Development of a national Department of Veterans Affairs mortality risk prediction model among patients with cirrhosis

Jejo David Koola [1,2,3] Samuel Ho,[2,4,5] Guanhua Chen,[6] Amy M Perkins,[7] Aize Cao,[1,8] Sharon E Davis,[1,8] Michael E Matheny[1,7,8,9]

For numbered affiliations see end of article.

Correspondence to
Dr Jejo David Koola, Medicine, University of California San Diego, La Jolla, California, United States; jkoola@ucsd.edu

ABSTRACT

Objective Cirrhotic patients are at high hospitalisation risk with subsequent high mortality. Current risk prediction models have varied performances with methodological room for improvement. We used current analytical techniques using automatically extractable variables from the electronic health record (EHR) to develop and validate a posthospitalisation mortality risk score for cirrhotic patients and compared performance with the model for end-stage liver disease (MELD), model for end-stage liver disease with sodium (MELD-Na), and the CLIF Consortium Acute Decompensation (CLIF-C AD) models.

Design We analysed a retrospective cohort of 73 976 patients comprising 247 650 hospitalisations between 2006 and 2013 at any of 123 Department of Veterans Affairs hospitals. Using 45 predictor variables, we built a time-dependent Cox proportional hazards model with all-cause mortality as the outcome. We compared performance to the three extant models and reported discrimination and calibration using bootstrapping. Furthermore, we analysed differential utility using the net reclassification index (NRI).

Results The C-statistic for the final model was 0.863, representing a significant improvement over the MELD, MELD-Na, and the CLIF-C AD, which had C-statistics of 0.655, 0.675, and 0.679, respectively. Multiple risk factors were significant in our model, including variables reflecting disease severity and haemodynamic compromise. The NRI showed a 24% improvement in predicting survival of low-risk patients and a 30% improvement in predicting death of high-risk patients.

Conclusion We developed a more accurate mortality risk prediction score using variables automatically extractable from an EHR that may be used to risk stratify patients with cirrhosis for targeted postdischarge management.

Summary box

What is already known about this subject?

► Cirrhosis has a high mortality and early risk stratification, especially in hospitalised patients, is important.

► Current risk models have widely varying performances depending on the study and cohort.

► Existing models are built assessing patients at a single point in time, for example, at hospital discharge.

What are the new findings?

► This paper builds a model with 73 976 patients comprising 247 650 hospitalisations with granular clinical data, resulting in a model with a C-statistic of 0.863.

► The Net Reclassification Index, a measure of model improvement, shows a 25% improvement in predicting survival of low-risk patients compared with existing models.

How might it impact on clinical practice in the foreseeable future?

► Increasing efforts to implement real-time models within electronic health record systems may allow more complex models to guide interventions such as care management and referral to hospice.

population.[3] Prognostication can help guide clinical decision making, transplant referral, care coordination, and hospice enrolment. Performance of the most common cirrhosis risk prediction models, including the model for end-stage liver disease (MELD),[4] model for end-stage liver disease with sodium (MELD-Na),[5] Chronic Liver Failure Consortium-Sequential Organ Failure Score,[6] CLIF Consortium acute decompensation (CLIF-C AD) score,[7] and the CLIF Consortium Acute on Chronic Liver Failure score,[8] has been varied.[7–10]

A systematic review of cirrhosis survival models by D'Amico and colleagues found 181 studies[11]; however, these studies were

INTRODUCTION

Cirrhosis has increased from being the fourteenth to being the eighth cause of death in the USA[1] with similar increases seen globally.[2] Patients with cirrhosis are at increased risk of hospital admission due to various causes, increased risk of readmission, and increased risk of death compared with the general

still limited by focusing on a selected group of patients with relatively small sample sizes, built on purely administrative databases with limited information, lacking validation, controlling for overfitting, or providing calibration metrics. Risk prediction models require revalidation and recalibration when applied in a new cohort as their performance often degrades due to changes in prevalence of risk factors and case mix.[12] Model performance can also degrade over time even when used within the same institution.[13] Because of widespread electronic health record (EHR) system adoption, very large datasets have become available for advanced analytics and machine learning.[14]

Due to these reasons, it is imperative that the models be anchored into the health system of use. It is also important to deploy many of these tools within an EHR in an automated fashion because of the increasing use of these tools and the need to minimise user burden and facilitate scalability. The Department of Veterans Affairs (VA) faces a higher burden of patients with cirrhosis than the general US population,[15] and the literature evaluating cirrhosis mortality in this cohort has been limited.[16–19] Opportunities exist for improved cirrhosis care for veterans,[19–21] and better mortality prediction may help motivate this care. Numerous studies have focused on posthospital discharge risk stratification in order to more effectively target care, either for the purpose of preventing inappropriate readmission or leveraging shared decision making to motivate palliative care referral.[22] We hypothesised that a model built using a large EHR database, using present-on-admission data and information collected automatically during the hospitalisation, could outperform traditional mortality risk measures.

METHODS
Study population
This work was part of a multiyear, multisite study to improve cirrhosis care at the Department of VA. We have previously published our efforts on predicting hospital readmission, and we refer readers to that study for further details on this cohort.[23] We analysed a retrospective cohort of patients with cirrhosis hospitalised for any cause from among 123 medical centres in the Department of VA between 1 January 2006 and 31 December 2013, with historical data from 1 January 2005 to allow for variable ascertainment. Patients were identified by using International Classification of Diseases, Ninth Revision (ICD-9) codes 571.2 (alcoholic cirrhosis), 571.5 (nonalcoholic cirrhosis), or any code identifying a history of a cirrhosis complication (varices, hepatic encephalopathy, hepatorenal syndrome, or portal hypertension). The full list of ICD-9 codes used is provided in online supplementary appendix table 1. Previous studies have shown that using administrative codes for cirrhosis or one of its cardinal complications can accurately identify a retrospective cohort with a positive predictive value (PPV) ranging from 84% to 92%.[24]
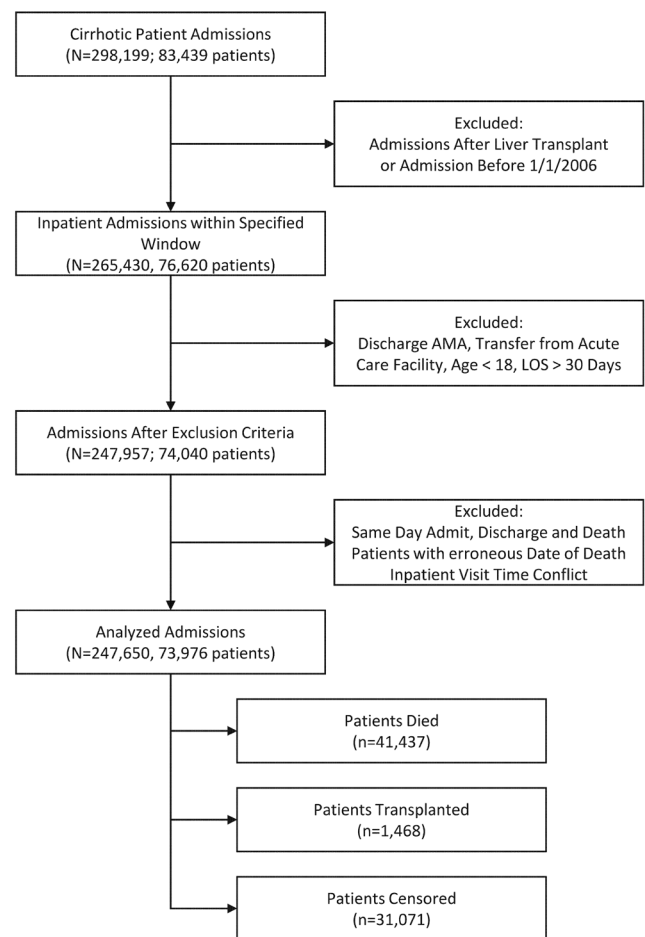

Figure 1 Flow of patients from total number of patients before exclusion criteria to total number of patients included in the study. AMA, against medical discharge; LOS, length of stay.

We included hospitalisations from patients who had the previously mentioned cirrhosis or cirrhosis complication codes at any time prior to the index hospitalisation. We excluded hospitalisations if the patient was discharged against medical advice, all hospitalisations after liver transplant (including the transplant hospitalisation itself), if the patient was transferred from another acute care hospital, paediatric patients, or if the hospital length of stay was greater than 30 days. We excluded hospitalisations with lengths of stay greater than 30 days because they were related frequently to issues with identifying a discharge disposition for the patient, rather than severity of illness. Refer to figure 1 for cohort flow diagram.

Data collection
The Veterans Health Administration is America's largest integrated healthcare system, serving nine million enrolled military veterans each year, and including acute inpatient hospitals, outpatient primary care and subspecialist clinics, outpatient pharmacies, rehabilitation facilities, and long-term care facilities. All VA personnel use the same EHR for documentation and clinical care.[25] The VA Informatics and Computing Infrastructure project
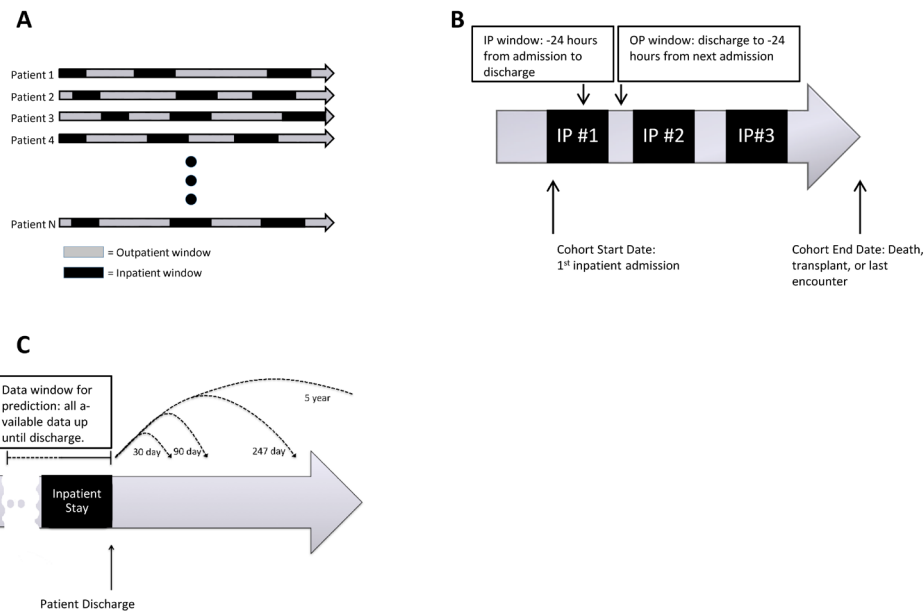
Figure 2 Outpatient (OP) versus inpatient (IP) time windows and variable ascertainment. Each patient's clinical course is summarised by a series of IP and OP time periods. The patient's clinical course, represented by all of the variables in the model, was calculated during the following time points: start of every IP stay, discharge of every IP stay, patient death, and patient censoring, allowing for time-varying coefficients in the Cox proportional hazards model. (A,B) Data used for model creation. (C) Model being used for prediction, where a clinical prediction can be made for any time point using the same survival model.IP, inpatient; OP, outpatient.

has colocated and harmonised data from all VA sites into a corporate data warehouse.[26]

Predictor variables

We initially evaluated a broad range of variables encompassing demographics, medications, laboratory values, diagnoses and procedures, vital signs, and healthcare use. To eliminate noise variables and to reduce overfitting, we performed variable selection using a penalised Cox proportional hazards model, using the L_1 penalty (least absolute shrinkage and selection operator (LASSO)), to select a subset of the predictor variables.[27] Our final model contained 45 variables.

The creatinine value was transformed with the natural logarithm. Restricted cubic splines modelled three continuous variables (age, Body Mass Index (BMI), and creatinine) to take into account the non-linear effect on the hazard. Medications were represented by their drug class, for example, 'beta blockers', using VA drug class codes.[28] We also represented certain medications frequently used to treat cirrhosis-related complications, for example, lactulose, as separate, individual variables.

Each variable was calculated during the following time points: start of every inpatient stay, discharge of every inpatient stay, patient death, and patient censoring. This pattern of repeated sampling of the predictor variables in our time-dependent model led to the natural occurrence of two time windows: the outpatient and the inpatient windows (refer to figure 2). We performed multiple imputation for missing values using non-negative matrix factorisation[29] via the R Non-Negative Linear Models

(NNLM) package.[30] We refer the reader to online supplementary appendix table 2 and Further clarification subsection, for further details regarding the methods and a description of all candidate variables.

Statistical analysis

We constructed an unpenalised time-dependent covariate Cox proportional hazards model[31] with the primary outcome being all-cause death and using the 45 features identified by the variable selection procedure; we censored at liver transplant, date of last encounter with the VA health system, or study end. We used a time-dependent covariate model to incorporate information from multiple time points from the patient's clinical course for improved model performance. We assessed overall discrimination using Harrell's C-statistic.[32]

Unlike logistic regression, survival models allowed us to perform prediction at any postdischarge time point (refer to figure 2). To better contrast our model against extant models, in addition to global performance, we specifically assessed performance at predicting 90-day mortality. We evaluated discrimination and calibration using the area under the receiver operating characteristic curve (AUC) and the Estimated Calibration Index (ECI). The ECI looks at the squared difference between the predicted probability and an estimated observed probability, ranging between 0 and 100, with 0 meaning perfect calibration.[33] Additionally, we graphically analysed calibration by investigating the smoothed observed-to-predicted probability plot.[33]

We internally validated our model by conducting 100 bootstrap evaluations to build the 95% bootstrap CI for the overall C-statistic, 90-day AUC, 90-day ECI, and Net Reclassification Index (NRI). We refer the reader to the online supplementary appendix figure 1 and Further clarification subsection for a graphical overview of our methods and further details. All statistical analyses were performed using the R statistical programming suite V.3.5.1.

Model comparison

We compared our model against the MELD,[4] MELD-Na,[5] and the CLIF-C AD[7] scores calculated at discharge. We recalibrated the three scores by constructing separate univariate survival models. We tailored the three extant models to the validation cohort because of differences in mortality and risk factors among VA patients.[12]

To demonstrate clinical utility, we analysed performance for two use cases: (1) identifying patients at very low risk of dying, <5%; and (2) finding very high-risk patients, >40% risk of dying.[34 35] We report the PPV and the NRI of our model compared with the extant models. The NRI offers a global assessment of the trade-off between true positives and false positives, with values of >0 indicating improved prediction performance.

Sensitivity analyses

We performed four sensitivity analyses (1) treating death and liver transplant as a composite outcome, (2) assessing model performance for cirrhosis-related admissions, (3) comparing model performance for patients with and without heart failure, and (4) comparing model performance for patients with and without diabetes. Although treating transplant as a competing risk may be optimal, it does not easily extend to a time-dependent model.[36] Because a minority of our patients underwent a transplant (1468, 2%) we chose to treat it as a censoring event; however, we report the sensitivity analysis to assess for risk of bias. Importantly, in order to provide additional directly comparable results to pre-existing models, we evaluated a new and prior model performance on cirrhosis-related admissions (definition in online supplementary appendix, Further clarification subsection).

RESULTS

Study population

After applying inclusion and exclusion criteria, 73 976 patients were included in the study with a total of 247 650 hospitalisations. Men represented 97.8% of the total admissions, with an age of 60.7±9.0 (mean±SD). Caucasian and African–American patients accounted for the majority of hospital admissions (73.7% and 18.3%, respectively). The aetiology of cirrhosis was mainly alcoholic (30.9%), viral hepatitis (14.2%), or alcoholic and viral (35.7%). In the remaining patients, the causes of cirrhosis were NAFLD (30 921, 12.5%), other/cryptogenic (40 309, 16.3%), primary biliary cirrhosis (1096, 0.4%), haemochromatosis (1087, 0.4%), and autoimmune hepatitis (393, 0.2%) . The average MELD score across all hospitalisations was 12.7±5.2 (mean±SD), though of note,we had 39 529 admissions with a MELD score of ≥18. Refer to table 1 for a description of the cohort.

There was a median follow-up of 474 days (IQR 111–1159) for a total of 149 232 patient-years. Adjusting for bias from early deaths using the Kaplan-Meier estimate of potential follow-up, follow-up improved to 1054 days (IQR 359–1993).[37] The median survival from time of first hospitalisation was 1064 days (IQR 207–2444). There were 41 437 events for an overall mortality of 56.0%. Figure 2 within the online supplementary appendix depicts the stratified survival curves.

Predictors of mortality

Table 2 presents the statistically significant HRs for the variables in the model. Refer to online supplementary appendix table 3 for HRs for all variables. All cirrhosis complications, vitals indicating haemodynamic instability, and lower BMI increased mortality. Many of the laboratory variables were statistically significant but had an overall weak effect size when predicting mortality except for albumin and total bilirubin. Every gram per decilitre increase in serum albumin concentration decreased the HR by 30% (HR 0.70, 95% CI 0.68 to 0.71) and every gram per decilitre increase in bilirubin increased the HR by 5% (HR 1.05, 95% CI 1.05 to 1.05). Using splines allowed us to investigate the non-linear effect of serum creatinine on mortality, with increasing mortality at the extremes (refer to online supplementary appendix figure 3). Compared with a baseline creatinine of 1.0 mg/dL, a creatinine of 2.0 and 0.4 has HRs of 1.43 (95% CI 1.39 to 1.47) and 1.12 (95% CI 1.05 to 1.20), respectively. Discharge to any location other than home significantly increased mortality from 1.30 (95% CI 1.01 to 1.67) for unknown discharge disposition up to 4.73 (95% CI 3.66 to 6.10) for discharge to hospice. Medications had varied effects on mortality, though overall themes included medications being used to treat complications, for example, lactulose (HR 1.24, 95% CI 1.19 to 1.28), associated with higher mortality.

Mortality risk model performance

The model presented good discrimination with a C-statistic of 0.863 (95% CI 0.863 to 0.864). For the specific use case of predicting 90-day mortality, our model showed good discrimination, with an AUC of 0.79 (95% CI 0.79 to 0.79). Figure 3 demonstrates the observed-to-expected probability plot for the 90-day prediction. We see excellent calibration for predicted probabilities less than 0.25, which represented 23 286/50 108 (46.5%) of the observations in our model validation dataset. The three extant models, however, show systematic underprediction and overprediction for probabilities less than and over ~0.60, respectively.

Table 1 Demographic, clinical and laboratory variables of included patients across all admissions

Variables	All patients, all admissions (N=247 650)
Age (years), mean (SD)	60.7 (9.0)
Gender (male), n (%)	242 088 (97.8)
Race, n (%)	
Caucasian	182 638 (73.7)
African–American	45 317 (18.3)
Asian–Hawaiian–Pacific Islander	4018 (1.6)
American Indian–Alaskan Native	3940 (1.6)
Unknown	11 737 (4.7)
Aetiology, n (%)	
Alcoholic	76 591 (30.9)
Viral (hepatitis B and C)	35 189 (14.2)
Alcoholic and viral	88 501 (35.7)
Non-alcoholic fatty liver disease	30 921 (12.5)
Haemochromatosis	1087 (0.4)
Autoimmune hepatitis	393 (0.2)
Biliary cirrhosis	1096 (0.4)
Other/cryptogenic	40 309 (16.3)
Healthcare use (past 1 year)	
ER visits	3.3 (6.4)
Inpatient hospitalisations	2.3 (3.5)
Outpatient visits	46.1 (50.6)
Non-face-to-face communication	8.0 (10.2)
Congestive heart failure, n (%)	57 290 (23.1)
Diabetes mellitus, n (%)	103 260 (41.7)
History of cirrhosis complications, n (%)	
Hepatic encephalopathy	50 086 (20.2)
Varices	48 016 (19.4)
Spontaneous bacterial peritonitis	11 529 (4.7)
Ascites	75 358 (30.4)
Hepatocellular carcinoma	21 684 (8.8)
Hepatorenal syndrome	4559 (1.8)
Vitals	
Systolic blood pressure	125.8 (19.3)
Diastolic blood pressure	73.1 (12.5)
Labs, median (IQR)	
Creatinine	1.0 (0.8–1.3)
Blood urea nitrogen	15.0 (10.0–21.0)
Sodium	137.0 (134.0–139.0)
Total bilirubin	1.1 (0.6–2.0)
Albumin	3.2 (2.7–3.7)

Continued

Table 1 Continued

Variables	All patients, all admissions (N=247 650)
International normalized ratio	1.2 (1.1–1.4)
White blood cell	6.0 (4.6–7.9)
Platelets	132.0 (84.4–200.0)
Alanine aminotransferase	34.0 (21.0–58.0)
Aspartate aminotransferase	48.0 (29.0–85.0)
Risk scores	
MELD, mean (SD)	12.7 (5.2)
MELD<12, n (%)	135 287 (54.6)
MELD≥12 and <18, n (%)	72 834 (29.4)
MELD≥18, n (%)	39 529 (16.0)
MELD-Na, mean (SD)	15.1 (5.6)
CLIF-C AD, mean (SD)	50.3 (8.1)
Disposition, n (%)	
Home	213 694 (86.3)
Hospice	185 (0.1)
Hospital	4938 (2.0)
In-hospital death	10 630 (4.3)
Nursing home	17 432 (7.0)
Other house	179 (0.1)
Unknown	1008 (0.4)

Comparison to existing models

Our model's performance, as measured by the C-statistic, was significantly better than the MELD, MELD-Na, and CLIF-C AD scores: 0.863 vs 0.655, 0.675, and 0.679, respectively (table 3). Global prediction error rate, when compared with other scores, improved by 27.1%–31.8% (refer to online supplementary appendix, Further clarification subsection, for details). When looking at the classification error, our model had a PPV of 9816/10 092 (97.3%) when identifying low-risk patients (predicted mortality at 90 days of <5%). Due to overprediction, the MELD and MELD-Na models did not generate any predictions of <5%. The CLIF-C AD had a PPV of 88.5% for identifying low-risk patients; however, the CLIF-C AD only predicted low risk for 407 out of the 50 108 patient validation cohort, resulting in a sensitivity of 1.0%. Similarly, for high-risk patients, our model afforded the highest PPV (54.5%) at the highest sensitivity (28.5%) compared with the other models. Refer to table 3 for details.

Using the NRI, our model achieved a 24% improvement in predicting survival of patients at low mortality risk and a 29%–31% improvement in predicting death accurately for high-risk patients. Refer to the online supplementary appendix figure 4 for details.

Sensitivity analyses

The sensitivity analyses, treating transplant as a composite outcome with death and cirrhosis-related readmission performance, demonstrated little change in model

Table 2 Statistically significant HRs from the time-dependent Cox proportional hazards model

Risk factor	Beta (SE)	HR (95% CI)	P value
Demographics			
Race (reference: Caucasian)			
Unknown	0.339 (0.025)	1.404 (1.336 to 1.476)	<0.001
African–American	−0.101 (0.018)	0.904 (0.872 to 0.937)	<0.001
Asian–Hawaiian–Pacific Islander	−0.077 (0.052)	0.926 (0.836 to 1.026)	0.14
American Indian–Alaskan Native	−0.043 (0.056)	0.958 (0.858 to 1.069)	0.444
Age	0.035 (0.004)	1.036 (1.027 to 1.045)	<0.001
Age*	−0.021 (0.035)	0.979 (0.914 to 1.049)	0.546
Age*	0.071 (0.146)	1.074 (0.806 to 1.431)	0.625
Age*	−0.081 (0.217)	0.922 (0.602 to 1.412)	0.709
Age contrasts (reference: 45)			
35 vs 45		0.703 (0.646 to 0.764)	
50 vs 45		1.193 (1.145 to 1.243)	
60 vs 45		1.646 (1.550 to 1.748)	
70 vs 45		2.265 (2.122 to 2.418)	
History of complications			
Hepatorenal syndrome	0.221 (0.032)	1.248 (1.171 to 1.329)	<0.001
Hepatic encephalopathy	0.242 (0.017)	1.274 (1.232 to 1.318)	<0.001
Hepatocellular carcinoma	0.666 (0.023)	1.947 (1.860 to 2.038)	<0.001
Paracentesis	0.175 (0.019)	1.191 (1.147 to 1.236)	<0.001
Ascites	0.140 (0.018)	1.150 (1.111 to 1.191)	<0.001
Healthcare use (HRs are per visit/communication in the past year)			
Number of inpatient visits	0.032 (0.002)	1.032 (1.027 to 1.037)	<0.001
Number of CT images	0.006 (0.003)	1.006 (1.000 to 1.013)	0.048
LACE score	−0.005 (0.002)	0.995 (0.992 to 0.999)	0.011
Labs			
Albumin	−0.362 (0.011)	0.697 (0.682 to 0.712)	<0.001
Alkaline phosphatase	0.001 (0.000)	1.001 (1.001 to 1.001)	<0.001
Alanine aminotransferase	0.000 (0.000)	1.000 (1.000 to 1.000)	0.046
Aminotransferase	0.000 (0.000)	1.000 (1.000 to 1.000)	<0.001
Total bilirubin	0.049 (0.001)	1.050 (1.047 to 1.052)	<0.001
Serum bicarbonate	−0.011 (0.002)	0.989 (0.986 to 0.992)	<0.001
International normalised ratio	0.204 (0.024)	1.226 (1.169 to 1.286)	<0.001
Potassium	0.135 (0.012)	1.145 (1.118 to 1.172)	<0.001
Log (creatinine)	−0.376 (0.053)	0.687 (0.618 to 0.762)	<0.001
Log (creatinine)*	4.901 (0.381)	134.426 (63.656 to 283.876)	<0.001
Log (creatinine)*	−13.066 (1.053)	0.000 (0.000 to 0.000)	<0.001
Creatinine contrasts (reference: 1.0)			
0.4 vs 1.0		1.121 (1.048 to 1.198)	
2.0 vs 1.0		1.429 (1.389 to 1.471)	
Sodium	−0.019 (0.002)	0.981 (0.978 to 0.984)	<0.001
Platelets	−0.002 (0.000)	0.998 (0.998 to 0.999)	<0.001
Prothrombin time	0.005 (0.002)	1.005 (1.001 to 1.010)	0.027
White blood cell count	0.018 (0.001)	1.018 (1.016 to 1.020)	<0.001

Continued

Table 2	Continued		
Risk factor	**Beta (SE)**	**HR (95% CI)**	**P value**
Meds			
Human albumin	0.627 (0.026)	1.871 (1.780 to 1.968)	<0.001
Cephalosporins, first generation	−0.447 (0.039)	0.640 (0.592 to 0.691)	<0.001
Glucocorticoids	0.269 (0.023)	1.309 (1.251 to 1.369)	<0.001
Lactulose	0.211 (0.017)	1.235 (1.194 to 1.277)	<0.001
Midodrine	0.581 (0.042)	1.787 (1.647 to 1.940)	<0.001
Opioids	0.147 (0.014)	1.159 (1.128 to 1.191)	<0.001
HMG Co-A reductase inhibitors	−0.186 (0.021)	0.831 (0.798 to 0.865)	<0.001
Comorbidities			
Congestive heart failure	0.190 (0.017)	1.210 (1.170 to 1.251)	<0.001
Fluid and electrolyte disorder	0.089 (0.015)	1.093 (1.061 to 1.126)	<0.001
Metastatic cancer	0.886 (0.025)	2.424 (2.308 to 2.547)	<0.001
Solid tumour without metastasis	0.220 (0.020)	1.246 (1.199 to 1.294)	<0.001
Weight loss	0.068 (0.018)	1.071 (1.033 to 1.109)	<0.001
Disposition (reference: home)			
Hospice	1.553 (0.130)	4.727 (3.661 to 6.103)	<0.001
Hospital	0.548 (0.061)	1.729 (1.535 to 1.947)	<0.001
Inpatient	2.852 (0.018)	17.320 (16.730 to 17.932)	<0.001
Nursing home	0.989 (0.022)	2.690 (2.575 to 2.810)	<0.001
Other	−0.035 (0.409)	0.965 (0.433 to 2.150)	0.931
Unknown	0.261 (0.128)	1.298 (1.010 to 1.667)	0.041
Missing	−0.046 (0.235)	0.955 (0.602 to 1.515)	0.845
Vitals			
Systolic blood pressure	−0.004 (0.000)	0.996 (0.995 to 0.997)	<0.001
Pulse oximetry	−0.024 (0.002)	0.976 (0.972 to 0.981)	<0.001
Pulse	0.011 (0.000)	1.011 (1.010 to 1.012)	<0.001
BMI	−0.068 (0.005)	0.934 (0.925 to 0.943)	<0.001
BMI*	0.193 (0.044)	1.213 (1.113 to 1.323)	<0.001
BMI*	−0.444 (0.191)	0.642 (0.441 to 0.932)	0.02
BMI*	0.272 (0.238)	1.312 (0.823 to 2.091)	0.253
BMI contrasts (reference: 20)			
14 vs 20		1.506 (1.425 to 1.591)	
16 vs 20		1.313 (1.266 to 1.363)	
18 vs 20		1.146 (1.125 to 1.167)	
30 vs 20		0.724 (0.700 to 0.750)	
40 vs 20		0.745 (0.716 to 0.775)	

'Inpatient' disposition is a dummy variable used to encode the inpatient time frame in the time-dependent covariate regression model.
*These variables represent splines.
BMI, Body Mass Index; LACE, length of stay, acuity of admission, Charlson Comorbidity Index and number of emergency department visits.

discrimination. Model performance for predicting 90-day mortality did not differ between patients with and without heart failure (AUC of 0.78 (95% CI 0.78 to 0.79) vs 0.79 (95% CI 0.78 to 0.79)). There was a slight degradation in performance in predicting 90-day mortality for patients with versus without diabetes (AUC of 0.77 (95% CI 0.77 to 0.78) vs 0.80 (95% CI 0.79–0.80)). We refer the reader to online supplementary appendix tables 4–7, online supplementary appendix figures 5 and 6 for details.

DISCUSSION

In this national VA cohort study of patients with cirrhosis, the overall mortality was 56%, with 90-day, 6-month,
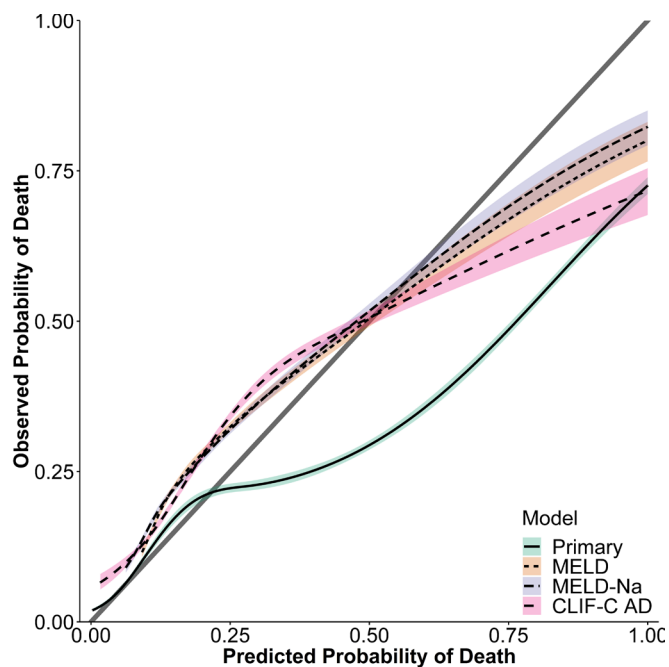
Figure 3 Observed-to-expected probability plot for 90-day prediction compared with MELD, MELD-Na, and CLIF-C AD. Perfect calibration lies along the identity line as depicted by the grey line. MELD, model for end-stage liver disease; MELD-Na, model for end-stage liver disease with sodium; CLIF-C AD, CLIF Consortium acute decompensation score.

and 1-year postdischarge mortalities of 18%, 24%, and 32%, respectively. We used data analytical techniques with the entire medical record to develop a model with good discrimination and calibration, and we were able to identify a large group of patients in the model validation cohort (10 092/50 108, 20.1%) with a very low 3-month mortality (predicted probability of death of <5%, observed survival rate of 97.3%), suggesting possible reallocation of healthcare resources for these patients. Though calibration was modest for high-risk patients, the model still outperformed the MELD, MELD-Na, and CLIF-C AD at identifying patients with poor prognosis who should be targeted for increased scrutiny and case management, in order to either prevent early readmission or motivate hospice referral.[38]

Though the MELD and the MELD-Na have become the de facto standards for mortality prediction in cirrhosis, their performance in subsequent studies has been highly varied.[7–10 39] Because the models were built using small-sized to modest-sized cohorts, changes in case mix and unmeasured factors can have significant effects on subsequent model performance. Despite recalibrating the MELD, MELD-Na, and CLIF-C AD scores to our VA cohort and limiting the analysis to comparable subcohorts, they performed modestly in this population. Our NRI analysis shows that our model would correctly classify an additional 24 out of every 100 discharged hospital

Table 3 Predictive performance for our model versus the MELD, MELD-Na and CLIF-C AD

	Our model	MELD	MELD-Na	CLIF-C AD
Global performance				
C-statistic	0.863 (0.863–0.864)	0.655 (0.655–0.655)	0.675 (0.675–0.675)	0.679 (0.679–0.679)
Mortality prediction at 90 days				
AUC (95% CI)	0.79 (0.79 to 0.79)	0.65 (0.65 to 0.65)	0.67 (0.67 to 0.67)	0.68 (0.68 to 0.68)
ECI (95% CI)	2.46 (2.22 to 2.72)	0.40 (0.36 to 0.45)	0.37 (0.33 to 0.41)	0.42 (0.38 to 0.47)
Classification error for identifying low-risk patients (predicted mortality <5%)				
Sensitivity (%)	26.3	n/a	n/a	1.0
Specificity (%)	97.8	100	100	99.6
PPV (%)	97.3	n/a	n/a	88.5
NPV (%)	31.1	25.4	25.4	25.5
Classification error for identifying high-risk patients (predicted mortality >40%)				
Sensitivity (%)	28.5	3.8	4.9	4.4
Specificity (%)	29.9	86.8	82.8	85.6
PPV (%)	54.5	45.9	45.5	47.4
NPV (%)	12.5	23.5	22.9	23.4

Overall model performance is described by the C-statistic. Additionally, we analyse the discrimination and calibration for the specific use case of predicting mortality at 90 days as measured by the AUC and ECI. We defined low risk as discharged patients with <5% 90-day mortality and high risk as discharged patients with >40% 90-day mortality. We used these thresholds as they identified potentially clinically significant thresholds. For example, low-risk patients may be targeted for early discharge, whereas high-risk patients may benefit from early outpatient follow-up or even hospice referral. The MELD and MELD-NA models failed to generate risk scores <5% for any patients; that is, they could not identify any very low-risk patients, and therefore sensitivity and PPV could not be calculated. Therefore, their classification errors were incalculable.

AUC, area under the curve; CLIF-C AD, CLIF Consortium acute decompensation score; ECI, Estimated Calibration Index; MELD, model for end-stage liver disease; MELD-Na, model for end-stage liver disease with sodium; n/a, not applicable; NPV, negative predictive value; PPV, positive predictive value.

patients as low risk compared with existing models while keeping false positives at <3%.

Compared with these traditional risk scores, opportunities exist for advanced clinical decision support using state-of-the-art models, potentially involving tens or hundreds of variables. Our database of over 250 000 hospitalisations allowed us to evaluate a wide range of predictors. More importantly the EHR allows automated calculation and integration of risk stratification into the clinical workflow for decision support. Similar to recent work on advanced predictive analytics, our model is not intended to be directly calculated by the clinician at the bedside, but instead automatically assessed by the EHR and compiled into dashboards for care coordination efforts.[40 41] For example, Amarasingham *et al*[42] reduced heart failure readmissions by 27% by integrating a complex, automated predictive model with the EHR.

Our analysis has limitations. Because the Veterans Health Administration care system is limited to American military veterans, our cohort largely comprises male patients and may not generalise to a population with a greater proportion of female patients. Second, the models we used were internally validated, and the generalisability of those models will need to be assessed through external validation in other populations. We used variables that are present in all EHRs and healthcare delivery environments, making generalisability more likely. Third, our model's predictor variables were extracted from the EHR and susceptible to coding errors and may have to be revalidated for ICD-10 codes. Fourth, the majority of VA patients seek care within the VA's integrated care system; however, there will be a small minority who are hospitalised at other facilities. However, of note, the VA clinical data warehouse does capture administrative claims data for outside hospitalisations if the VA acts as the payer.

In summary, this study identified a high mortality rate in patients with advanced liver disease. To our knowledge, this is the largest study predicting cirrhosis mortality using granular clinical data. Our model, one of the first employing a time-dependent covariate survival model for cirrhosis mortality, also allows predictions at any time point within our study's 4-year follow-up time frame. We demonstrate the promise of big data analytics over traditional risk measures and suggest opportunities for an EHR-derived risk algorithm that can help stratify patients to personalise care.

Author affiliations
[1]Veteran's Health Administration, VA Tennessee Valley Healthcare System, Nashville, Tennessee, USA
[2]Department of Medicine, University of California San Diego, La Jolla, California, USA
[3]UC San Diego Health Department of Biomedical Informatics, University of California San Diego, La Jolla, California, United States
[4]Department of Medicine, VA San Diego Healthcare System, San Diego, California, USA
[5]Department of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
[6]Department of Biostatistics and Medical Informatics, University of Wisconsin Madison, Madison, Wisconsin, USA
[7]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[8]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[9]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Twitter Jejo David Koola @JejoKoola

ORCID iD
Jejo David Koola http://orcid.org/0000-0001-5171-8475

REFERENCES

1 Murray CJL, Atkinson C, Bhalla K, *et al*. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA* 2013;310:591–606.
2 Lozano R, Naghavi M, Foreman K, *et al*. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 2012;380:2095–128.
3 Fleming KM, Aithal GP, Card TR, *et al*. All-Cause mortality in people with cirrhosis compared with the general population: a population-based cohort study. *Liver Int* 2012;32:79–84.
4 Kamath PS, Wiesner RH, Malinchoc M, *et al*. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001;33:464–70.
5 Kim WR, Biggins SW, Kremers WK, *et al*. Hyponatremia and mortality among patients on the liver-transplant waiting list. *N Engl J Med* 2008;359:1018–26.
6 Moreau R, Jalan R, Gines P, *et al*. Acute-On-Chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology* 2013;144:1426–37.
7 Jalan R, Pavesi M, Saliba F, *et al*. The CLIF Consortium acute decompensation score (CLIF-C ads) for prognosis of hospitalised cirrhotic patients without acute-on-chronic liver failure. *J Hepatol* 2015;62:831–40.
8 Jalan R, Saliba F, Pavesi M, *et al*. Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. *J Hepatol* 2014;61:1038–47.
9 Tandon P, Reddy KR, O'Leary JG, *et al*. A Karnofsky performance status-based score predicts death after hospital discharge in patients with cirrhosis. *Hepatology* 2017;65:217–24.

10 Peng Y, Qi X, Guo X. Child–Pugh versus MELD score for the assessment of prognosis in liver cirrhosis. *Medicine* 2016;95:e2877.

11 D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* 2006;44:217–31.

12 Moons KGM, Kengne AP, Grobbee DE, *et al*. Risk prediction models: II. external validation, model updating, and impact assessment. *Heart* 2012;98:691–8.

13 Davis SE, Lasko TA, Chen G, *et al*. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc* 2017;24:1052–61.

14 Amarasingham R, Patzer RE, Huesch M, *et al*. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff* 2014;33:1148–54.

15 Beste LA, Leipertz SL, Green PK, *et al*. Trends in burden of cirrhosis and hepatocellular carcinoma by underlying liver disease in US veterans, 2001-2013. *Gastroenterology* 2015;149:1471–82.

16 Heuman DM, Abou-Assi SG, Habib A, *et al*. Persistent ascites and low serum sodium identify patients with cirrhosis and low MELD scores who are at high risk for early death. *Hepatology* 2004;40:802–10.

17 Ahmad J, Downey KK, Akoad M, *et al*. Impact of the MELD score on waiting time and disease severity in liver transplantation in United States veterans. *Liver Transpl* 2007;13:1564–9.

18 Goldberg DS, French B, Forde KA, *et al*. Association of distance from a transplant center with access to waitlist placement, receipt of liver transplantation, and survival among US veterans. *JAMA* 2014;311:1234–43.

19 Kanwal F, Kramer JR, Buchanan P, *et al*. The quality of care provided to patients with cirrhosis and ascites in the Department of Veterans Affairs. *Gastroenterology* 2012;143:70–7.

20 Davila JA, Henderson L, Kramer JR, *et al*. Utilization of surveillance for hepatocellular carcinoma among hepatitis C virus-infected veterans in the United States. *Ann Intern Med* 2011;154:85–93.

21 Julapalli VR, Kramer JR, El-Serag HB, *et al*. Evaluation for liver transplantation: adherence to AASLD referral guidelines in a large Veterans Affairs center. *Liver Transpl* 2005;11:1370–8.

22 Coventry PA, Grande GE, Richards DA, *et al*. Prediction of appropriate timing of palliative care for older adults with non-malignant life-threatening disease: a systematic review. *Age Ageing* 2005;34:218–27.

23 Koola JD, Ho SB, Cao A, *et al*. Predicting 30-day Hospital readmission risk in a national cohort of patients with cirrhosis. *Dig Dis Sci* 2019;36.

24 Nehra MS, Ma Y, Clark C, *et al*. Use of administrative claims data for identifying patients with cirrhosis. *J Clin Gastroenterol* 2013;47:e50–4.

25 Brown S*et al*. VistA—U.S. Department of Veterans Affairs national-scale his. *Int J Med Inform* 2003;69:135–56.

26 Fihn SD, Francis J, Clancy C, *et al*. Insights from advanced analytics at the Veterans health administration. *Health Aff* 2014;33:1203–11.

27 Tibshirani R. The LASSO method for variable selection in the COX model. *Stat Med* 1997;16:385–95.

28 VA National Drug File - Data.gov. Available: https://catalog.data.gov/dataset/va-national-drug-file-may-2015 [Accessed 13 Jun 2017].

29 Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999;401:788–91.

30 Lin X, Boutros PC. NNLM: fast and versatile non-negative matrix factorization, 2016. Available: https://cran.r-project.org/web/packages/NNLM/index.html [Accessed 25 Apr 2017].

31 Therneau TM, Grambsch PM. *Modeling survival data: extending the COX model*. Springer Science & Business Media, 2013.

32 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.

33 Van Hoorde K, Van Huffel S, Timmerman D, *et al*. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *J Biomed Inform* 2015;54:283–93.

34 Pencina MJ, D'Agostino RB, D'Agostino RB, *et al*. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.

35 Kerr KF, Wang Z, Janes H, *et al*. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014;25:114–21.

36 Beyersmann J, Schumacher M. Time-Dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics* 2008;9:765–76.

37 Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996;17:343–6.

38 Salpeter SR, Luo EJ, Malter DS, *et al*. Systematic review of noncancer presentations with a median survival of 6 months or less. *Am J Med* 2012;125:512.e1–512.e16.

39 Cholongitas E, Marelli L, Shusang V, *et al*. A systematic review of the performance of the model for end-stage liver disease (MELD) in the setting of liver transplantation. *Liver Transpl* 2006;12:1049–61.

40 Kuzniewicz MW, Puopolo KM, Fischer A, *et al*. A quantitative, Risk-Based approach to the management of neonatal early-onset sepsis. *JAMA Pediatr* 2017;171:365–71.

41 Cronin PR, Greenwald JL, Crevensten GC, *et al*. Development and implementation of a real-time 30-day readmission predictive model. *AMIA Annu Symp Proc* 2014;2014:424–31.

42 Amarasingham R, Patel PC, Toto K, *et al*. Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study. *BMJ Qual Saf* 2013;22:998–1005.