“Beyond MELD” – Emerging strategies and technologies for improving mortality prediction, organ allocation and outcomes in liver transplantation

Jin Ge1, W. Ray Kim2,*, Jennifer C. Lai1, Allison J. Kwong2

Keywords: MELD; Prognostication; Allocation; Frailty; Sarcopenia; EHR; OMOP; Clinical Decision Support.

Received 2 December 2021; received in revised form 24 February 2022; accepted 4 March 2022

1Division of Gastroenterology and Hepatology, Department of Medicine, University of California – San Francisco, San Francisco, CA, USA; 2Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

• Corresponding author. Address: 430 Broadway Street, Floor 3, Redwood City, CA 94063-3132, USA. Tel.: +1-650-723-5135, fax: +1-650-723-5488.
E-mail address: wrkim@stanford.edu (W.R. Kim).
https://doi.org/10.1016/j.jhep.2022.03.003

Summary
In this review article, we discuss the model for end-stage liver disease (MELD) score and its dual purpose in general and transplant hepatology. As the landscape of liver disease and transplantation has evolved considerably since the advent of the MELD score, we summarise emerging concepts, methodologies, and technologies that may improve mortality prognostication in the future. Finally, we explore how these novel concepts and technologies may be incorporated into clinical practice.

© 2022 Published by Elsevier B.V. on behalf of European Association for the Study of the Liver.

Introduction
The deficit of available donor organs in relation to the number of patients in need of liver transplantation necessitates systems to allocate organs in an efficient yet equitable manner. The current principles of liver allocation in the United States,1 the Eurotransplant region,2,3 and elsewhere include determination of priority through objective and measurable medical criteria, ordered from most to least medically urgent.1,4 Urgency has been represented primarily by the model for end-stage liver disease (MELD) score, rather than the Child-Pugh score, to avoid subjective variables such as ascites and encephalopathy and to expand the scale (to reduce the number of candidates with identical scores).5,6

The MELD score, which is comprised of serum bilirubin, creatinine, and the international normalised ratio, has since served a dual purpose in general and transplant hepatology. It effectively predicts short-term (e.g., over 90 days) mortality among patients with chronic liver disease, thereby providing clinicians with a critical tool to prognosticate liver-related and waitlist mortality. It has been used to determine medical urgency (and hence priority) for liver transplant candidates since 2002 in the United States and 2006 in the Eurotransplant region, making it an essential tool for transparent and equitable organ allocation.7,8

The landscape of chronic liver disease and liver transplantation has evolved considerably in the last two decades. Both waitlist mortality prediction and transplant organ allocation require ongoing re-evaluation to ensure accurate prognostication and appropriate distribution of donor organs. In 2016, the MELD score was updated to include serum sodium, an objective biomarker that is often a surrogate indicator for ascites.9 A new update to and recalibration of the MELD score, MELD 3.0, was recently published with the inclusion of sex and serum albumin.10 At the same time, a substantial proportion of liver transplants are allocated by MELD “exception”, representing indications where the mortality risk and need for transplant are not well-represented by the MELD score.11

In addition, emerging technologies, new methodologies, and evolving conceptual frameworks for liver disease may improve clinicians’ ability to prognosticate and manage patients with end-stage liver disease. In this article, we present emerging tools and techniques “beyond MELD” for improvement in liver allocation, prognostication, and outcomes in patients with end-stage liver disease.

Beyond MELD – for liver allocation

Improving the MELD score
Over the past two decades of MELD score-based liver allocation, the demographics of chronic liver disease and indications for liver transplantation have changed dramatically worldwide. The widespread availability of effective direct-acting antiviral therapy for hepatitis C and the increasing prevalence of alcohol-associated liver disease and non-alcoholic steatohepatitis has fundamentally changed the population of patients awaiting liver transplantation.11,12 Throughout these changes, however, the MELD score has continued to provide robust predictions of short-term waitlist mortality that outperform most other clinical scoring systems, with c-statistics that exceed 0.80 in various cohorts.9,10,13 Still, it has been perceived that the predictive power of the MELD score may have diminished in recent years.14,15 The MELD score may not represent mortality risk as accurately for
patients with some of the most severe clinical complications of cirrhosis, such as acute-on-chronic liver failure (ACLF), refractory ascites/hepatic hydrothorax, recurrent variceal bleeding, and hepatocellular carcinoma.\textsuperscript{14,16}

In addition, the MELD score has historically underpredicted mortality risks for women.\textsuperscript{17,18} This sex disparity is multifactorial but in part stems from the reliance of the MELD score on the measurement of serum creatinine, which can vary by sex for the same degree of renal dysfunction.\textsuperscript{12,19,20} Women on average have lower muscle mass compared to men, leading to systematic underestimation of renal function by serum creatinine.\textsuperscript{21} Alternatives to the creatinine component of the MELD score have been proposed, including MDRD (modification of diet in renal disease)\textsuperscript{18,22} and cystatin C (\textsuperscript{25,26} but are still less-than-ideal owing to the lack of improvement in model performance, inclusion of age and/or race-based equations, or clinical availability (cystatin C) (Table 1). The most recent iteration of the MELD score, MELD 3.0, incorporates sex as an independent variable to correct for the sex disparity due to creatinine, while also updating coefficients, adding serum albumin and adjusting the creatinine to a lower cap of 3.0 mg/dl.\textsuperscript{20} Other factors contributing to the sex disparity, including anthropometric differences and thus fewer opportunities for size-appropriate organs or the allocation of exception points, may require other types of adjustments to fully address the differences in outcomes and access to transplant between sexes.\textsuperscript{17,19,27,28}

While the MELD score remains a reliable indicator of mortality risk in liver disease, it can certainly benefit from further refinement. In so doing, the selection of variables should be carefully considered. Older age, medical co-morbidities, or certain aetiologies of liver disease may be associated with increased mortality risk, yet there is no consensus that these variables should influence waitlist priority or access to liver transplantation. Race may also be predictive, but this variable in clinical prediction scores can be problematic, as racial differences among populations in large datasets are often not genetic or biological, but rather reflect socioeconomics and healthcare policy.\textsuperscript{28} Race adjustment in these situations, while well-intentioned, can exacerbate inequity. Lastly, variables should be objective, verifiable, and readily available. Although addition of such variables may generate better prediction of waitlist mortality, they are not necessarily appropriate for use in organ allocation. Systems for organ distribution also need to be interpretable and transparent with regards to how changes of a specific variable would impact allocation.

**Emerging concepts to improve allocation**

The rationale behind organ allocation systems is to maximise the use of available organs and reduce deaths on the waiting list. Organ allocation may be driven by 3 important principles:

- **Urgency** – Allocation to the patient estimated to have the shortest survival without a transplant.
- **Utility** – Allocation to the patient estimated to have the longest post-transplant survival.
- **Transplant benefit** – Allocation based on the difference between the mean survival estimates with and without a transplant.

In the past two decades, liver allocation in the United States and parts of Europe has been based almost entirely on the principle of urgency – in other words, by risk of death as determined by the MELD score.\textsuperscript{78} Although the Final Rule instituted by the Department of Health and Human Services in the United States also provides for consideration of utility and survival benefit – to make the best use of donated organs, to avoid wasting organs, and to avoid futile transplants.\textsuperscript{1} However, acceptable standards and thresholds for post-transplant longevity and futility have been challenging to define,\textsuperscript{30} and current models for post-transplant survival do not perform well enough alone to be used in allocation.\textsuperscript{31–33} Moreover, the net benefit of liver transplant, defined by the difference between survival with and without transplant, is largely driven by waitlist mortality, where the candidates with the highest MELD score gained the most life-years from transplant.\textsuperscript{34,35}

In many MELD-based liver allocation systems, exception points grant waitlist priority and thus access to transplant for patients whose mortality risk and need for transplant is not well-represented by the MELD score, the most common exception being for hepatocellular carcinoma.\textsuperscript{36} Calibration of these exception points to approximate the mortality risk and urgency for transplant and to equitably allocate organs has turned out to be a moving target as patient characteristics and management of various conditions have shifted over time. Ensuring equitable allocation for this population may require additional solutions, including integration of transplant benefit and flexibility for donor-recipient matching in certain cases.\textsuperscript{37} For example, the United States allocation system does consider utility in the specific contexts of hepatocellular carcinoma or cholangiocarcinoma, by which patients exceeding certain criteria do not receive standard priority for liver transplant, owing to the excess risk of post-transplant recurrence and thus lower transplant benefit.\textsuperscript{39} Such rules may set a precedent for utility to be considered in future liver allocation policies.

Disparities in waitlist outcomes also arise from unequal access to transplant. Patients with the

---

**Key point**

While the MELD score remains a reliable indicator of mortality risk in liver disease, further refinements, exception points, and continuous distribution are required as we move toward truly fair and equitable organ allocation.
same medical urgency should have an equal opportunity of receiving a liver transplant, yet this is currently not the case. Upcoming changes in allocation in the United States include not only optimisation of the MELD score but also continuous distribution, a composite point scoring system that will enable the consideration of additional variables, including height, body surface area, blood type, geography, paediatric status, and travel efficiency, and indication for transplant (i.e. exceptions), to move closer to fair and equitable organ allocation. Under the proposed framework defined by the Organ Procurement and Transplantation Network (OPTN) in the United States, continuous distribution will attempt to balance 5 goals: medical urgency, post-transplant survival, candidate biology, patient access, and placement efficiency, although the specific attributes ultimately included and their respective weighting will depend on feedback from the transplant community and modelling and analysis. The system is envisioned to provide a more dynamic reflection of patient-related factors and thereby improve access.

Consensus processes, such as that described by the Italian liver transplant community, may help to develop allocation policy that fairly balances the various priorities of liver transplantation, including urgency, utility, and transplant benefit.

### Beyond MELD – For prognostication

#### Muscle dysfunction as a clinical marker for assessing disease severity in patients with cirrhosis

Emerging factors that have not classically been reflected by the MELD score, such as malnutrition, frailty, and sarcopenia, have improved our ability to dynamically characterise the morbidity and mortality associated with cirrhosis. Malnutrition represents a spectrum of nutritional deficiencies that cause adverse effects on physiologic function or clinical outcomes. It contributes to and is interdependent with measurable clinical manifestations of muscle dysfunction: frailty and sarcopenia.

---

**Table 1. Limitations of existing and proposed waitlist mortality risk scores to be used in liver allocation.**

<table>
<thead>
<tr>
<th>Score</th>
<th>Components</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child–Pugh score</td>
<td>Bilirubin, INR, albumin, ascites, encephalopathy</td>
<td>Established minimal listing criteria for liver transplant candidates</td>
<td>Inclusion of potentially subjective variables i.e. ascites and encephalopathy</td>
</tr>
<tr>
<td>MELD</td>
<td>Bilirubin, INR, creatinine</td>
<td>Adequate discriminative ability</td>
<td>Underestimation of renal dysfunction in women compared to men</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use of objective and widely available tests</td>
<td>Does not accurately represent transplant urgency for certain disease etiologies such as hepatocellular carcinoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improved waitlist mortality, equity in liver allocation</td>
<td>May not accurately represent mortality risk for complications such as hepatic encephalopathy or acute-on-chronic liver failure</td>
</tr>
<tr>
<td>MELD-Na</td>
<td>Bilirubin, INR, creatinine, sodium</td>
<td>Addition of sodium as a surrogate for ascites</td>
<td></td>
</tr>
<tr>
<td>MELD-Plus</td>
<td>Bilirubin, INR, creatinine, sodium, albumin, total cholesterol, WBC, age, length of stay</td>
<td>Improved mortality prediction compared to MELD-Na after hospital admission</td>
<td>Only calculated after a cirrhosis-related hospital admission</td>
</tr>
<tr>
<td>MELD-lactate</td>
<td>Bilirubin, INR, creatinine, sodium, lactate</td>
<td>Improved in-hospital mortality prediction compared to MELD or MELD-Na in patients hospitalised for infection or MELD ≤15</td>
<td>Only calculated during a hospital admission</td>
</tr>
<tr>
<td>MELD-Na-MDRD</td>
<td>Bilirubin, INR, creatinine, age, sex, race</td>
<td>More accurate estimation of renal function accounting for potential differences in muscle mass</td>
<td>Did not improve mortality prediction</td>
</tr>
<tr>
<td>MELD-GRAIL-Na</td>
<td>Bilirubin, INR, creatinine, blood urea nitrogen, age, sex, race, albumin, sodium</td>
<td>Estimation of renal function developed for liver disease with better accuracy and precision compared to standard eGFR calculations</td>
<td>Inclusion of age and race could lead to bias in allocation</td>
</tr>
<tr>
<td>MELD-Cystatin C</td>
<td>Bilirubin, INR, creatinine, cystatin C</td>
<td>Biomarker of renal function less susceptible to differences in muscle mass</td>
<td>Lack of clinical availability</td>
</tr>
<tr>
<td>MELD-Na-Shift</td>
<td>Bilirubin, INR, creatinine, sodium, albumin</td>
<td>Adds 0–1 MELD points for women</td>
<td>Addition of points for women at arbitrary levels</td>
</tr>
<tr>
<td>MELD 3.0</td>
<td>Bilirubin, INR, creatinine, sodium, sex, albumin</td>
<td>Addition of 1.33 points for women</td>
<td>Calculation somewhat more complex</td>
</tr>
</tbody>
</table>

eGFR, estimated glomerular filtration rate; GRAIL, glomerular filtration rate assessment in liver disease; INR, international normalised ratio; MDRD, modification of diet in renal disease; MELD, model for end-stage liver disease; WBC, white blood cell.
Frailty is classically defined as the clinical state of decreased physiologic reserve and increased vulnerability to health stressors. In patients with cirrhosis, this manifests as the phenotypic representation of impaired muscle contractile function. Frailty is estimated to be present in 17% to 43% of patients with cirrhosis based on different measurement standards; it worsens in patients with cirrhosis over time and has been strongly associated with waitlist and post-transplant mortality. For instance, frailty was associated with a nearly 2-fold higher adjusted risk of death in 1,044 ambulatory patients with cirrhosis awaiting liver transplantation in a multicentre study in the United States. Moreover, frailty is linked with increased healthcare utilisation both in the ambulatory and hospitalised settings. Given frailty’s strong association with post-transplant outcomes, the concept of “prehabilitation” or intervening to modify physical reserve prior to surgery has gained traction in both transplant and non-transplant surgical fields.

Sarcopenia is defined as the progressive and generalised loss of skeletal muscles associated with increased likelihood of adverse outcomes. Sarcopenia is also common in adults with cirrhosis, affecting 30% to 70% of patients with strong sex-based differences in prevalence. The gold standard for sarcopenia assessment is computed tomography imaging; since abdominal imaging is commonly performed for clinical reasons, muscle mass measurements are often obtainable. Sarcopenia has a robust association with waitlist mortality before and after transplantation, as well as with hepatic decompensation. Sarcopenia is progressive in patients with cirrhosis, and serial/longitudinal measures of muscle loss have been associated with clinical outcomes including waitlist mortality.

**Electronic health data and multicentre electronic consortia**

Recent advances in computing power in conjunction with the availability of large databases and analytical methodologies have dramatically increased the tools available for clinical research in hepatology. Historically, the predominant forms of large clinical research databases in the United States and Europe have been based on either patient registries, such as the Scientific Registry of Transplant Recipients or Eurotransplant databases, multicentre curated cohorts, or administrative claims databases. Beyond these large databases, there has been a growing movement towards aggregation of longitudinal electronic health records (EHRs) across multiple institutions and health systems.

In the United States and the European Union, EHRs now have greater than 96% penetration in acute care hospital and physicians’ offices. EHR data, gathered as the transactional record of health care delivery and operations, are now viewed as a key resource to generate unique insights. Novel applications of data science and clinical informatics on EHR data have the potential to accelerate clinical research and improve patient care. One of the key advantages of EHR data is its dynamic longitudinal nature with data acquisition occurring at every interaction that the patient has with the healthcare system. Correctly harnessed, integration of longitudinal data could produce more comprehensive reflections of patients’ clinical trajectory.

For instance, incorporation of time-variant variables, such as laboratory values and vital signs, captured in EHRs have enabled continuous prediction of the development of acute kidney injury during inpatient admissions. Moreover, the use of longitudinal and sequential data elements gathered from EHR flowsheets, medication administrations, physician notes, and radiology reports have enabled the construction of deep-learning models to more accurately predict inhospital mortality, 30-day readmissions, and prolonged length of stay. In clinical hepatology, the integration of longitudinal EHR elements, such as structured flowsheet entries, medication administration, procedure orders, vital signs, and laboratory values, has enabled dynamic calculations of the North American Consortium for the Study of End-Stage Liver Disease-ACLF and Chronic Liver Failure Consortium-ACLF prognostication scores in hospitalised patients with ACLF.

Despite the potential for longitudinal EHR data to improve outcome prediction, the lack of standards, lack of semantic interoperability, and disparate EHR systems/implementations have historically limited large multi-institution collaborations. Early regional-based EHR consortia, such as HealthLNK based in the Chicago area, have demonstrated the value of multicentre EHR data in predicting factors associated with mortality in patients with cirrhosis.

The development and wider availability of common data models, such as the observational medical outcomes partnership (OMOP) model and the fast healthcare interoperability resources (FHIR) model, may now facilitate larger EHR-based collaboratives. Examples of such large EHR-based research collaboratives include the Observational Health Data Sciences and Informatics group based in the United States and the European Health Data and Evidence Network based in the
European Union. While the trend towards common data models and centralised EHR data for observational research had already been underway, the COVID–19 pandemic drastically accelerated this movement with the creation of the National COVID Cohort Collaborative (N3C)

N3C is a novel, centralised, and harmonised repository of EHR data from more than 64 sites from across the United States built on the OMOP platform, formed in response to the need for rapid accrual and analyses of clinical data during the COVID–19 pandemic. Its effective use has allowed for the rapid generation of insights into the mortality risk of SARS-CoV-2 infection among patients with cirrhosis. The work highlights the prospect of transplant hepatology-specific multicentre EHR collaboratives with deep clinical content expertise, which may accelerate the development of comprehensive models for mortality prediction in patients with end-stage liver disease.

**Novel modelling methodologies for mortality risk prediction**

While high-dimension multicentre EHR data has tremendous potential, their “big data” nature may require the use of novel analytical techniques. “Big data” is an amorphous term that is classically defined in terms of the 5 “Vs” (volume, velocity, variety, veracity, and value) to describe large datasets that may be more effectively analysed using artificial intelligence-based methods, such as machine learning (ML), which permit data-driven rather than hypothesis-driven discovery. The most prevalent ML algorithms are divided into supervised (classification) and unsupervised (sorting) methods (Table 2).

There is often some overlap between traditional statistical and ML approaches: Logistic regression is such an example of a methodology common to both. In general, classification trees and neural network-based methods have generally been the predominant ML algorithms applied to contemporary hepatology research. The cirrhosis mortality model, developed from the United States Veterans Affairs Corporate Data Warehouse (VHACDW) using a combination of gradient boosting and logistic regression methods, offered significantly improved discrimination compared to the MELD score. Of particular interest are artificial neural networks (ANNs), which are learning algorithms that can be employed for both supervised and unsupervised tasks. Neural networks are inspired by neuroanatomy—each neuron is a computing unit, and all neurons are connected to build a network. Signals travel from input layer to the output layer going through multiple hidden layers—which represent higher complexity. Deep neural networks, characterised by multiple layers between the input and output layers, have been utilised for longitudinal analyses of EHR data to predict outcomes of cirrhosis.

In liver transplant, ML methodologies have been used to explore waitlist mortality and organ allocation. One of the first ML models in transplant hepatology developed in 2003 was an ANN model to predict 1-year mortality in a cohort of 92 patients. While limited in scale, this ANN model outperformed logistic regression and the Child-Pugh score. Similarly, an ANN-based mortality model derived from patients awaiting liver transplantation in Italy and validated in the United Kingdom showed better predictive ability than the original MELD score. The optimised prediction of mortality model—developed in 2019 and trained on OPTN data using ML optimal classification trees—demonstrated superior mortality prediction vs. the MELD score, and led to decreased mortality and increased survival benefit across all candidate demographics, diagnoses, and geographic regions in liver simulated allocation model simulations.

Despite these encouraging results, ML models for waitlist mortality have several limitations, including interoperability and complexity. In addition, many early applications of ML methodologies have only considered binary outcomes rather than a time-dependent survival function which is key in the accurate determination of transplant urgency and waitlist priority. Due to these limitations and challenges in practical implementation, waitlist mortality models based on ML have yet to gain much traction in organ allocation. ML models have the potential to better predict post-transplant outcomes through the real-time considerations of longitudinal candidate variables, donor variables, and the interaction of donor-candidate matching, which may play a role in continuous distribution.

**Potential pitfalls of algorithms for clinical prediction**

While there is substantial potential for ML to influence clinical practice in transplant hepatology and potentially improve patient outcomes, limitations of these technologies must be recognised. First, additional complexity may not improve predictive performance if underlying data and variables are the same. When comparing the ability of ML models (support vector classification and random forest) vs. logistic regression to predict readmission and death in 2,179 North American patients with ACLF, ML model accuracies were equivalent to models generated using only the MELD score. The performance of future ML modelling may improve if higher density data incorporating novel variables, such as sarcopenia and frailty, are available.

Second, despite harmonisation and rationalisation of different ontologies and semantics, data quality, shift, and reproducibility are still ongoing.
issues in the modelling of EHR data. Dataset shift describes the changes in model performance due to temporal or spatial shifts between the population used for training and the population upon which the algorithm is deployed. One prominent recent example is the University of Michigan’s deactivation of a proprietary sepsis-alert model due to shifts in patient populations during the COVID-19 pandemic. Dataset shift is not exclusive to ML algorithms but also to other clinical prediction scoring systems. Periodic audits and updating of scoring systems, such as the update of MELD to MELD 3.0, are necessary to adapt our clinical tools to changing conditions.

Third, underlying bias can be amplified by clinical prediction and ML-based algorithms. The most prominent example in transplantation is the incorporation of race in estimated glomerular filtration rate (eGFR) calculations, which have disadvantaged racial minorities in listing practices and allocation for kidney transplant. In transplant hepatology, eGFR has been avoided in clinical prognostication modelling due to its potential for exacerbating race- and sex-based disparities. Human intelligence, in addition to artificial intelligence, remains critically important for the thoughtful and deliberate selection of data features, variables and analytic methodologies.

Fourth, structured data, which forms the basis for most classical models and ML algorithms at this time, are limited by coding. For example, efforts to diagnose Fontan-associated liver disease were limited by the lack of specific structured diagnostic codes across multiple clinical databases. The volume of unstructured data far exceeds structured data, with an estimated 90% of digital data in healthcare being unstructured. Incorporating or converting unstructured data elements in the EHR, such as imaging reports, pathology reports, and clinical documentation, into structured or tagged...

### Table 2. Common machine learning algorithms used in clinical research.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Summary</th>
<th>Application example in hepatology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supervised learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear regression</td>
<td>Relationship modelling between a response variable and one or more explanatory variables</td>
<td>Prediction of liver fat fraction from the presence of metabolic syndrome, type 2 diabetes, and laboratory markers</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Prediction of the probability of a target variable in binary classification</td>
<td>Prediction of 30-day readmissions for acute-on-chronic liver failure patients</td>
</tr>
<tr>
<td>Decision tree</td>
<td>Classification or regression of data based on simple rules splitting values of input variables</td>
<td>Prediction of acute kidney injury after liver transplantation utilising scoring systems</td>
</tr>
<tr>
<td>Random forest</td>
<td>Ensemble of multiple decision trees operating as a committee</td>
<td>Personalised surveillance model for development of hepatocellular carcinoma in patients with hepatitis C cirrhosis</td>
</tr>
<tr>
<td>Gradient boosted trees</td>
<td>Ensemble method of building weaker prediction models sequentially where each model predicts leftover error</td>
<td>Risk stratification of mortality for patients with cirrhosis in the United States Veteran Health Administration</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>Linear classification by finding the hyperplane that maximises the margins between 2 classes</td>
<td>Prediction of 30-day readmissions for acute-on-chronic liver failure patients</td>
</tr>
<tr>
<td>K-Nearest neighbor</td>
<td>Classification of new data or cases based on similarity or distance between input features</td>
<td>Identification of molecular signature associated with development of hepatocellular carcinoma</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>Use of Bayes theorem to predict membership probability assuming independence among predictors</td>
<td>Prediction of hepatitis B cirrhosis utilising serum biomarkers</td>
</tr>
<tr>
<td><strong>Unsupervised learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-Means</td>
<td>Partition observations into k clusters in each observation belongs to the cluster with nearest center</td>
<td>Classification of cirrhosis based on un-labelled MRI data</td>
</tr>
<tr>
<td>Principal component analysis</td>
<td>Reduce dimensionality by converting correlated variables into a set of uncorrelated variables</td>
<td>Identification of splanchnic and clinical characteristics associated with hyperdynamic circulation in patients with cirrhosis</td>
</tr>
<tr>
<td>Gaussian mixture</td>
<td>Probabilistic model that assumes all data are generated from a finite set of Gaussian distributions</td>
<td>Detection of hepatocellular carcinoma from computed tomography images</td>
</tr>
<tr>
<td>Hidden Markov</td>
<td>System is assumed to be a Markov model with unobservable states</td>
<td>Progression from cirrhosis to hepatocellular carcinoma based on clinical covariates and diagnostic codes</td>
</tr>
<tr>
<td><strong>Neural network algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artificial neural networks</td>
<td>Group of interconnected nodes/computing units that form a network</td>
<td>Quantification of skeletal muscle mass from computed tomography scans</td>
</tr>
<tr>
<td>Convolutional neural networks</td>
<td>Neural network with nodes designed to resemble visual cortices</td>
<td>Prediction of hepatocellular carcinoma development among patients with hepatitis C cirrhosis</td>
</tr>
<tr>
<td>Recurrent neural networks</td>
<td>Neural network where connections between nodes are based on temporal sequences</td>
<td>Prediction of 1-year mortality in patients with cirrhosis utilising EHR data</td>
</tr>
<tr>
<td>Deep neural networks</td>
<td>Multiple layers between the input and output layers</td>
<td>Longitudinal analyses of EHR data elements to predict hospitalisation outcomes</td>
</tr>
</tbody>
</table>

EHR, electronic health record.
features remains challenging. Transformation of such data into structured data requires substantial cleaning, splitting, merging, validating, and sorting, but does improve clinical representation in predictive analytics.\textsuperscript{109}

Finally, algorithms are not anticipated to completely replace the “subjective” judgment of clinicians involved in the care of the peri-transplant patient.\textsuperscript{110} For instance, significant technical expertise is required to conduct split liver transplantation,\textsuperscript{111} to use donor organs with technical variants or higher risk features,\textsuperscript{112} or to successfully transplant patients with complex surgical histories.\textsuperscript{113} These institution- and clinician-specific knowledge and skills are often ill-captured and ill-evaluated by algorithms.

For these reasons, the application of ML-based artificial intelligence has received a mixed reception from both clinicians and the general population.\textsuperscript{114–116} Among clinicians, there are latent fears that algorithms may ultimately replace their skills or functions.\textsuperscript{116,117} In addition, many clinicians are uncomfortable with “black box” ML tools, even though examples of similar opacity abound in other diagnostic and therapeutic areas of clinical medicine.\textsuperscript{118} Among providers and patients, there is a concern about the loss of patient-provider relationships, privacy in data use, and accountability – namely who is responsible for adverse outcomes due to clinical decisions influenced or augmented by artificial intelligence.\textsuperscript{114,115,119} There is an increasing recognition that transparency, interpretability, and explainability are necessary for long-term acceptance of artificial intelligence tools. Ante hoc systems, which are interpretable by design, and post hoc systems, which provide local and reproducible explanations for algorithm outputs, are now commonly utilised to enable greater trust in ML algorithms.\textsuperscript{116,120} Similarly, active incorporation of human knowledge, or expert-augmentation, in the algorithm construction process is another strategy to improve “explainability.”\textsuperscript{121} To begin to address these concerns, the development of standardised tools and evaluations on transplant reporting and assessments of bias in applied ML techniques is currently underway.\textsuperscript{102,122}

**Beyond MELD – for improvement in patient outcomes**

Emerging technologies to actively manage waitlist mortality risk

One technology to overcome issues with unstructured data is natural language processing (NLP), which is a suite of related techniques to convert unstructured or narrative text into tagged or structured elements for analysis.\textsuperscript{123,124} There has been particular interest in utilising NLP for the diagnosis of non-alcoholic fatty liver disease as this condition is poorly documented in structured EHR data.\textsuperscript{125,126} NLP has been used on abdominal ultrasound, computerised tomography, and magnetic resonance imaging reports from the VHACDW to rapidly screen patients with radiographic evidence of fatty liver disease.\textsuperscript{126} In an analysis of clinical notes available for 38,575 patients enrolled in the Mount Sinai BioMe cohort, NLP methods outperformed ICD codes and text search.\textsuperscript{125}

Real-time clinical decision support (CDS) and prospective risk modelling are also emerging areas of research/implementation in the management of

---

**Key point**

There is an increasing push to develop data-driven machine learning-based algorithms to further improve outcome prediction in patients with liver disease.

---

**Key point**

Clinical decision support and prospective risk modelling are emerging areas of research that are hoped to lead to improvements in the management of patients with cirrhosis and those on the liver transplant waiting list.

---

**Fig. 1. Rapid-cycle testing in ‘electronic’ randomised controlled trials.** Schematic of rapid-cycle ‘electronic’ randomised controlled trials could be implemented using CDS systems: computational risk assessment allows a patient to be randomised for an intervention associated with a CDS, the results of which could then be used to iteratively modify the risk stratification algorithm or the CDS system. CDS, clinical decision support.
patients with cirrhosis. Simple decision support tools have been implemented to support targeted quality improvement efforts, such as the proper use of ceruloplasmin in liver disease evaluation, improving hepatitis C screening, and albumin utilisation. The substitutable medical applications and reusable technologies on FHIR (SMART-on-FHIR) application programming interface allows for the development of more complex and prospective CDS systems by securely and automatically pulling in relevant patient data from disparate locations in the EHR. Previous SMART-on-FHIR CDS applications created to support the American Academic of Pediatrics guideline on management of neonatal hyperbilirubinemia showed to have excellent usability and improved order rates for clinically appropriate phototherapy. SMART-on-FHIR CDS applications have yet to be widely pilot tested or implemented in the care of patients with cirrhosis.

Potential applications of encounter-level CDS include improving adherence to guideline-recommended care in cirrhosis, prompting timely intervention before anticipated/forecasted clinical decompensation, or aiding immunosuppression surveillance in the post-transplant setting. On a patient or precision-level, CDS could allow for the calculation of “personalised” risk models for progression of fibrosis to cirrhosis, development of hepatocellular carcinoma, and risk of waitlist dropout. The use of these models and CDS systems may help inform decisions surrounding organ allocation and acceptance in the future. Prospective implementation of such CDS systems could allow for real-world “electronic” experiments or clinical trials. These concepts remain unexplored in chronic liver disease and liver transplantation, but may generate significant real-world evidence that could be used to optimise organ allocation and reduce waitlist mortality.

**Conclusions**

While the demographics and epidemiology of chronic liver diseases have changed dramatically in the past two decades, the MELD score and its successors have continued to provide robust predictions of short-term waitlist mortality. Continued refinements of the MELD score, such as MELD 3.0, improve its predictive ability and actively address deficiencies such as sex-based differences in waitlist mortality. Continuous distribution has emerged as a conceptual framework to optimise organ allocation by weighing factors beyond waitlist mortality. The selection of variables for changes to the liver allocation system, however, remains fraught with challenges, requiring careful consideration of objectivity, verifiability, and availability.

In the management of patients with cirrhosis and hepatic decompensation, more accurate, comprehensive, and real-time prediction of mortality, based on availability of the large amounts of information in EHRs, has the potential to dramatically change how we approach the clinical care of patients with cirrhosis and its complications. In addition, novel concepts and emerging technologies may play a major role in refining mortality prediction in an individual patient. For example, the prognosis of a patient with cirrhosis may be accurately assessed by deep neural network-based algorithms incorporating past clinical data in the EHR, current MELD 3.0, frailty measurements, and muscle mass volume derived from a computed tomography scan on an integrated SMART-on-FHIR application in the EHR system. We hope that, sometime in the near future, these novel tools will provide clinically actionable information to alter a patient’s outcome, well beyond determining a patient’s priority ranking for liver allocation.

**Abbreviations**

ACLF, acute-on-chronic liver failure; ANNs, artificial neural networks; CDS, clinical decision support; eGFR, estimated glomerular filtration rate; EHR, electronic health record; FHIR, fast healthcare interoperability resources; GRAIL, glomerular filtration rate assessment in liver disease; MELD, model for end-stage liver disease; MDRD, modification of diet in renal disease; ML, machine learning; N3C, National COVID Cohort Collaborative; NLP, natural language processing; OMOP, observational medical outcomes partnership; OPTN, Organ Procurement and Transplantation Network; SMART-on-FHIR, substitutable medical applications and reusable technologies on fast health interoperability resources; VHACDW, Veterans Affairs Corporate Data Warehouse.

**Financial support**

The authors of this study were supported by the AASLD Anna S. Lok Advanced/Transplant Hepatology Award (AASLD Foundation, Ge), P30DK026743 (UCSF Liver Center Grant, Ge and Lai), R01DK127224 (National Institute of Diabetes and Digestive and Kidney Diseases, Kim), R01AG059183 (National Institute on Aging, Lai), the Clinical, Translational, and Outcomes Research Award (AASLD Foundation, Kwong), and K23AA029197 (National Institute on Alcohol Abuse and Alcoholism, Kwong). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or any other funding agencies. The funding agencies played no role in the analysis of the data or the preparation of this manuscript.
Conflict of interest
The authors of this manuscript have no conflicts of interest to disclose as described by the Journal of Hepatology.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors’ contributions
Authorship was determined using ICMJE recommendations. Ge: Drafting of manuscript; critical revision of the manuscript for important intellectual content. Kim: Drafting of manuscript; critical revision of the manuscript for important intellectual content. Kwong: Drafting of manuscript; critical revision of the manuscript for important intellectual content.

Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2022.03.003.

References
Author names in bold designate shared co-first authorship


