HEPATOLOGY
Official Journal of the American Association for the Study of Liver Diseases

Improvement of Predictive Models of Risk of Disease Progression in Chronic Hepatitis C by Incorporating Longitudinal Data

Monica A. Konerman,[1] Yiwei Zhang,[1] Ji Zhu,[1] Peter D.R. Higgins,[1] Anna S.F. Lok,[1] and Akbar K. Waljee[1,2]

Existing predictive models of risk of disease progression in chronic hepatitis C have limited accuracy. The aim of this study was to improve upon existing models by applying novel statistical methods that incorporate longitudinal data. Patients in the Hepatitis C Antiviral Long-term Treatment Against Cirrhosis trial were analyzed. Outcomes of interest were (1) fibrosis progression (increase of two or more Ishak stages) and (2) liver-related clinical outcomes (liver-related death, hepatic decompensation, hepatocellular carcinoma, liver transplant, or increase in Child-Turcotte-Pugh score to ≥7). Predictors included longitudinal clinical, laboratory, and histologic data. Models were constructed using logistic regression and two machine learning methods (random forest and boosting) to predict an outcome in the next 12 months. The control arm was used as the training data set (n = 349 clinical, n = 184 fibrosis) and the interferon arm, for internal validation. The area under the receiver operating characteristic curve for longitudinal models of fibrosis progression was 0.78 (95% confidence interval [CI] 0.74-0.83) using logistic regression, 0.79 (95% CI 0.77-0.81) using random forest, and 0.79 (95% CI 0.77-0.82) using boosting. The area under the receiver operating characteristic curve for longitudinal models of clinical progression was 0.79 (95% CI 0.77-0.82) using logistic regression, 0.86 (95% CI 0.85-0.87) using random forest, and 0.84 (95% CI 0.82-0.86) using boosting. Longitudinal models outperformed baseline models for both outcomes (P < 0.0001). Longitudinal machine learning models had negative predictive values of 94% for both outcomes. *Conclusions:* Prediction models that incorporate longitudinal data can capture nonlinear disease progression in chronic hepatitis C and thus outperform baseline models. Machine learning methods can capture complex relationships between predictors and outcomes, yielding more accurate predictions; our models can help target costly therapies to patients with the most urgent need, guide the intensity of clinical monitoring required, and provide prognostic information to patients. (HEPATOLOGY 2015;61:1832-1841)

The marked improvement in efficacy and side effect profile of the direct-acting antiviral agents have dramatically altered the approach to treatment decision making for chronic hepatitis C (CHC).[1,2] The availability of short courses of well-tolerated all-oral therapy with sustained virologic response rates >90% has prompted recommendations that all patients with CHC should be considered for treatment. There has simultaneously been a focus on improving hepatitis C viral (HCV) infection outcomes at the public health level. The Centers for Disease Control and Prevention, the Institute of Medicine, and the US Preventive Services Task Force have advocated for HCV screening as well as treatment as a means of

Abbreviations: APRI, aspartate aminotransferase to platelet ratio index; AUROC, area under the receiver operating characteristic curve; CHC, chronic hepatitis C; CI, confidence interval; HALT-C, Hepatitis C Antiviral Long-Term Treatment Against Cirrhosis; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; IFN, interferon; LR, logistic regression; ML, machine learning; NPV, negative predictive value; RF, random forest.

From the [1]Division of Gastroenterology, Department of Internal Medicine, University of Michigan Health System, Ann Arbor, MI; [2]VA Ann Arbor Health Services Research and Development Center of Clinical Management Research, Ann Arbor, MI

Received December 17, 2014; accepted February 10, 2015.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.27750/suppinfo

Supported by National Institutes of Health training grant T32DK062708 (to M.A.K.) and grant R01 GM097117 (to P.D.R.H.) and by a VA Health Services Research and Development CDA-2 Career Development Award (1IK2HX000775, to A.K.W.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the health care centers, the National Institutes of Health, or the Veterans Administration.

disease prevention.[3-5] The high prevalence of CHC in the United States paired with the high cost of direct-acting antiviral agents has created notable logistical and financial barriers to universal treatment of patients with CHC. The barriers are even more pronounced in resource-limited countries, many of which have a much higher prevalence of CHC than in Western countries.[6]

If clinicians were better able to predict which patients are at the highest risk for disease progression, these costly therapies could be targeted to patients who have the most urgent need for treatment. Risk prediction models for disease progression would also provide clinicians with valuable information to help guide the intensity of clinical monitoring required and meaningful prognostic information irrespective of treatment decision making. Most published predictive models for disease progression in CHC are based on data on a few variables collected at baseline, with a small number of models incorporating selected data at a single follow-up time point.[7] These rigid models do not mirror clinical practice where assessments of risk of disease progression incorporate a patient's test results over time. In addition, models with only baseline variables cannot distinguish between patients with similar initial data but who go on to have distinct disease courses and outcomes. As such, the aim of this study was to improve upon existing models by incorporating longitudinal data that capture the nonlinear nature of disease progression in CHC. Data from the Hepatitis C Antiviral Long-term Treatment Against Cirrhosis (HALT-C) trial were used for this purpose. We believe that our approach is applicable to other areas of medicine as most chronic diseases do not progress at a linear rate, and it is important for physicians to be able to utilize longitudinal data to refine prognostication as they follow patients so that they can adapt the management plan.

Patients and Methods

Study Population and Data Collection. The design of the HALT-C trial has been described in detail previously.[8] To briefly summarize, the trial enrolled patients with CHC with Ishak fibrosis score ≥ 3 on liver biopsy and prior nonresponse to interferon (IFN) therapies. Patients with a prior history of hepatic decompensation or hepatocellular carcinoma (HCC) were excluded. Patients were randomized to maintenance therapy with pegylated-IFN or to no treatment for the next 3.5 years. Following completion of the randomized phase, patients were followed without treatment until October 2009. For this analysis, we included patients randomized to no treatment in the training set. This selection criterion was chosen given that IFN therapy can have an effect on laboratory results, which in turn may impact their predictive value. Liver biopsies were performed at baseline and repeated at 1.5 and 3.5 years. All biopsy specimens were reviewed for fibrosis, inflammation, steatosis, and iron by a panel of hepatic pathologists. Patients were seen every 3 months during the randomized phase of the trial and every 6 months thereafter. During each visit blood tests were performed and patients were assessed for clinical outcomes.

Definition of Outcomes. Outcomes of interest included (1) histologic progression and (2) liver-related clinical outcomes. Histologic progression was defined as an increase of two or more stages in Ishak fibrosis score from baseline liver biopsy. Any patient with Ishak >4 at baseline was excluded from this part of the analysis. Liver-related clinical outcomes included any of the following: liver-related death, hepatic decompensation (variceal bleeding, ascites, spontaneous bacterial peritonitis, or hepatic encephalopathy), HCC or presumed HCC, liver transplant, or increase in Child-Turcotte-Pugh score to ≥ 7 points on two consecutive time points 3 months apart.[8] Diagnostic criteria were established for each clinical outcome, and an outcomes review panel adjudicated each outcome report per the HALT-C study protocol. Only the first clinical outcome for each patient was included in the analysis.

Predictor Variables. A detailed description of the variables assessed is provided in Table 1. Predictors evaluated included demographics, viral characteristics, clinical characteristics (including relevant comorbidities), laboratory test results, and histology. In order to capture the extensive longitudinal data, for each predictor we created five variables: mean, maximum, mean of differential, maximum of differential, and

Address reprint requests to: Monica A. Konerman, M.D., M.Sc., University of Michigan Health System, Division of Gastroenterology, 3912 Taubman Center, 1500 East Medical Center Drive, SPC 5362; Ann Arbor, MI 48109. E-mail: konerman@med.umich.edu; tel: +1-734-936-0241; fax: +1-734-763-7834.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.27750

Potential conflict of interest: Dr. Lok received grants from AbbVie, Bristol-Myers Squibb, Idenix, and Gilead.

Table 1. Predictor Variables Assessed

Comprehensive Model	
Baseline variables	Demographics: Age, gender, race
	Viral characteristics: HCV genotype, IL28B genotype, HCV RNA, prior HCV treatment regimens, estimated duration of HCV infection
	Clinical characteristics: alcohol use (lifetime drinks and current use), tobacco use, BMI, waist circumference, history of diabetes, presence and grade of esophageal varices on upper endoscopy, beta-blocker use, antihypertensive use, evidence of portal hypertension
	Labs: WBC with differential, hemoglobin, platelets, AST, ALT, AST/ALT, total bilirubin, albumin, alkaline phosphatase, APRI, AFP, INR, MELD, creatinine, BUN, glucose, triglycerides, insulin, HOMA2 IR, iron level, iron saturation, total iron binding capacity, ferritin
	Histology: Ishak score, histologic activity index, steatosis score, biopsy length, biopsy fragmentation, iron score
Longitudinal variables	Viral characteristics: HCV RNA
	Clinical characteristics: BMI
	Labs: WBC with differential, hemoglobin, platelets, AST, ALT, AST/ALT, alkaline phosphatase, total bilirubin, albumin, INR, AFP, APRI, MELD, CTP score (for fibrosis progression model only), BUN, creatinine, eGFR, urinary protein, glucose, triglycerides, iron, total iron binding capacity, ferritin,
	Histology: Ishak score, histologic activity index, steatosis score, biopsy length, biopsy fragmentation, iron score
Condensed Model	
Baseline variables	Demographics: Age, gender, race
	Viral characteristics: HCV genotype, HCV RNA
	Clinical characteristics: BMI, history of diabetes
	Labs: WBC, hemoglobin, platelets, AST, ALT, AST/ALT, total bilirubin, albumin, alkaline phosphatase, APRI, AFP, INR, MELD, creatinine, BUN, glucose
Longitudinal variables	Clinical characteristics: BMI
	Labs: WBC, hemoglobin, platelets, BUN, creatinine, glucose, AST, ALT, AST/ALT, alkaline phosphatase, total bilirubin, albumin, INR, AFP, APRI, MELD, CTP score (for fibrosis progression model only)

Abbreviations: AFP, alpha-fetoprotein; ALT, alanine aminotransferase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; BUN, blood urea nitrogen; CTP, Child-Turcotte-Pugh; eGFR, estimated glomerular filtration rate; HCV, hepatitis C virus; HOMA2 IR, homeostatic model assessment of insulin resistance; IL, interleukin; INR, international normalized ratio; MELD, model of end-stage liver disease; WBC, white blood cell count.

mean of acceleration. These variables were defined as follows: mean was defined as the mean of the observed values; maximum was defined as the maximum of the observed values; mean of the differential was defined as the mean of the difference between sequential observed values divided by the sequential observation time; maximum of the differential was defined as the maximum of the difference between sequential observed values divided by the sequential observation time; and mean of acceleration was defined as the mean of the difference between sequential differential observed values divided by the difference between sequential differential observation time ($\Delta x/\Delta t$).Results of all predictors until 12 months prior to time of prediction were included. For fibrosis progression, outcomes could only be assessed at the fixed intervals of year 1.5 and 3.5 when biopsies were obtained per study protocol.

A second condensed clinical outcomes prediction model was also created. The predictor variables included in the condensed clinical model were chosen based on their availability in clinical practice and taking into account the results of the variable importance graphs generated from the comprehensive clinical model and the results of our systematic review of the literature on predictors of clinical outcomes.[7]

Development of Regression Model. We first developed a predictive logistic regression (LR) model for both outcomes within the next 12 months. We generated a model using baseline variables only and a model that included baseline and longitudinal data. Because regression models do not converge when the number of predictors is large, we used a lasso technique to limit the predictor variables to those with the highest predictive value.[9] A 10-fold cross-validation was performed by dividing the data into 10 roughly equal smaller data sets (folds). The model (including variable selection) is then run 10 times, with the data in each fold being held out in each run. The cross-validation was then repeated 50 times to give an estimate of the performance characteristics.

Development of Machine Learning Models. An in-depth description of the machine learning (ML) algorithms and model construction is provided in the Supporting Information. Briefly, we used two ML methods, random forest (RF) analysis and boosting, to build prediction models.[10-12] Boosting and RF are two decision tree–based ensemble statistical methods that can build classification and regression prediction models. Compared to the commonly used predictive models, these two ML methods are able to incorporate many predictor variables without compromising the accuracy of the risk prediction.

In RF, as each decision tree is built, only a random subset of the predictor variables are considered as possible splitters for each binary partitioning. The predictions from each tree are used as "votes" in classification, and the outcome with the most votes is considered the dichotomous outcome prediction for that sample. Using this method, multiple decision trees were constructed to create the final classification prediction model and to determine overall variable importance. Variable importance identifies the most important variables based on their contribution to the predictive accuracy of the model. The most important variables are identified as those that most frequently result in early splitting of the decision trees. Boosting, in comparison to RF, is an iterative process that focuses on the misclassified data such that each tree is based on a weighted average of the data points and the weights are calculated based on the previous model in the iterative process. The ML methods were also validated using a 10-fold cross-validation and 50 replications.

Assessing and Comparing Model Performance and Internal Validation. We compared the performance of the ML models and the classic LR model for both fibrosis progression and clinical outcomes with area under the receiver operating characteristic curve (AUROC) analysis and 95% confidence intervals (CIs). We then compared the longitudinal models with models built on baseline predictors alone for each outcome. We performed internal validation of the longitudinal prediction models using the maintenance pegylated-IFN treatment arm of the HALT-C trial. The receiver operating characteristic curves were used to identify optimal risk cutoffs to maximize the model sensitivity and specificity and define a high-risk and a low-risk group. We assessed the ability of each model to differentiate the risk of fibrosis progression or clinical outcomes among low-risk and high-risk patients. Brier scores, which capture both calibration and discrimination, were also reported as an overall measure of model performance. Brier scores can range from 0 to 1, with lower scores being consistent with more accurate and better model performance. In order to assess the performance of our longitudinal ML model in the setting of missing data as may occur in the clinical setting, we then applied the model using imputation for missing predictors. The MissForest method of imputation for missing laboratory data was used.[13]

All ML methods were performed using the statistical language R (version 3.0.2), with the package randomForest, Adaboost, and gbm (by Y.Z. and J.Z.).[11,12,14] Additional analyses were conducted using STATA statistical software. Two-sided P values <0.05 were considered statistically significant.

Results

Predicting Fibrosis Progression. A total of 274 patients in the no-treatment arm had an Ishak score of <5 on the baseline biopsy and at least one of the two subsequent protocol follow-up liver biopsies. For this analysis, we included 184 patients who did not have any missing data for any of the predictor variables. At baseline biopsy, 22 patients had Ishak fibrosis stage 2, 105 had Ishak stage 3, and 57 had Ishak stage 4. Fifty (27.1%) patients had fibrosis progression. Baseline characteristics of patients who did and those who did not have a $\geq$2-point increase in Ishak score are shown in Table 2. These findings were similar to those of the larger cohort that included patients with missing data (Supporting Table S1).

The AUROC results for the three separate prediction models created using either baseline or longitudinal data to differentiate patients with fibrosis progression are displayed in Fig. 1A. For models with longitudinal data, the AUROCs were 0.78 (95% CI 0.74-0.83) using LR, 0.79 (95% CI 0.77-0.81) using RF, and 0.79 (95% CI 0.77-0.82) using boosting. The difference between the longitudinal AUROCs of the two ML models and the LR model, calculated using the 50-replication approach, were statistically significant ($P = 0.002$ for RF, $P = 0.0006$ for boosting). Each of the three longitudinal models had statistically higher AUROCs than their respective models with baseline data alone ($P < 0.0001$).

The variable importance graph for the RF ML longitudinal model is shown in Fig. 2A. The most important variables in differentiating patients who developed fibrosis progression and those who did not were as follows: mean aspartate aminotransferase, mean and differential mean aspartate aminotransferase to platelet ratio index (APRI), mean alanine aminotransferase, and baseline model of end-stage liver disease score.

Predicting Clinical Outcomes. A total of 533 patients were assessed for clinical outcomes. For this analysis, we included the 349 patients who did not have any missing data for any of the predictor variables. A total of 100 patients (28.6%) met predefined criteria for the combined clinical outcome. Baseline characteristics of those patients who did and those who did not have a clinical outcome are shown in Table 2.

The AUROC results for the three separate prediction models created using baseline or longitudinal data

Table 2. Baseline Characteristics of Patients by Outcome: Training Cohort

	Fibrosis Progression (n = 184)			Clinical Outcome (n = 349)		
Variable	No (n = 134) Mean or %	Yes (n = 50) Mean or %	P	No (n = 249) Mean or %	Yes (n = 100) Mean or %	P
Age (years)	49.6	48.6	0.37	49.2	49.6	0.63
% Female	27.6	38.0	0.17	28.9	27	0.72
Race (% white)	71.6	76.0	0.19	71.9	74	0.15
% HCV genotype 1	92.5	90	0.20	92	91	0.59
Duration of Infection (years)	25.9	26.8	0.49	26.3	28.1	0.06
BMI	29.3	31.5	0.02	29.6	20.6	0.12
Diabetes (%)	13.4	22	0.16	15.2	18	0.53
Alcohol intake/day (g)	28.9	28.0	0.89	27.5	32.6	0.35
Tobacco use (pack-years)	13.9	17.0	0.27	15.3	12.1	0.12
Log HCV RNA (log_{10} IU/mL)	6.5	6.4	0.10	6.5	6.3	0.003
Platelet count (1000/mm^3)	201	173	0.008	185	123	<0.0001
INR	0.99	1.03	0.008	1.02	1.08	<0.0001
AST ratio to ULN*	1.75	2.40	0.009	2.03	2.46	0.01
ALT ratio to ULN*	2.13	2.81	0.05	2.37	2.34	0.88
AST/ALT	0.78	0.81	0.44	0.78	0.97	<0.0001
Alkaline phosphatase ratio to ULN*	0.78	0.85	0.22	0.79	0.95	0.0002
Albumin (g/dL)	3.97	3.87	0.04	3.94	3.67	<0.0001
Total bilirubin (mg/dL)	0.66	0.81	0.008	0.73	0.91	<0.0001
AFP ratio to ULN*	1.01	1.66	0.03	1.17	2.64	<0.0001
MELD	6.5	7.2	0.0001	6.8	7.5	0.0001
APRI	0.99	1.77	0.0004	1.34	2.33	<0.0001
Ishak	3.1	3.3	0.14	3.83	4.14	<0.0001
HAI	7.23	7.06	0.60	7.39	7.47	0.22
Steatosis (0-4)	1.14	1.72	0.0002	1.33	1.38	0.65

*Variable expressed relative to the ULN to account for differences in reference ranges for normal results among different clinical trial sites.

Abbreviations: AFP, alpha-fetoprotein; ALT, alanine aminotransferase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; HAI, histologic activity index; HCV, hepatitis C virus; INR, international normalized ratio; MELD, model of end-stage liver disease; RNA, ribonucleic acid; ULN, upper limit of normal.

to differentiate patients who did or did not develop a clinical outcome are displayed in Fig. 1B. For models with longitudinal data, the AUROCs were 0.79 (95% CI 0.78-0.82) using LR, 0.86 (95% CI 0.85-0.87) using RF, and 0.84 (95% CI 0.82-0.86) using boosting. The ML models had significantly better discriminative accuracy than the LR model for clinical outcomes ($P < 0.0001$). The longitudinal models outperformed the related baseline models for all three methods ($P < 0.0001$).

The variable importance graph for the longitudinal RF ML model in predicting clinical outcomes is shown in Fig. 2B. The most important independent variables in differentiating patients who developed clinical outcomes and those who did not were as follows: mean APRI, maximum baseline and mean platelet count, and mean albumin. To assess whether our models were more accurate at predicting any of the five combined clinical outcomes, additional sensitivity analyses were performed by removing one clinical outcome from the combined clinical outcome at a time. Neither the AUROC nor the variable importance results significantly changed. Of note, removing HCC as one of the combined clinical outcomes did not significantly alter the AUROC or the variable importance (Supporting Fig. S1).

Performance of Prediction Models in the Internal Validation Cohort. Validation of the prediction models was performed using data from the treatment arm of the HALT-C trial. The baseline characteristics of the patients in the treatment arm are displayed in Supporting Table S2. A total of 183 patients in the IFN treatment arm had no missing data for any of the predictor variables and were included in this analysis for histologic and clinical outcomes. In the internal validation cohort 46 (25.1%) patients had fibrosis progression and 31 (17%) had a clinical outcome. The features associated with developing an outcome on univariate analysis in the internal validation cohort were similar, though not identical, to results in the control arm of the HALT-C study (Supporting Table S2).

In the internal validation cohort, the longitudinal fibrosis progression models had the following AUROCs: 0.79 (95% CI 0.71-0.87) using LR, 0.88 (95% CI 0.83-0.93) using RF, and 0.86 (95% CI

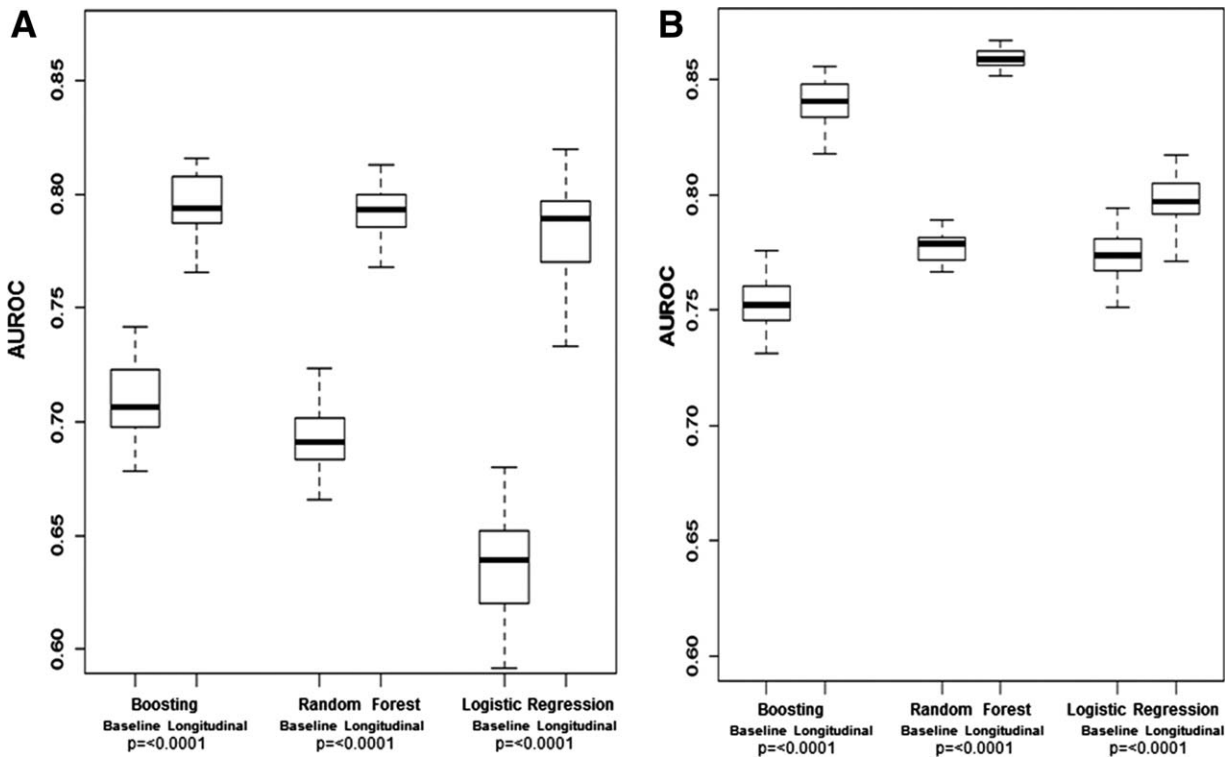

Fig. 1. Areas under the receiver operating characteristic curve in the training cohort for (A) fibrosis progression and (B) clinical outcomes. Abbreviation: AUROC, area under the receiver operating characteristic curve.

0.80-0.91) using boosting (Fig. 3A). The longitudinal predictive models for clinical outcomes had the following AUROCs in the internal validation cohort: 0.76 (95% CI 0.67-0.86) using LR, 0.81 (95% CI 0.73-0.90) using RF, and 0.80 (95% CI 0.70-0.90) using boosting (Fig. 3B). An additional analysis was

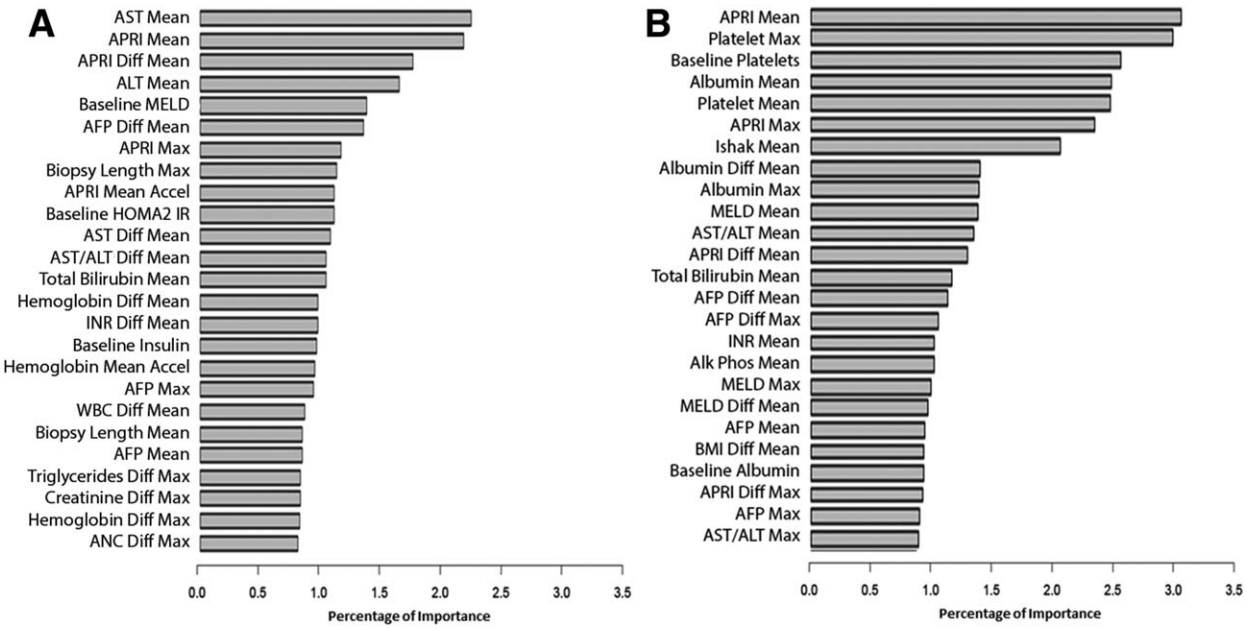

Fig. 2. Longitudinal RF variable importance in the training cohort for (A) fibrosis progression and (B) clinical outcomes. Abbreviations: Accel, acceleration; AFP, alpha-fetoprotein; Alk Phos, alkaline phosphatase; ALT, alanine aminotransferase; ANC, absolute neutrophil count; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; CTP, Child-Turcotte-Pugh; Diff, differential; HOMA2 IR, homeostatic model assessment of insulin resistance; INR, international normalized ratio; MELD, model of end-stage liver disease; WBC, white blood cell count.

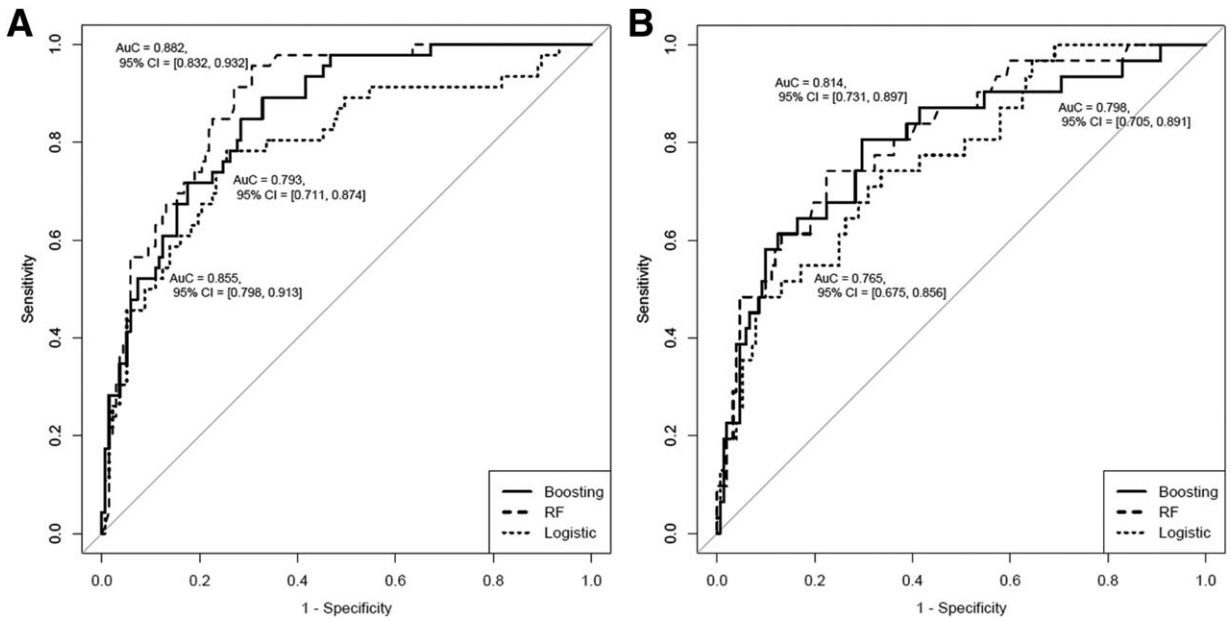

Fig. 3. Area under the receiver operating characteristic curves in the internal validation cohort of (A) longitudinal models for fibrosis progression and (B) longitudinal models for clinical outcomes. Abbreviations: AuC, area under the receiver operating characteristic curve; RF, random forest.

performed using the entire validation cohort including patients with missing data for the predictors which yielded similar results.

The proportion of patients correctly classified as high versus low risk and the associated Brier score are displayed in Table 3 and illustrated in Fig. 4. For fibrosis progression, the ML models were 85% sensitive, 71%-77% specific with a negative predictive value (NPV) of 94%. For clinical outcomes, the ML models had a sensitivity of 74%-81%, a specificity of 70%-78%, and an NPV of 94%.

Performance of the Condensed Clinical Prediction Model. The results of the more condensed clinical prediction model built with only variables routinely available in clinical practice yielded similar results (Fig. 5A). Once again, the longitudinal models outperformed the related baseline models for all three methods ($P < 0.0001$). The variables that contributed most to the predictive accuracy of the condensed model were similar to the comprehensive model and were as follows: mean APRI, maximum mean and baseline platelets, and mean albumin (Supporting Fig. S2). In

Table 3. Misclassification Table for Longitudinal Predictive Models of Fibrosis Progression and Clinical Outcomes: Internal Validation Cohort

Fibrosis Progression								
		Fibrosis Progressors (n = 46)		Fibrosis Nonprogressors (n = 137)				
	Cutoff	Predicted Fibrosis Progression	Predicted No Fibrosis Progression	Predicted Fibrosis Progression	Predicted No Fibrosis Progression	Brier score	NPV	PPV
Random forest	0.353	39 (84.8%)	7 (15.2%)	31 (22.6%)	106 (77.4%)	0.208	93.8%	55.7%
Boosting	−10.47	39 (84.8%)	7 (15.2%)	39 (28.5%)	98 (71.5%)	0.251	93.3%	50.0%
Logistic regression	−1.19	36 (78.3%)	10 (21.7%)	35 (25.5%)	102 (74.5%)	0.246	91.1%	50.7%

Clinical Outcomes								
		Clinical Progressors (n = 31)		Clinical Nonprogressors (n = 152)				
	Cutoff	Predicted Clinical Progression	Predicted No Clinical Progression	Predicted Clinical Progression	Predicted No Clinical Progression	Brier score	NPV	PPV
Random forest	0.291	23 (74.2%)	8 (25.8%)	34 (22.4%)	118 (77.6%)	0.230	93.7%	40.4%
Boosting	−12.29	25 (80.7%)	6 (19.3%)	45 (29.6%)	107 (70.4%)	0.279	94.7%	35.7%
Logistic regression	−1.77	23 (74.2%)	8 (25.8%)	51 (33.6%)	101 (66.4%)	0.322	92.7%	31.1%

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

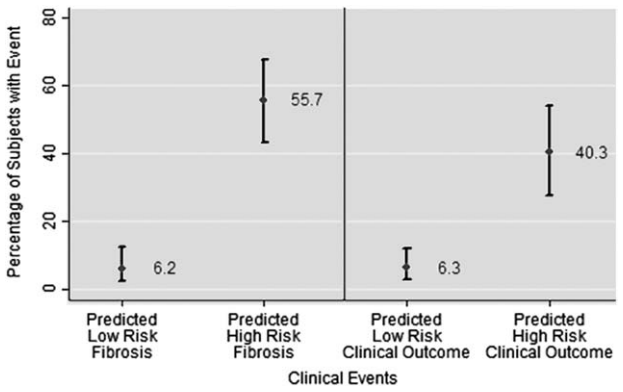

Fig. 4. Outcome incidence by risk strata: internal validation cohort.

the internal validation cohort, the results of the condensed longitudinal clinical progression models were essentially unchanged compared to the more comprehensive models (Fig. 5B). The proportion of patients correctly classified as high versus low risk were also very similar though slightly less accurate compared to the original comprehensive model. For clinical outcomes, the condensed longitudinal ML models had a sensitivity of 76%-78%, a specificity of 66%-70%, and an NPV that remained high at 94% (Supporting Table S3).

Discussion

Recent advances in the treatment of CHC have revolutionized the approach to treatment decision making and reinvigorated the public health initiatives to identify patients with CHC. The pool of potential treatment candidates is expected to continue to expand, and the economic impact of these highly efficacious but extremely costly therapies could potentially cripple health care budgets. With over 3.2 million of the US population estimated to have CHC and a single 12-week course of therapy with sofosbuvir priced at $84,000, universal treatment would cost over $268 billion, not accounting for the cost of other medications and any other associated costs.[15] In this context, our data related to improving prediction models of disease progression for patients with CHC provide clinically relevant and valuable tools. These models can help target HCV therapies to patients with the most urgent need for treatment until such time that logistic and financial solutions allow universal treatment. Our model also provides important prognostic information that can help inform patients and tailor the intensity of clinical monitoring required.

In this study, we demonstrated that prediction models that incorporate longitudinal data outperform models restricted to baseline data alone. Moreover, we demonstrated that ML techniques can overcome limitations of the classic forms of statistical analyses by virtue of their ability to incorporate large numbers of predictor variables without compromising the accuracy of the risk prediction. For fibrosis progression, the AUROCs of our longitudinal models of 0.86-0.88 were notably higher than those in prior studies of 0.66.[16] Our ML longitudinal prediction models also yielded very high NPVs of 94%; thus, very few patients classified as low risk of fibrosis progression ultimately developed an outcome. Our findings confirm the utility of liver enzymes and other noninvasive markers of liver fibrosis, specifically APRI, particularly when the results of these tests are used in aggregate.[17-21] From a clinical practice and health policy standpoint the results of our clinical outcome prediction models are even more relevant. Our models were able to accurately discriminate high- versus low-risk patients with a sensitivity of 74%, a specificity of 78%, and an NPV of 94%. As expected, the variables that contributed most importantly to the predictive capability of the model were longitudinal laboratory markers of advanced liver disease including changes in platelet count, APRI, and albumin. Of interest, when removing HCC/presumed HCC from the composite clinical outcomes, neither the AUROC nor the variable importance significantly changed. This is somewhat surprising given that other studies have identified different predictors for hepatic decompensation and HCC.[22,23]

The major strength of our study is the application of novel statistical approaches to analyze longitudinal data, which improved the accuracy of prediction estimates; however, there are several limitations to our findings. These stem from the constraints on the generalizability of our results given the enrollment criteria for the HALT-C study, which only enrolled patients with advanced fibrosis and prior HCV treatment failure. Moreover, the HALT-C cohort was primarily composed of middle-aged Caucasian men with genotype 1 infection and, thus, represents only a portion of the overall population of patients with CHC. Future studies would benefit from evaluating cohorts that include more diverse ranges of baseline liver disease, demographic characteristics, as well as other HCV genotypes. In addition, our endpoint of interest was a composite of liver-related clinical outcomes, and our models may not be as accurate for prediction of individual outcomes. Sensitivity analyses did show that our models performed equally well when each outcome was removed one at a time.

In conclusion, our findings build upon the existing tools by providing novel approaches to analyze

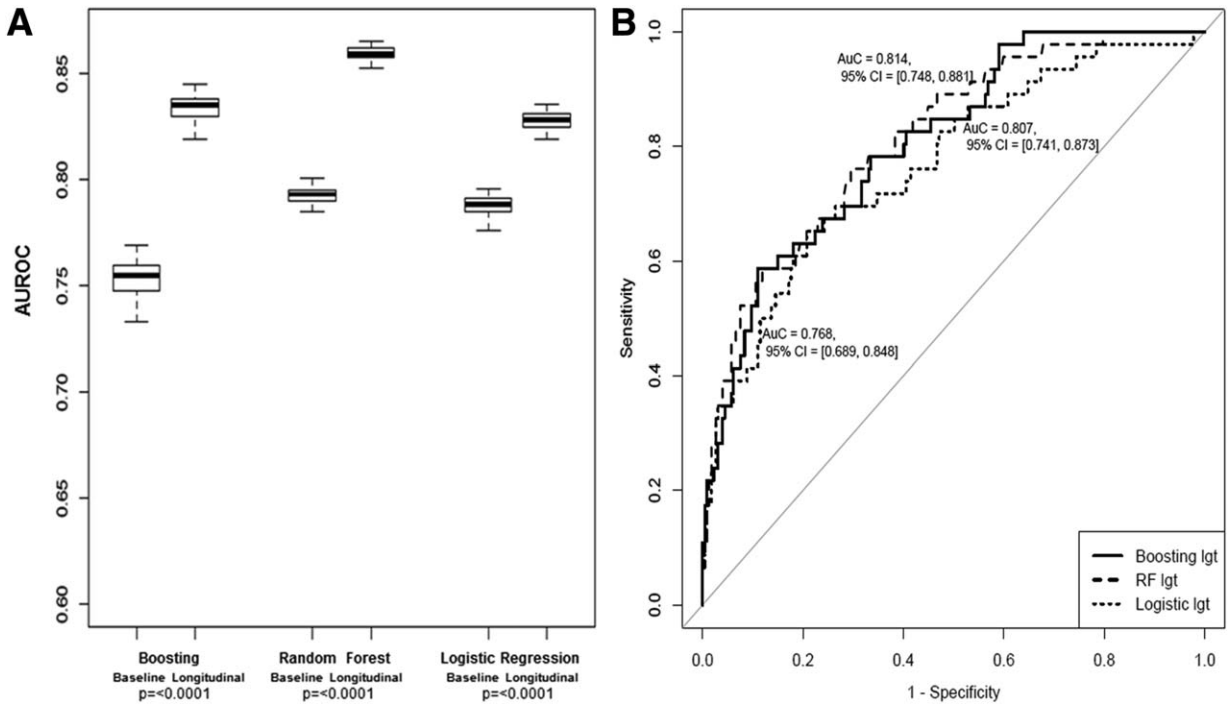

Fig. 5. Area under the receiver operating characteristic curves for (A) clinical outcomes in training cohort: condensed model and (B) condensed clinical outcomes model: internal validation cohort. Abbreviations: AUROC/AuC, area under the receiver operating characteristic curve; RF, random forest.

individual patient results over time in order to more accurately assess the risk of disease progression from CHC. Machine learning methods of analysis have long been successfully applied in other fields such as business and marketing and, as demonstrated here, provide a significant opportunity for application in clinical settings.[24] In our proposed models, we demonstrate that accurate risk predictions can be made based on data routinely available in clinical practice. In its present form, our model can easily be implemented into existing electronic medical records as a clinical decision tool. Developing our model as a universally accessible web-based tool would further increase its accessibility and uptake in clinical practice and is an ultimate goal from an implementation standpoint. Similar to our prediction models in inflammatory bowel disease, we anticipate a tool which would pull data from an individual patient's electronic medical records or a web-based platform where physicians will input and store serial laboratory results from individual patients and an update of the prediction of high or low risk for an outcome in the next 12 months can be run at each clinic visit.[25] This result can then be discussed with patients by the clinician to help inform decisions regarding treatment initiation and the intensity of clinical monitoring (such as frequency of clinic visits and outpatient testing). In the current era of highly efficacious therapy for CHC, ideally we would treat all patients who do not have an absolute contraindication to therapy. Unfortunately, until society can solve the logistic and financial barriers, clinicians and policy makers are faced with the arduous task of trying to target these therapies to patients with the most urgent need. Herein we illustrate that it is possible to create predictive models of risk of disease progression that accurately identify those patients at highest risk for adverse outcomes. Offering immediate treatment to patients identified as high risk for clinical outcomes would reduce the immediate cost burden of HCV treatment without jeopardizing the outcomes of other patients as long as they continue to be monitored and risk assessments are updated at each clinic visit. Future studies are needed to externally validate our results in broader patient populations. Ultimately, we hope treatment will be affordable and accessible to all patients with CHC.

References

1. Lawitz E, Mangia A, Wyles D, Rodriguez-Torres M, Hassanein T, Gordon SC, et al. Sofosbuvir for previously untreated chronic hepatitis C infection. N Engl J Med 2013;368:1878-1887.
2. Olysio (simeprevir): Full prescribing information. Titusville, NJ: Janssen Therapeutics; 2013.

3. Colvin HM, Mitchell AE, eds.; Committee on the Prevention and Control of Viral Hepatitis Infections, Institute of Medicine. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C. Washington, DC: National Academies Press; 2010.
4. Moyer VA. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2013;159:349-357.
5. Smith BD, Morgan RL, Beckett GA, Falck-Ytter Y, Holtzman D, Teo CG, et al. Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945-1965. MMWR Recomm Rep 2012;61:1-32.
6. Lavanchy D. Evolving epidemiology of hepatitis C virus. Clin Microbiol Infect 2011;17:107-115.
7. Konerman MA, Yapali S, Lok AS. Systematic Review: Identifying patients in need of early treatment and intensive monitoring-predictors and predictive models of disease progression in chronic hepatitis C. Aliment Pharmacol Ther 2014;40:863-879.
8. Lee WM, Dienstag JL, Lindsay KL, Lok AS, Bonkovsky HL, Shiffman ML, et al. Evolution of the HALT-C Trial: pegylated interferon as maintenance therapy for chronic hepatitis C in previous interferon nonresponders. Control Clin Trials 2004;25:472-492.
9. Tibshirani R. Regression shrinkage and selection via the lasso. J R Statist Soc 1996;58:267-288.
10. Friedman J HT, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat 1998;28:2000.
11. Liaw A, Wiener M. Classification and regression by random forest. R News 2002;2:18-22.
12. Schapire RE. The boosting approach to machine learning: an overview. In Denison DD, ed. Nonlinear Estimation and Classification. Lecture Notes in Statistics, vol. 171. New York: Springer Science+Business Media; 2003:149-171.
13. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013;3.
14. Brieiman L. Random forests. Machine Learning 2001;45:5-32.
15. U.S. Food and Drug Administration approves Gilead's Sovaldi (sofosbuvir) for the treatment of chronic hepatitis C [press release]. Foster City, CA: Gilead; December 6, 2013. Available at: http://www.gilead.com/news/press-releases/2013/12/us-food-and-drug-administration-approves-gileads-sovaldi-sofosbuvir-for-the-treatment-of-chronic-hepatitis-c.
16. Fontana RJ, Dienstag JL, Bonkovsky HL, Sterling RK, Naishadham D, Goodman ZD, et al. Serum fibrosis markers are associated with liver disease progression in non-responder patients with chronic hepatitis C. Gut 2010;59:1401-1409.
17. Baran B, Gulluoglu M, Soyer OM, Ormeci AC, Gokturk S, Evirgen S, et al. Treatment failure may lead to accelerated fibrosis progression in patients with chronic hepatitis C. J Viral Hepat 2014;21:111-120.
18. Ghany MG, Kleiner DE, Alter H, Doo E, Khokar F, Promrat K, et al. Progression of fibrosis in chronic hepatitis C. Gastroenterology 2003;124:97-104.
19. Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Komatsu N, et al. The presence of steatosis and elevation of alanine aminotransferase levels are associated with fibrosis progression in chronic hepatitis C with non-response to interferon therapy. J Hepatol 2008;48:736-742.
20. Mummadi RR, Petersen JR, Xiao SY, Snyder N. Role of simple biomarkers in predicting fibrosis progression in HCV infection. World J Gastroenterol 2010;16:5710-5715.
21. Williams MJ, Lang-Lenton M. Progression of initially mild hepatic fibrosis in patients with chronic hepatitis C infection. J Viral Hepat 2011;18:17-22.
22. Kwon H, Lok AS. Does antiviral therapy prevent hepatocellular carcinoma? Antivir Ther 2011;16:787-795.
23. Lok AS, Everhart JE, Wright EC, Di Bisceglie AM, Kim HY, Sterling RK, et al. Maintenence peginterferon therapy and other factors associated with hepatocellular carcinoma in patients with advanced hepatitis C. Gastroenterology 2011;140:840-849.
24. Waljee AK, Higgins PD. Machine learning in medicine: a primer for physicians. Am J Gastroenterol 2010;105:1224-1226.
25. Waljee AK, Joyce JC, Wang S, Saxena A, Hart M, Zhu J, et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. Clin Gastroenterol Hepatol 2010;8:143-150.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.27750/suppinfo