Published in final edited form as:

J Viral Hepat. 2016 June ; 23(6): 455–463. doi:10.1111/jvh.12509.

Dynamic prediction of risk of liver-related outcomes in chronic hepatitis C using routinely collected data

M. A. Konerman¹, M. Brown², Y. Zheng², and A. S. F Lok¹

¹Department of Internal Medicine, Division of Gastroenterology, University of Michigan Health System, Ann Arbor, MI, USA

²Fred Hutchinson Cancer Research Center, Seattle, WA, USA

SUMMARY

Accuracy of risk assessments for clinical outcomes in patients with chronic liver disease has been limited given the nonlinear nature of disease progression. Longitudinal prediction models may more accurately capture this dynamic risk. The aim of this study was to construct accurate models of short- and long-term risk of disease progression in patients with chronic hepatitis C by incorporating longitudinal clinical data. Data from the Hepatitis C Antiviral Long-term Treatment Against Cirrhosis trial were analysed ($n = 533$ training cohort; $n = 517$ validation cohort). Outcomes included a composite liver outcome (liver-related death, decompensation, hepatocellular carcinoma (HCC) or liver transplant), decompensation, HCC and overall mortality. Longitudinal models were constructed for risk of outcomes at 1, 3 and 5 years and compared with models using data at baseline only or baseline and a single follow-up time point. A total of 25.1% of patients in the training and 20.8% in the validation cohort had an outcome during a median follow-up of 6.5 years (range 0.5–9.2). The most important predictors were as follows: albumin, aspartate aminotransferase/alanine aminotransferase ratio, bilirubin, alpha-fetoprotein and platelets. Longitudinal models outperformed baseline models with higher true-positive rates and negative predictive values. The areas under the receiver-operating characteristic curve for the composite longitudinal model were 0.89 (0.80–0.96), 0.83 (0.76–0.88) and 0.81 (0.75–0.87) for 1-, 3-, and 5-year risk prediction, respectively. Model performance was retained for decompensation and overall mortality but not HCC. Longitudinal prediction models provide accurate risk assessments and identify patients in need of intensive monitoring and care.

Keywords

antiviral therapy; cirrhosis; hepatic decompensation; hepatocellular carcinoma

Correspondence: Monica A. Konerman, MD MSc, Division of Gastroenterology, University of Michigan Health System, 3912 Taubman Center, 1500 East Medical Center Drive, SPC 5362; Ann Arbor, Michigan 48109, USA. konerman@med.umich.edu.

DISCLOSURES

No personal interest relevant to this study. This study was funded in part by the National Institutes of Health T32DK062708 training grant (MAK).

INTRODUCTION

Chronic liver disease (CLD) represents the 12th leading cause of death in the United States and is estimated to cost \$1.6 billion annually in healthcare costs and lost productivity [1-3]. The clinical course and complications of CLD can vary widely across individuals. This has made risk stratification for adverse outcomes difficult thus limiting our ability to tailor management. Prediction models for risk of adverse clinical outcomes have traditionally been restricted to baseline data given the limitations of classical forms of statistical analysis [4]. Models that incorporate longitudinal data may yield more accurate risk assessments by capturing serial results that reflect the dynamic nature of disease progression [5].

The availability of highly efficacious and well-tolerated short courses of all oral therapy for chronic hepatitis C (CHC) has put it at the forefront of CLD management [6-9]. The high prevalence of CHC both nationally (>3 million persons in the United States) and globally (150 million persons worldwide) amplifies the impact of advances in care for this patient population [10,11]. Despite these revolutionary advances, logistical and financial barriers presently preclude universal access to care and treatment [12]. As a result, many payers have instituted policies to determine which patients will have their treatment covered now and which patients will have their treatment deferred. The criteria used to determine urgency of treatment vary among payers but most have relied on assessment of fibrosis stage at a single time point [13]. Given the nonlinear progression of CHC and the limitations of noninvasive assessment of liver fibrosis, accurate predictive models of risk of clinical progression would help guide decisions on how to target limited resources to patients with the highest risk of adverse clinical outcomes in the near future. These models would also provide prognostic information to patients and clinicians.

The Hepatitis C Antiviral Long-term Treatment Against Cirrhosis (HALT-C) trial provides a robust longitudinal cohort of patients with serial data points and adjudicated clinical outcomes amenable to constructing and testing the performance characteristics of prediction models for risk of clinical disease progression. We have previously applied machine learning algorithms to construct longitudinal models to predict risk of developing a composite liverrelated clinical outcome in the next 12 months and demonstrated their superiority compared to baseline models [5]. In this study, we aimed to develop accurate predictive models of risk of disease progression at longer time intervals (3 and 5 years) using only data routinely obtained in clinical practice as these models would have more clinical applicability. Additionally, we investigated the predictive capability for multiple outcomes including a composite liver outcome, hepatic decompensation alone, HCC alone and overall mortality. In doing so, we aim to demonstrate the advantages of longitudinal models in predicting both population and individual level risk of adverse outcomes such that similar methods can be applied to other forms of CLD that also have dynamic and heterogeneous rates of progression.

PATIENTS AND METHODS

Study Population and Data Collection

We used data from the Hepatitis C Antiviral Long-term Treatment Against Cirrhosis (HALT-C) trial for this study. The design of the HALT-C trial has been described in detail previously [14]. To briefly summarize, the trial enrolled patients with CHC with Ishak fibrosis score ≥ 3 and prior nonresponse to interferon-based (IFN) therapies. Patients with a prior history of hepatic decompensation or HCC were excluded. Patients were treated with full dose pegylated- IFN and ribavirin during the lead in phase of the trial. Patients with virological breakthrough or relapse and nonresponders were then randomized to maintenance therapy (pegylated-IFN alfa-2a 90ug weekly) or no treatment for the next 3.5 years. Following the randomized phase, patients were followed without treatment until October 2009. For this analysis, we included patients randomized to no treatment in the training cohort. This selection criterion was decided upon given that IFN therapy can have an effect on laboratory results which in turn may impact their predictive value. The IFN treatment arm of the study served as the internal validation cohort. Patients were seen every 3 months during the randomized phase of the trial and every 6 months thereafter. Liver biopsies were performed at baseline and repeated at 1.5 and 3.5 years. All biopsy specimens were reviewed for fibrosis (Ishak stage 0– 6), inflammation (histologic activity index 0–18) and steatosis (scored 0–4) by a panel of hepatic pathologists. During each visit, blood tests were performed and patients were assessed for clinical outcomes.

Predictors and Outcomes

Candidate variables of interest were selected based on results of prior studies aimed at identifying predictors of outcomes in patients with CHC [4]. Predictors of interest included demographics, viral characteristics (HCV genotype and HCV RNA) and clinical characteristics including body mass index (BMI), history of diabetes, alcohol use and estimated duration of HCV infection. Longitudinal predictive variables included: complete blood cell count, liver panel, alpha-fetoprotein (AFP), international normalized ratio (INR), model for end-stage liver disease (MELD) score, Child–Turcotte–Pugh (CTP) score, aspartate aminotransferase (AST) to platelet ratio index (APRI) and AST/alanine aminotransferase (ALT) ratio. Histologic data included Ishak fibrosis score, histologic activity index and steatosis score.

The primary outcome of interest was the time to a composite liver-related clinical outcome. This included liverrelated death, hepatic decompensation (variceal bleeding, ascites or hepatic encephalopathy), HCC or presumed HCC, or liver transplant [14]. Patients characterized as having presumed HCC met criteria similar to the currently accepted definition outlined in the American Association for the Study of Liver Diseases guidelines for nonhistologic diagnosis of HCC [15]. In the HALT-C trial, criteria for diagnosis of clinical outcomes were predefined and adjudicated by the study's clinical review panel. Only the first clinical outcome for each patient was included in the analysis. Individuals without an outcome were censored at the time of their last follow-up visit. In addition to the composite clinical outcome, we also constructed predictive models for HCC alone, hepatic decompensation alone and overall mortality. For overall mortality, we also included patients

who underwent liver transplantation given that in the absence of transplantation the patient may have died.

Development of Prediction Models

Descriptive statistics were used to report the characteristics of the patients in the study. Kaplan–Meier (KM) curves were used to present the event rates of various outcomes. To identify predictor variables that were independently associated with our outcome of interest, we performed univariate and multivariate Cox proportional hazards model and calculated hazards ratios (HRs) and corresponding 95% confidence intervals (CI). Throughout, robust variance was considered to account for correlations among repeated measurements from the same individual. Two-sided P values <0.05 were considered statistically significant.

Predictive models were constructed using the training cohort in two steps: first, for serial measures, we modelled each biomarker's trajectory through time using a mixed effect model with random and fixed effects for measurement time. This also allowed us to reduce random measurement error in observed values and accounted for correlations among measurements within an individual. Next, we fitted a partly conditional survival model to model the residual time to event from each of the clinical visit times. Covariates in the model include the fitted longitudinal biomarkers at the visit time, the length of clinical monitoring and other clinical information [16,17]. Baseline covariates (age, gender, race, history of diabetes and alcohol use) were not significantly associated with the outcome and were left out of the multivariate model. Using the final prediction model, we could then calculate the risk of an adverse outcome at a future time point of interest (prediction time) using all the data accumulated up to the time when a prediction is made (follow-up time). For comparison, a baseline prediction model using the same variables in the longitudinal model was developed using an ordinary multivariate proportional hazard regression model, with time from baseline to event/censoring time as the outcome.

Assessing and Comparing Model Performance

An array of metrics was considered to evaluate the model performance. Model calibration was evaluated with prediction errors by comparing the model-based predictive probabilities with the empirical estimates from the data. We calculated time-dependent receiver-operating characteristic (ROC) curves for longitudinal data and used the area under the ROC curve (AUROC) as a global measure of the discriminatory capacity of the model for separating individuals who experienced outcomes by 1, 3 or 5 years (cases) versus those who did not (controls) at different follow-up times [16,18]. As medical decision-making is often based on whether an individual's risk exceeds a preselected risk threshold, we calculated true- and false-positive rates (TPR and FPR), and negative and positive predictive values (NPV and PPV) at specific risk thresholds (0.1 for 1-year prediction, 0.25 for 3-year prediction and 0.5 for 5-year prediction, corresponding to high-risk thresholds at these times). Recognizing that such thresholds may vary by provider, we plotted risk distributions of cases (TPR) and controls (FPR) at subsequent 1, 3 or 5 years, respectively, over all possible risk thresholds at specific follow-up times. A superior prediction rule will select higher proportions of cases but a lower proportion of controls for treatment or more intensive monitoring across all risk thresholds.

We also compared the baseline and longitudinal models to a model using baseline plus a single follow-up time point (Year 2) previously developed by Ghany *et al.* [19] To make the results comparable across different models, we presented the performance of our longitudinal model for predicting outcomes in subsequent 1-, 3- and 5-year among individuals with 2 years of follow-up data.

We made model assessments with both the training set and the validation set. When calculating performance measures in the training set, cross-validation was considered to correct for potential bias due to using the training set for both model building and performing internal validation. IFN treatment arm data were used for an independent external assessment of the model performance.

RESULTS

Baseline Characteristics and Incidence of Outcomes

The baseline characteristics of patients are displayed in Table 1. The cohorts consisted primarily of middle-aged (50 and 51 years) Caucasian (71–72%) men (70–72%) with HCV genotype 1 infection (92–95%). The majority was overweight but not obese and 17–18% had diabetes. The baseline MELD score was 7 and 41% had cirrhosis. A total of 134 patients (25.3%) in the training cohort and 108 patients (20.8%) in the validation cohort developed one of the composite liver outcomes over a median followup of 6.5 years (range 0.5–9.2). When specific events were used as the initial outcome, there were 78 patients in the training cohort and 59 in the validation cohort with hepatic decompensation; 51 patients in training cohort and 37 in the validation cohort with HCC; and 107 patients in training cohort and 104 in the validation cohort meeting our definition for overall mortality. Figure S1 shows KM curves of cumulative probability of composite liver outcomes, hepatic decompensation only, HCC only and overall mortality in the entire cohort.

Predictor Variables Associated with Outcomes

Predictors associated with developing the composite clinical outcome, hepatic decompensation alone, HCC alone and overall mortality on multivariate analysis using the training cohort are displayed in Table 2. As compared to the longitudinal markers, none of the baseline measures were influential in the prediction models. The longitudinal markers that were the most important independent predictors of the composite clinical outcome were as follows: albumin, AST/ALT ratio, total bilirubin, AFP and platelet count. Most but not all of these five markers were retained in the models predicting other outcomes. The importance of these variables is further demonstrated in Fig. 1 where the averaged trend of each of these 5 variables over time distinguishes patients who did versus those who did not have the composite outcome (data from training cohort). As an example, the red line shows in patients who had an outcome, albumin levels were similar at baseline but dropped more during follow-up than those who did not have an outcome (green line).

We developed models to predict risk of outcomes at various time points during follow-up using accumulated longitudinal information for individual patients. Figure 2 shows two patients from the training cohort with similar biomarker values and similar predicted risk of

outcome at baseline. Subject B had worsening of liver disease over time evidenced by decreasing platelet and albumin and increasing bilirubin, AST/ALT ratio and AFP during follow-up while marker values remained stable in Subject A. The predicted risk of outcomes assessed after 1 and 2 years of follow-up increased in Subject B but remained unchanged in Subject A. These predictions are consistent with actual outcomes with Subject B meeting criteria for outcome 5 years after enrolment, whereas Subject A had not developed the outcome of interest at the time of last clinical visit, 6.5 years after enrolment.

Composite Outcome Model Performance – Training Cohort

The longitudinal model calibrated the training cohort data well, with averaged model-based predicted probabilities among a set of subjects matching well to the empirical rates for the same set of subjects (Figure S2). Across follow-up times up to 3 years, AUROCs for the longitudinal models were close to or above 0.8 for 1-, 3- and 5-year prediction (Figure S3A).

In the training set, we also compared the performance of the longitudinal model in predicting outcomes at 1, 3 and 5 years with a baseline model and an existing model that used baseline and Year 2 data on the subset of patients who had at least 2 years of follow-up [19]. For the baseline model, predictions were made using only marker information at enrolment. For all three models, outcomes at Years 3, 5 and 7 from enrolment corresponding to 1, 3 and 5 years from Year 2 were considered. The longitudinal model outperformed the baseline and baseline with Year 2 data model with improved calibration and overall bias-corrected predictive accuracy. The results of the risk prediction accuracy measures for the models developed using the training cohort are displayed in Table S1.

Composite Outcome Model Performance – Validation Cohort

When the longitudinal model was validated using the treatment arm of the study, the performance was retained with AUROCs around 0.8 for 3-year predictions at followup times up to 3 years from enrolment and for 5-year predictions at follow-up times up to 2.5 years from enrolment (Figure S3B).

The overall and threshold-specific accuracy measures were similar (Table 3), compared to those observed in training set. The discriminatory capacity of markers measured up to 2 years using data from the validation cohort for 1- (left), 3- (middle) and 5-(right) year predictions were compared across the 3 models using ROC curves and are displayed in Fig. 3(a). For predictions of 3-year risk, the AUROC was 0.81 (95%CI: 0.74–0.85) for the baseline model, 0.79 (95% CI: 0.72–0.85) for the baseline plus Year 2 data model and 0.83 (95%CI: 0.76–0.88) for the longitudinal model. At a 3-year risk threshold of 0.25, substantial improvement was observed across all predictive accuracy parameters including higher NPV, PPV and TPR with lower prediction error and lower FPR (Table 3). The accuracy of the longitudinal model tended to be better for 1 or 3 year prediction, compared with a more long-term 5-year prediction (Fig. 3a, Table 3).

To further illustrate how we can use the models in clinical practice, we plotted the TPR and FPR at different prediction intervals as a function of risk threshold above which clinical decision-making would be altered (e.g. increase intensity of clinical monitoring) (Fig. 3b). As demonstrated by these figures, at any given risk threshold, the TPR of the longitudinal

models are significantly higher compared to the baseline models and baseline + 1 followup time point models while the FPR remain similar.

Additionally, we also evaluated the longitudinal model performance when AFP was removed as a predictor variable as this marker is not always obtained in practice, particularly in patients with early stage disease. There was no to minimal (change of 0–0.01 for AUROC) change in model performance for the composite outcome with removal of AFP, at all three prediction intervals (Table S2).

Secondary Outcomes Model Performance – Validation Cohort

The longitudinal model performance was retained for predicting risk of decompensation alone (3-year AUROC = 0.92, 95% CI: 0.88–0.95) but not for risk of HCC alone (3-year AUROC = 0.59, 95% CI: 0.43–0.74) (Figure S4). When assessing risk of overall mortality, the model performance was intermediate with a 3-year AUROC of 0.75 (95% CI 0.66–0.82) (Figure S4). For secondary outcomes models that included AFP (HCC and overall mortality), we once again evaluated the impact of removing AFP as a predictor variable. Similar to the effect on the composite outcome, the impact on AUROC and prediction error were minimal (Table S2).

DISCUSSION

Chronic liver disease continues to represent a major cause of death and disability globally [2]. Chronic hepatitis C has become a focal point within hepatology in the setting of recent therapeutic advances. Given the high prevalence of CHC and the high cost of currently available DAAs, an accurate risk prediction model could provide clinicians with a tool to more objectively identify patients at highest risk for disease progression until such time when universal treatment can be implemented [20]. In this study, we used a robust longitudinal database to demonstrate a novel statistical approach to construct highly accurate prediction models using only routinely collected data. Herein we demonstrate the ability of longitudinal models to accurately risk stratify patients at both the individual and population level, in particular their ability to distinguish between patients with similar baseline data who go on to have distinct clinical outcomes.

We previously reviewed existing clinical prediction models for CHC, highlighting the limitations in model performance due to restricting data to baseline or baseline and a single follow-up time point [4]. In a recent study, we demonstrated the superiority of longitudinal models compared to baseline models in predicting risk of composite clinical outcomes in the next 12 months using machine learning algorithms [5]. In this study, we aimed to construct models using only data routinely obtained in clinical practice in order to optimize implementation of these models into real world settings. In addition, we investigated the predictive capability for multiple outcomes of interest including hepatic decompensation alone, HCC alone and overall mortality given that SVR to HCV treatment has been demonstrated to improve overall quality of life and other diseases. Lastly, we evaluated different time frames of risk prediction (1, 3 and 5 years) because in clinical practice we are interested in not only short-term but also long-term risk of disease progression. Our data demonstrate that longitudinal models outperform baseline models at both the individual and

population levels. Longitudinal models had higher AUROCs (0.89 vs 0.79–0.81), NPVs (99% vs. 96–98%) and TPRs with similar FPRs compared to the baseline models. The risk prediction accuracy was retained at longer prediction time frames (i.e. 3 years), although the superiority of risk prediction waned at more distant time frames (5 years). These accuracy measures remained similar when predicting decompensation alone or overall mortality, but not for predicting HCC alone.

Evaluating our results in further detail revealed several notable findings. First, the variables identified as being independently predictive of outcomes were in line with results of prior studies that have highlighted these laboratory markers as indicators of progression of liver disease. It was of interest that AFP was identified as an independent predictor not only for HCC alone outcome but also for the composite and overall mortality outcome, although removal of AFP from the models had minimal impact on the performance of the models. This finding is important because in clinical practice AFP is less often obtained, particularly in patients with early stage disease. A unique feature of our models is the ability to distinguish patients with similar baseline data but who have distinct clinical courses and outcomes during follow-up. As expected for all predictions, the performance of the models was less robust for longer prediction time frames. The lower accuracy for longer time frames may also be secondary to the smaller proportion of patients in the HALT-C cohort who were followed beyond 5 years. Our models were most accurate in predicting decompensation alone followed by the composite outcome, overall mortality and then HCC. Given that the composite outcome included HCC (which is more difficult to predict) and that overall mortality includes death from nonliver causes (which we do not expect we will be able to predict using markers of liver disease), the model performance falls into a logical order. Predicting HCC remains challenging and has been mirrored by results of other investigators [21–23].

The major strength of our study is the ability to make dynamic predictions of risk of clinical outcomes for patients using routinely collected data at any time during follow-up and over a range of prediction time frames. There are several limitations to our study. The primary limitation stems from the enrolment criteria for the HALT-C study which only enrolled patients with advanced disease and prior treatment failure thus limiting the generalizability of our results. In addition, the HALT-C cohort comprised primarily middle-aged Caucasian men with genotype 1 infection and thus represents only a portion of the overall population of patients with CHC. Development of models that can be applied to a broader patient population with high degree of accuracy in predicting not only hepatic decompensation but also HCC and earlier events such as cirrhosis and advanced fibrosis would be important given our goal of universal treatment. Although our models were only tested in a cohort of patients with more advanced CHC, the statistical approach to construct these models could be applied to broader cohorts of patients with CHC as well as other CLD that would similarly benefit from objective risk stratification for clinical outcomes.

In conclusion, our findings demonstrate that predictive models for risk of disease progression in CHC have improved individual and population accuracy when built on longitudinal instead of baseline data. The models retain their accuracy when risk predictions are made for shortterm (1-year) or long-term (3-year) time frames. Furthermore, we have

shown that accurate risk predictions can be made using results from a few routinely collected laboratory markers. As such, our models can be applied to real world clinical practice through existing electronic medical records (EMR) or as a broader web based application. Our models have particular clinical utility in resource limited countries where the need to identify high-risk patients is more critical. These methods to construct predictive models can be applied to other forms of CLD such as nonalcoholic fatty liver disease where these tools would be similarly useful to help guide treatment and intensity of clinical monitoring required. Future studies are needed to externally validate our results in broader patient populations, specifically those with early stage CHC and other forms of CLD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. NIH releases action plan for liver disease research. *J Investig Med*. 2005; 53(2):63–64.
2. Murphy SL, Kochanek KD, Xu J, Heron M. Deaths: Final Data for 2012. *Natl Vital Stat Rep*. 2015; 63(9):1–117.
3. Udompap P, Kim D, Kim WR. Current and future burden of chronic nonmalignant liver disease. *Clin Gastroenterol Hepatol*. 2015; 13(12):2031–41. [PubMed: 26291665]
4. Konerman MA, Yapali S, Lok AS. Systematic review: identifying patients with chronic hepatitis C in need of early treatment and intensive monitoring—predictors and predictive models of disease progression. *Aliment Pharmacol Ther*. 2014; 40(8):863–879. [PubMed: 25164152]
5. Konerman MA, Zhang Y, Zhu J, Higgins PD, Lok AS, Waljee AK. Improvement of predictive models of risk of disease progression in chronic hepatitis C by incorporating longitudinal data. *Hepatology (Baltimore, MD)*. 2015; 61(6):1832–1841.
6. Kowdley KV, Gordon SC, Reddy KR, et al. Ledipasvir and sofosbuvir for 8 or 12 weeks for chronic HCV without cirrhosis. *New England J Med*. 2014; 370(20):1879–1888. [PubMed: 24720702]
7. Colvin, HM., Mitchell, AE., editors. Institute of Medicine (US) Committee on the Prevention and Control of Viral Hepatitis Infection. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C. 2010.
8. Moyer VA. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2013; 159(5):349–357. [PubMed: 23798026]
9. Smith BD, Morgan RL, Beckett GA, et al. Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945–1965. *MMWR Recomm Rep*. 2012; 4:1–32.
10. Lavanchy D. Evolving epidemiology of hepatitis C virus. *Clin Microbiol Infect*. 2011; 17(2):107–115. [PubMed: 21091831]
11. Murphy EL. The increasing burden of mortality from viral hepatitis in the United States. *Ann Intern Med*. 2012; 157(2):149–150. [PubMed: 22801684]
12. Gilead. [March 30, 2015] U.S. Food and Drug Administration Approves Gilead's Sovaldi (Sofosbuvir) for the Treatment of Chronic Hepatitis C. Dec 6. 2013 Available at: <http://www.gilead.com/news/pressreleases/2013/12/us-food-and-drugadministration-approves-gileadssovaldi-sofosbuvir-for-the-treatmentof-chronic-hepatitis-c>
13. Barua S, Greenwald R, Grebely J, Dore GJ, Swan T, Taylor LE. Restrictions for medicaid reimbursement of sofosbuvir for the treatment of hepatitis C virus infection in the United States. *Ann Intern Med*. 2015; 163(3):215–23. [PubMed: 26120969]
14. Lee WM, Dienstag JL, Lindsay KL, 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(5):472–492. [PubMed: 15465617]

15. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* (Baltimore, MD). 2011; 53(3):1020–1022.
16. Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. 2005; 61(2):379–391. [PubMed: 16011684]
17. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005; 61(1):92–105. [PubMed: 15737082]
18. Zheng Y, Heagerty PJ. Prospective accuracy for longitudinal markers. *Biometrics*. 2007; 63(2): 332–341. [PubMed: 17688486]
19. Ghany MG, Kim HY, Stoddard A, Wright EC, Seeff LB, Lok AS. Predicting clinical outcomes using baseline and follow-up laboratory data from the hepatitis C long-term treatment against cirrhosis trial. *Hepatology* (Baltimore, MD). 2011; 54(5):1527–1537.
20. Leidner AJ, Chesson HW, Xu F, Ward JW, Spradling PR, Holmberg SD. Cost-effectiveness of hepatitis C treatment for patients in early stages of liver disease. *Hepatology* (Baltimore, MD). 2015; 61(6):1860–1869.
21. Chang KC, Wu YY, Hung CH, et al. Clinical-guide risk prediction of hepatocellular carcinoma development in chronic hepatitis C patients after interferon-based therapy. *Br J Cancer*. 2013; 109(9):2481–2488. [PubMed: 24084770]
22. Lee MH, Lu SN, Yuan Y, et al. Development and validation of a clinical scoring system for predicting risk of HCC in asymptomatic individuals seropositive for anti- HCV antibodies. *PLoS ONE*. 2014; 9(5):e94760. [PubMed: 24801353]
23. Lok AS, Everhart JE, Wright EC, et al. Maintenance peginterferon therapy and other factors associated with hepatocellular carcinoma in patients with advanced hepatitis C. *Gastroenterology*. 2011; 140:840–849. [PubMed: 21129375]

Abbreviations

AFP	alpha-fetoprotein
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUROC	area under the receiver-operating characteristic curve
BMI	body mass index
CHC	chronic hepatitis C
CI	confidence intervals
CLD	chronic liver disease
CTP	Child–Turcotte–Pugh
EMR	electronic medical records
FPR	false-positive rate
HALT-C	Hepatitis C Antiviral Long-term Treatment Against Cirrhosis
HCC	hepatocellular carcinoma
HR	hazard ratio

INR	international normalized ratio
MELD	model for end-stage liver disease
NPV	negative predictive value
PPV	positive predictive value
ROC	receiver-operating characteristic
TPR and FPR	true- and false-positive rates
ULN	upper limit of normal

Fig. 1. Longitudinal Variables Independently Predictive of Composite Outcome – Training Cohort. Observed longitudinal variables by year and outcome. Individual marker value trajectories are shown in green when no event was observed and red when the composite outcome was observed. Smoothed lines show averaged marker levels by groups of patients with and without outcomes. All marker values are on a log scale.

Fig. 2.

Individual Level Composite Outcome Risk Prediction Based on Longitudinal Clinical Data at Multiple Prediction Time Points – Training Cohort. Top panel: Marker values by time are shown for two individuals (a and b). Observed values are shown as points in addition to the smoothed values shown as a line. Marker values are on a log scale. The vertical grey lines indicate prediction times, where marker information up to the time of prediction is used to predict individual level risk. Bottom panel: For each individual prediction time indicated by a vertical line, we predict future risk of developing the outcome.

Fig. 3.

(a). ROC Curves Comparing Baseline, the Baseline + 1 follow-up time point, and Longitudinal Models for Predicting Composite Outcome in the next 1 (left), 3 (middle) and 5 years (right) at 2 years follow-up time – Validation Cohort. (b). Comparison of True-Positive and False-Positive Rates Across Risk Thresholds Among Baseline, Baseline + 1 follow-up time point, and Longitudinal Models in Predicting Composite Outcome in the next 1 (left), 3 (middle) and 5 years (right) at 2 years follow-up time – Validation Cohort

Table 1

Baseline characteristics and incidence of outcomes *

	Training cohort (n = 533)	Validation cohort (n = 517)
Demographics		
Age (year)	50 (7)	51 (7)
		<i>P</i> = 0.033
Sex, Female	28%	30%
Race, White	71%	72%
Viral Characteristics		
HCV genotype 1	92%	95%
		<i>P</i> = 0.026
HCV RNA (log ₁₀ IU/ml)	6.4 (0.5)	6.4 (0.5)
Duration of Infection (years)	27 (8)	29 (8)
		<i>P</i> = 0.004
Clinical Characteristics		
BMI (kg/m ²)	30 (6)	30 (5)
Diabetes	17%	18%
Alcohol intake/day (gm)	28 (46)	22 (32)
Tobacco Use (pack year)	15 (17)	15 (16)
Labs		
Platelet count (1000/mm ³)	165 (68)	165 (63)
INR	1.0 (0.1)	1.0 (0.1)
AST ratio to ULN	2.1 (1.5)	2.1 (1.5)
ALT ratio to ULN	2.2 (1.7)	2.1 (1.5)
AST/ALT ratio	0.9 (0.3)	0.9 (0.3)
Alkaline Phosphatase ratio to ULN	0.8 (0.4)	0.8 (0.4)
Albumin (g/dL)	3.9 (0.4)	3.9 (0.4)
Total Bilirubin (mg/dL)	0.8 (0.4)	0.8 (0.4)
AFP ratio to ULN	1.8 (3.3)	1.7 (3.0)
MELD	7.1 (1.4)	7.0 (1.4)
APRI	1.6 (1.6)	1.6 (1.5)
Histology		
Ishak fibrosis (5)	41%	41%
HAI	7.5 (2.0)	7.6 (2.1)
Steatosis score (2)	39%	43%
Number of observed outcome		
Composite Outcome	134	108
Hepatic Decompensation	78	59
HCC	51	37
Median follow-up time (years) [range]	6.5 [0.5–9.2]	6.5 [0.5–9.1]

* Mean (SD) of variables at baseline in the training and validation cohorts. *P* values only shown for significant results. HCV, hepatitis C virus; BMI, body mass index; INR, international normalized ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ULN, upper limit of

normal; AFP, alpha-fetoprotein; MELD, model for end-stage liver disease; APRI, aspartate to platelet ratio index; HAI, histologic activity index; HCC, hepatocellular carcinoma.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Predictors of outcomes – training cohort *

Predictor	Composite outcome	HCC	Hepatic decompensation	Overall mortality
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Platelet count (1000/mm ³)	0.33 (0.20, 0.54)	0.15 (0.08, 0.3)	0.45 (0.24, 0.83)	0.24 (0.14, 0.42)
Albumin	0.01 (0.002, 0.08)		0.01 (0.001, 0.06)	0.01 (0.001, 0.07)
Total Bilirubin	1.46 (0.93, 2.31)	0.51 (0.27, 0.95)	2.70 (1.50, 4.85)	
AFP ratio to ULN	1.26 (1.07, 1.48)	1.46 (1.16, 1.84)		1.37 (1.14, 1.64)
AST/ALT ratio	1.80 (0.98, 3.32)		2.26 (0.93, 5.53)	

AFP, alpha-fetoprotein; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HCC, hepatocellular carcinoma; HR, hazard ratio; CI, confidence interval; ULN, upper limit of normal.

* Cox proportional hazard results. Model fit using log transformation of smoothed longitudinal values and adjusted for measurement time.

Table 3

Risk prediction accuracy summary measures misclassification table for baseline and longitudinal predictive models of composite outcome – validation cohort*

	Longitudinal (95% CI)	Baseline + 1 Follow-up time point (95% CI)	Baseline only (95% CI)
5-Year risk prediction			
Prediction Error	0.24 (0.20–0.30)	0.25 (0.21–0.29)	0.25 (0.21–0.30)
TPR	0.67 (0.56–0.76)	0.48 (0.36–0.60)	0.43 (0.32–0.52)
FPR	0.13 (0.08–0.18)	0.10 (0.06–0.15)	0.08 (0.04–0.12)
NPV	0.91 (0.87–0.94)	0.87 (0.83–0.91)	0.84 (0.80–0.88)
PPV	0.57 (0.45–0.69)	0.54 (0.41–0.67)	0.63 (0.47–0.78)
AUROC	0.81 (0.75–0.87)	0.77 (0.71–0.83)	0.82 (0.76–0.86)
3-Year Risk Prediction			
Prediction Error	0.10 (0.08–0.12)	0.11 (0.09–0.14)	0.14 (0.11–0.16)
TPR	0.76 (0.65–0.85)	0.62 (0.49–0.73)	0.56 (0.40–0.65)
FPR	0.14 (0.10–0.17)	0.18 (0.15–0.22)	0.16 (0.12–0.20)
NPV	0.96 (0.93–0.97)	0.93 (0.90–0.95)	0.91 (0.87–0.94)
PPV	0.47 (0.37–0.57)	0.35 (0.26–0.43)	0.41 (0.28–0.49)
AUROC	0.83 (0.76–0.88)	0.79 (0.72–0.85)	0.81 (0.74–0.85)
1-Year Risk Prediction			
Prediction Error	0.03 (0.02–0.05)	0.04 (0.02–0.05)	0.07 (0.05–0.09)
TPR	0.81 (0.64–0.98)	0.63 (0.40–0.85)	0.60 (0.39–0.76)
FPR	0.95 (0.07–0.12)	0.11 (0.08–0.14)	0.18 (0.14–0.22)
NPV	0.99 (0.98–1.00)	0.98 (0.97–0.99)	0.96 (0.94–0.98)
PPV	0.29 (0.18–0.40)	0.20 (0.10–0.29)	0.20 (0.12–0.29)
AUROC	0.89 (0.80–0.96)	0.79 (0.66–0.89)	0.81 (0.72–0.87)

* Risk thresholds for 1-, 3- and 5-year predictions were 0.1, 0.25 and 0.5.

For this analysis, predictions were made at Year 2 for occurrence of events 1, 3 and 5 years from Year 2, that is at Years 3, 5 and 7 from enrolment.

TPR, true-positive rate; FPR, false-positive rate; NPV, negative predictive value; PPV, positive predictive value; AUROC, area under the receiver-operating characteristic curve.