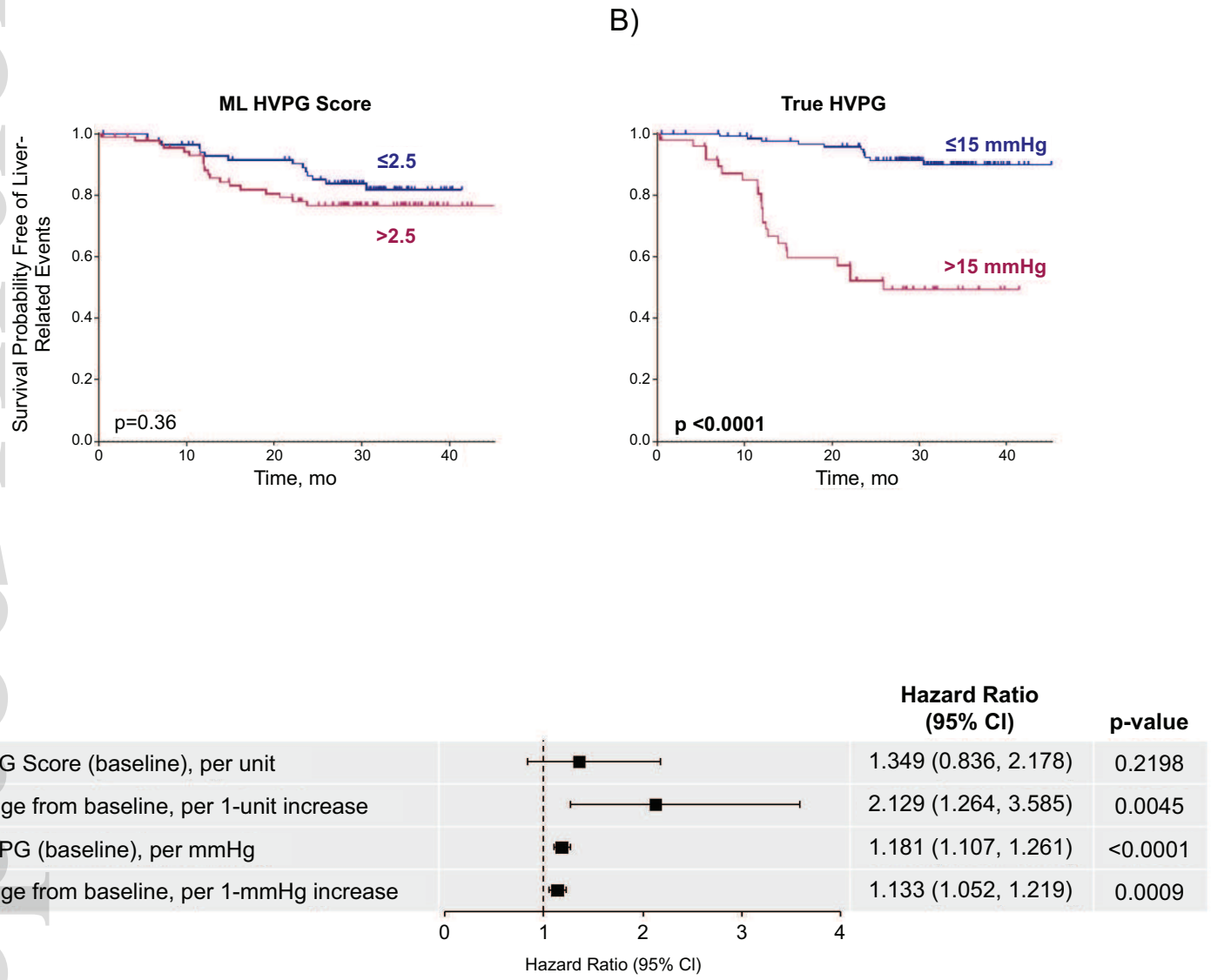


Figure 5. Association between ML HVPG Score and true HVPG with liver-related clinical events



hep_32087_f5.eps