The Model for End-Stage Liver Disease (MELD) was initially created to predict survival in patients with complications of portal hypertension undergoing elective placement of transjugular intrahepatic portosystemic shunts. The MELD which uses only objective variables was validated subsequently as an accurate predictor of survival among different populations of patients with advanced liver disease. The major use of the MELD score has been in allocation of organs for liver transplantation. However, the MELD score has also been shown to predict survival in patients with cirrhosis who have infections, variceal bleeding, as well as in patients with fulminant hepatic failure and alcoholic hepatitis. MELD may be used in selection of patients for surgery other than liver transplantation and in determining optimal treatment for patients with hepatocellular carcinoma who are not candidates for liver transplantation. Despite the many advantages of the MELD score, there are approximately 15%-20% of patients whose survival cannot be accurately predicted by the MELD score. It is possible that the addition of variables that are better determinants of liver and renal function may improve the predictive accuracy of the model. Efforts at further refinement and validation of the MELD score will continue. (HEPATOLOGY 2007;45:797-805.)

February 27, 2007 marked the fifth anniversary of the Model for End-Stage Liver Disease (MELD) becoming the standard by which prioritizes in donor liver allocation were determined. Since the score was first derived in a relatively small number of patients undergoing the transjugular intrahepatic portosystemic shunts (TIPS) procedure, it has been validated in many different populations of patients with liver disease. Within a relatively short period of time, MELD became a common metric by which the severity of liver disease could be accurately described.

In this paper, we review the initial development and validation of the MELD score, its application in organ allocation and management of patients with a variety of liver conditions, its strengths and limitation, and current and future efforts to refine and improve it further.

Creation and Validation of MELD

MELD was initially created to predict survival following elective placement of TIPS.1 The model was subsequently validated as a predictor of survival in several cohorts of patients with varying levels of liver disease severity (e.g., hospitalized and ambulatory patients), as well as patients of geographically and temporally diverse origin.2 The survival model was initially termed the “Mayo End-Stage Liver Disease” or “MELD” model to acknowledge the affiliation of the investigators who created the model. During discussions leading to the establishment of MELD as the basis for prioritization of organs for liver transplantation,3,4 we changed the name to “Model for End-Stage Liver Disease” which maintained the acronym “MELD”, but removed the association with a particular institution, a process that was thought would lead to wider acceptance of the model.

MELD incorporates 3 widely available laboratory variables including the international normalized ratio (INR), serum creatinine, and serum bilirubin. The original mathematical formula for MELD is: MELD = 9.57 × log₅(INR) + 3.78 × log₅(total bilirubin) + 11.2 × log₅(creatinine) + 6.43.

The score can be calculated on handheld computing devices, and is available at www.mayoclinic.org/gi-rst/mayomodel5.html. When the model was initially created the etiology of cirrhosis was also included. In the TIPS
population, the etiology of liver disease was important in that patients with alcoholic liver disease and cholestatic liver disease undergoing TIPS procedures had a better survival than those patients with viral hepatitis-related or other causes of cirrhosis.

In subsequent studies, we confirmed that the etiology of cirrhosis was a less important variable in determining survival in other patient cohorts with end-stage liver disease. Therefore, etiology of liver disease was removed as a variable from the model while still preserving its accuracy. The advantage of dropping etiology of cirrhosis as a variable was that the subjective element in determining etiology was removed, and the model could be based purely on objective laboratory variables.

MELD has been validated as a predictor of survival in independent groups of patients with a wide variety of liver diseases. In these studies, the accuracy of MELD was evaluated by its ability to rank patients according to risk for mortality determined by the “concordance statistic” or the “c” statistic. For example, a c-statistic of 0.7 indicates that patients with a higher MELD score will die earlier than patients with a lower MELD score 7 of 10 times. A c-statistic of 0.7 is thought to have reasonable clinical utility, while a c-statistic of ≥0.8 in a prediction model lends strong support to its accuracy. Most studies that evaluated MELD to rank patients according to their risk of mortality have yielded “c”-statistics upwards of 0.8, and usually superior to the Child-Turcotte-Pugh (CTP) class. Addition of complications such as ascites, encephalopathy, variceal bleed, and SPB do not improve MELD significantly; quantitative tests of liver function are also not superior to MELD in predicting survival.

Application of MELD

MELD in Liver Transplantation

Prior to February 27, 2002, patients were prioritized for receiving organs for liver transplantation based on their United Network of Organ Sharing (UNOS) status, a reflection of their CTP score. Given that the waiting list for liver transplantation approached 20,000 patients, and there were only 3 categories on the waiting list for patients with cirrhosis, namely Status 2A, Status 2B, and Status 3, time spent on the waiting list became the major determinant of who would receive a liver transplant. Therefore, this policy placed at a disadvantage patients who were at a high risk for mortality but who were listed late in their disease course and had not accrued enough waiting time.

In 1998, the Institute of Medicine decreed that a new allocation policy be put in place based on objective variables and which de-emphasized waiting time, that is, the “sickest first” policy. This led to MELD which is based on objective variables and could accurately rank patients with cirrhosis according to risk of mortality, replacing the then current CTP-based organ allocation system. In applying MELD to organ allocation, UNOS made several changes to how the MELD score was to be calculated. The lower limit for serum creatinine, serum bilirubin, and INR was fixed at 1 so that there would be no negative scores; the upper limit of serum creatinine was capped at 4 mg/dl.

Implementation of MELD led to an immediate reduction in liver transplant waiting list registrations for the first time in history of liver transplantation (12% decrease in 2002) because accrual of waiting time was no longer necessary. Accurate prediction of short-term mortality in the vast majority (83%-87%) of wait-listed candidates led to a reduction of almost 15% in the mortality on the waiting list. The number of deaths of patients on the wait list increased up to 2001 (Fig. 1); since implementation of MELD in 2002, the number of deaths showed a substantial decrease from 2046 in 2001 to 1364 in 2005. Although this reduction in mortality is in part attributable to a modest increase in available organs (4,671 in 2001 versus 5,160 in 2005), there is a wide consensus that MELD has made a significant contribution to reducing the mortality on the waiting list. A recent analysis showed that the reduction in mortality occurred only among patients with chronic liver disease (in whom MELD is used to allocate organs), but not among patients with fulminant liver disease in status 1 (in whom MELD is not used), suggesting that at least part of the decrease in waitlist mortality may be attributed to MELD-based organ allocation. Moreover, the median waiting time to liver transplantation decreased from 656 days to 416 days in the MELD era. Several countries have replaced the CTP score with MELD to rank patients according to...
mortality risk.14 In a study from Australia, clinical judgment, which is often used in centers to determine which patient should rank higher on a waiting list for mortality, has been proven to be inferior to the MELD score in determining survival.15

In contrast to the clear benefit of accurately estimating mortality on the waiting list, MELD has not been found to be as useful in predicting mortality following liver transplantation.16-19 Mortality in the post-transplantation period is related not only to the degree of liver dysfunction prior to transplantation, but to other factors, such as donor characteristics, experience of the transplantation team, and random postoperative complications which cannot be predicted. Moreover, patient selection by physicians will tend to negate the effect of pre-transplant MELD on post transplant survival. Consequently, almost all models which attempt to determine post transplant survival are not clinically useful. Thus, MELD score before liver transplant is not predictive of post liver transplant outcome because of relatively poor correlation between pre-transplant disease severity and post transplant outcome.

Likewise, in small studies on living donor liver transplantation, pre-transplant MELD scores have had little impact on post transplant survival.20 There have been proposals that organ allocation take into account donor and recipient factors simultaneously such that marginal grafts not be used for patients with high MELD scores in the hopes to improve post liver transplantation outcome.21,22 Similarly, in patients with MELD scores ≤14, mortality with transplantation was found to be higher than that of patients with the same MELD score not transplanted.23 One must exercise caution, however, in interpreting these results, because they are simply observational (as opposed to randomized) data.

To the degree that disease severity before liver transplantation affects post transplantation morbidity and complications, healthcare resources used correlate with pre-transplantation MELD score. At the level of individual patient, resource utilization, as judged by days in intensive care, red cell transfusions, and duration of hospitalization was higher with higher pre-transplant MELD scores.24,25 On the other hand, on an aggregate level, MELD-based organ allocation has not increased healthcare resource utilization. Based on a nationally representative hospital utilization database, we have shown that resources used did not increase since MELD was implemented.26 Thus, MELD-based allocation consistently directs organs to the sickest patients so that those patients in the previous era who had high MELD but insufficient waiting time are undergoing transplantation earlier, reducing the number of “outliers” that incurred astronomical costs to the health care system.

Patients undergoing retransplantation are at a higher risk of mortality after transplantation than those undergoing primary liver transplantation. The 2-year survival after retransplantation is lower than that after primary liver transplantation, with the difference in survival being greatest in patients with MELD >25. Moreover, retransplantation more than 2 years after the primary transplantation is associated with poor survival independent of the MELD score.27 These data notwithstanding, the most immediate question about MELD in allocation for retransplantation is whether the risk of death on waiting list is different between primary and repeated liver transplant candidates. Retransplantation candidates are registered on the waiting list with significantly higher MELD than primary liver transplant candidates (22 versus 14) and thus retransplantation candidates as a whole experienced higher mortality rate on the waiting list.

It is important to note that although the modifications created by UNOS in calculating the MELD score have some rationale, the modifications have been empirical rather than based on validated studies. For instance, the upper value of serum creatinine has been capped at as high as 4 mg/dl, allowing inclusion of patients with intrinsic renal disease. Thus, there has been a higher rate of combined kidney and liver transplants than previously, but without significant change in survival.28 The need for renal replacement therapy is higher if MELD scores are greater than 24,29 even though there is a suggestion that the prevalence of chronic renal disease up to 2 years after transplantation has not increased.30

Another area which requires more validation is the assignment of MELD scores to patients with hepatocellular carcinoma. The initial recommendation for allocation of 24 MELD points for patients with Stage 2 HCC was found to be too high, and this was later reduced to 22.31 It remains to be seen whether this reduction in score is more accurately reflective of the patients with hepatocellular carcinoma removed from the waiting list because they exceeded the Milan criteria.

Prediction of Long-term Survival in Patients with Cirrhosis

In a recent systematic review of 118 studies outlining the natural history and prognostic indicators of survival in cirrhosis,32 MELD and CTP score were recognized as predictors of long-term survival in patients with decompensated cirrhosis. Recently, both MELD and HVPG were shown to be independent predictors of survival. The “c” statistic for predicting survival was 0.71 for the MELD score, and the addition of HVPG and age in-
increased the predictive mortality to 0.76, which was not statistically significant. MELD is an accurate predictor of long-term mortality, even in patients with chronic hepatitis B, with or without cirrhosis.33 Whereas patients with NASH-related cirrhosis had a lower risk of mortality than patients with HCV-related cirrhosis, even when “decompensated”; MELD was an accurate predictor of survival in both groups of patients.34

**Application of MELD Beyond ESLD**

**Variceal Bleeding.** Retrospective studies have shown that both MELD and CTP are accurate predictors of survival in patients with variceal bleeding.35 MELD and hepatocellular carcinoma were better predictors of survival than CTP score in a study of 172 patients with cirrhosis and first variceal bleed, with MELD score discriminating between patients at high risk (≥15) and low risk (<15) for mortality.36 Our own unpublished data confirms the superiority of MELD over the CTP score in determining mortality following a variceal bleed.

**Infections in Patients with Cirrhosis.** Both studies addressing the predictability of MELD in determining mortality in patients with infections leading to renal failure have demonstrated that the only predictors of survival in these patients were the MELD score, and the type of hepatorenal syndrome; CTP score was not independently predictive of mortality.37,38 These data would suggest that patients with a high MELD score are either more likely to get infected, or that patients with a high MELD score are more likely to die as a result of spontaneous bacterial peritonitis (SBP) because outpatients with SBP typically have lower MELD scores than inpatients with SBP, and better survival.39

**Fulminant Hepatic Failure.** Among 312 patients with non-acetaminophen-induced FHF in the UNOS database, MELD was a highly significant predictor of 30-day mortality (P < 0.0001).40 Of note, patients with primary graft nonfunction and patients with hepatic artery thrombosis had a lower risk of 30-day mortality than patients with FHF and did not show a significant association between transplantation and survival. On the other hand, patients with FHF experienced the greater benefit with liver transplantation, survival improving from 58% at 30 days without transplantation to 91% with liver transplantation (P < 0.0001). This raises the question as to whether MELD should also be used to prioritize candidates with FHF, that is, UNOS Status 1, for liver transplantation.

In a large prospective study from Denmark, the utility of MELD in determining the onset of FHF in patients with acetaminophen-induced liver failure was demonstrated (“c” statistic: 0.92). However, once patients developed FHF, MELD was less accurate in predicting mortality.41 In a U.S. multicenter study in patients with FHF secondary to hepatitis A, MELD had a “c” statistic of only 0.7 in ranking patients according to mortality risk.42 These studies would suggest that, in patients with FHF in addition to the MELD score, there are other variables such as the degree of intracranial pressure important in predicting mortality. The U.S. study demonstrated that a model using serum creatinine, serum bilirubin, need for endotracheal intubation, and vasopressor support had a “c” statistic of 0.89. However, there were only 4 deaths, and only 9 of the 29 patients underwent liver transplantation. Therefore, the limited number of endpoints for a 4 variable model (typically, 40 events would be necessary for internal validation of a 4 variable model) is likely to be associated with “model over-fitting”, and needs validation in an independent data set before being widely used.

**Alcoholic Hepatitis.** In all 3 studies that compared the MELD score to the Maddrey score,43 MELD was a more accurate predictor of mortality.44-46 This is not surprising because the MELD score has both the variables included in the Maddrey score, namely the prothrombin time and bilirubin but, in addition, has an index of renal dysfunction which is the serum creatinine. The cut-off of the MELD score for determining severe alcoholic hepatitis is >21 which is associated with 3-month mortality of 20%, whereas patients with MELD score ≤11 have excellent survival.45

**Other Chronic Liver Diseases.** MELD has been shown to be a useful predictor even in patients n whom cirrhosis was not clearly documented. In patients with chronic hepatitis B, MELD and antiviral treatments (lamivudine) were independent predictors of survival.43 In patients admitted to the Intensive Care Unit, the “Royal Free Model” had similar discrimination ability as the MELD and Sequential Organ Failure (SOFA) score47 in predicting mortality. All three scores were superior to APACHE II or CTP scores.48 It must be pointed out that both the Royal Free Model and the SOFA score include variables reflecting multiple system organ failure, and are more likely to “reflect” the dying process rather than being predictors of mortality. Similarly, in patients with acute on chronic liver failure, MELD and hepatic encephalopathy predicted survival,49 especially when MELD was ≥30.17

**MELD in the Management of Patients with ESLD**

**Transjugular Intrahepatic Portosystemic Shunts.**

MELD was originally developed in patients undergoing TIPS and may be used most appropriately to predict probability of survival after the procedure. The indication for TIPS, whether refractory ascites or variceal bleeding, is
not an independent variable determining survival. Emergency TIPS, that is TIPS carried out in a patient actively bleeding in spite of 2 sessions of endoscopic therapy carried out within 24 hours, is a predictor of mortality, but these patients also usually have a higher MELD score.\textsuperscript{1} It is not clear whether TIPS adds additional mortality to that predicted by the MELD score in patients with complications of portal hypertension. That is, it is not clear, for example, whether a patient with a MELD score of 24 undergoing TIPS has a higher mortality than a patient with similar complications and a MELD score of 24 not undergoing TIPS. Nonetheless, TIPS is associated with a higher risk of mortality than seen in patients on the waiting list for liver transplantation with identical MELD scores, at least a small additional risk of procedure-related mortality,\textsuperscript{50} and a risk of morbidity as well as additional expense.

The importance of MELD as an independent predictor of death in patients undergoing TIPS has been confirmed by other studies.\textsuperscript{38,51} The MELD score within the UNOS region at which patients are likely to receive a transplant may also be used in determining whether a TIPS should be carried out before liver transplantation. For instance, if a patient with a MELD score of 28 and refractory ascites is in a UNOS region where organs are available to patients whose MELD score is 28-30, continued conservative measures rather than TIPS might be recommended. In general, a MELD score of >24 is associated with an increased risk of 3-month post TIPS mortality,\textsuperscript{52,53} and consequently, TIPS should be avoided in such patients unless they are candidates for liver transplantation.

The MELD score has been compared to both the Emory score,\textsuperscript{35} as well as the CTP score for prediction of long-term survival in patients undergoing TIPS. The “c” statistic for the MELD score was superior to the Emory score but only slightly superior to or no better than the CTP classification in predicting post procedure mortality.\textsuperscript{38}

**Hepatocellular Carcinoma.** In our experience, hepatic resection for HCC can be carried out safely in patients with cirrhosis and MELD score ≤ 8. Neither minor hepatic resections (≤3 segment resection) nor major resection (≥4 segment) were associated with any mortality 30 days postoperatively if MELD score was ≤ 8. Moreover, patients with HCC smaller than 5 cm in diameter and MELD score ≤ 8, had a 5-year survival of 80%. In patients undergoing ablation therapies for unresectable HCC, patients with MELD score ≤10 and CLIP score ≤2 had the best outcome. In patients with MELD score ≥10 and CLIP ≥2 the outcome was poor; therefore, local therapies for HCC should probably be considered only in patients with MELD score ≤10 and CLIP score ≤2.\textsuperscript{54}

**Selection of Patients for Surgery Other Than Liver Transplantation.** Traditionally, the CTP score has been used to determine risk of postoperative mortality. Additional risk factors for mortality have included serum creatinine concentration, the American Society of Anesthesiologists (ASA) physical status class, and cardiopulmonary comorbidity.\textsuperscript{55} However, these variables have not been put together to create a model to quantitate the risk of postoperative mortality.

We have demonstrated that patients with cirrhosis are at low risk of mortality after hepatic resection for hepatocellular carcinoma (HCC) if their MELD score is 8 or less.\textsuperscript{56} This MELD cutoff of 8 for carrying out hepatic resection for HCC has been confirmed by a study from Italy.\textsuperscript{57} The utility of MELD in determining postoperative mortality has also been confirmed in patients undergoing cardiac surgery,\textsuperscript{58} as well as in abdominal operations including cholecystectomy.\textsuperscript{59-61} However, MELD was an inaccurate predictor of mortality in patients without cirrhosis undergoing liver resection.\textsuperscript{62} In our experience, MELD, the ASA physical status, and age can be used to determine mortality following surgery independent of the procedure performed. We have demonstrated a close relationship between MELD score and mortality, with the relationship persisting both short-term and long-term following surgery, irrespective of the type of surgery being performed. Emergency surgery too was not an independent predictor of mortality independent of the MELD score.\textsuperscript{63} The MELD score, ASA physical status, and age may be used in determining whether elective surgical procedures should be carried out before or following liver transplantation.

**Strengths and Limitations of MELD**

In many ways, MELD is an ideal survival model in comparison to either models/scores used in patients with liver disease (Table 1). Its strengths derive from the robust statistical foundation in its development and the large number and variety of samples in which it was validated. The model is based on only objective variables that are readily obtained. Inclusion of creatinine incorporates a measure of renal function, a well-recognized predictor of survival in patients with liver disease.\textsuperscript{64-66} Whereas the usefulness of CTP has been appreciated by clinicians for many decades, it did not have much statistical basis in its development, nor did it undergo as rigorous validation as MELD. It also includes subjective variables such as ascites and encephalopathy. In our initial validation study, the c-statistic associated with the CTP score in the prediction of 3-month survival was 0.84 (95% CI 0.78-0.90), in
comparison to 0.87 for MELD. Thus, the MELD scale is thought to be at least as good as the CTP score in predicting short-term mortality, while it may overcome many limitations of the CTP score, at least for the purpose of prioritization in donor organ allocation. Several authors pointed out that MELD has not been proven to be superior to the CTP score in patients listed for liver transplantation or in a wider population of patients with cirrhosis.48,67 When the score designation with regard to ascites and encephalopathy is done consistently from one patient to the next by an experienced observer, the CTP score is probably as accurate and reproducible as MELD. However, one of the drawbacks of the CTP system is that it is much more subject to variability and interpretation than MELD. This was one of the factors that made MELD more attractive as a standard for organ allocation in that it minimizes the possibility of “gaming” the system. Another advantage of MELD over the CTP score is that it has a much wider range of possible scores and has a better precision with which to distinguish patients according to their mortality risk. Finally, even the most vocal skeptics of MELD agree that the serum creatinine is an important contribution of MELD in highlighting the importance of renal function in the assessment of mortality risk in patients with ESLD. A model developed by adding serum creatinine to the CTP score has not been proven to be more accurate than the MELD score.68

There are some cautions to be exercised in applying MELD in individual patients. First, one must remember that MELD was created and validated in a cohort of patients who were screened carefully with certain criteria, which included absence of acute, reversible complications, such as bacterial infection or azotemia associated with dehydration. In deriving the TIPS model, we used the prothrombin time and serum creatinine and bilirubin data recorded at the time when reversible factors had been excluded. This approach was taken because we were primarily interested in a measure most accurately reflective of the underlying liver function. Therefore, in patients on the waiting list for liver transplantation, the MELD score should, in principle, be calculated only after acute reversible processes are adequately treated.

Second, the primary role that was asked of MELD was to rank patients according to mortality risk in a relatively homogenous population of registrants on the liver transplant waiting list. Thus, depending on the population to which it is applied, mortality seen in patients with a given MELD score may not necessarily be the same. Similarly, hospitalized patients with cirrhosis who were not candidates for liver transplantation may have a higher mortality than candidates for liver transplantation who are younger and devoid of comorbidity. Thus, it is not possible to provide a universally applicable survival prediction by MELD.

Third, although the objectivity of the variables included in MELD is far superior to previous models, the variables used in the MELD score may be subject to some variability depending on how they are measured. The serum creatinine typically measured by a colorimetric alkaline picric Jaffe method may be less accurate than when the enzymatic method for measuring serum creatinine is used. When the serum bilirubin is above 25 mg/dl, the colorimetric method overestimates the serum creatinine. Accordingly, in patients with serum bilirubin >25 mg/dl the enzymatic method for measuring serum creatinine is recommended. A somewhat related question regarding bilirubin is whether the direct fraction of the serum bilirubin is a more accurate predictor of survival than the total serum bilirubin. In the absence of studies clarifying this issue, the total serum bilirubin remains the preferred index for expressing overall liver function.

The prothrombin time is also subject to variability. The thromboplastins available worldwide have an International Sensitivity Index range from 1-3, the lower numbers indicating a more sensitive thromboplastin. The prothrombin time is more prolonged if a sensitive thromboplastin is used as compared with a less sensitive thromboplastin. The INR for prothrombin time was introduced as a means of decreasing this variability when measuring prothrombin times in patients on warfarin anticoagulation. The accuracy of INR may be decreased as a measure of the coagulation status in patients with liver disease, since there are other abnormalities in the coagulation pathway in those patients. However, its role in predicting

---

**Table 1. Liver Disease: Features of Current Definitions/Scores**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Objective Parameters</th>
<th>Subjective Parameters</th>
<th>Parameters Readily Available</th>
<th>Prospectively Designed</th>
<th>Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>An Ideal Model</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FHF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SBP</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HRS</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Child-Pugh (“ Decompensated Liver Disease”)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MELD</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>(?)✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

According to the table, MELD is considered an ideal model with high accuracy, reproducibility, and a wide range of possible scores, providing a universally applicable survival prediction.
survival in patients with liver disease has been demonstrated repeatedly, especially when used to calculate the MELD score. It is possible that other methods of expressing prothrombin time such as the prothrombin index are more accurate reflectors of liver function, but these methods have yet to be validated as being more accurate than the INR, both as indices of liver function and of coagulation status. In our opinion, measures of the prothrombin time, the INR is superior to others because of its wide availability as well as the track record as a survival indicator in patients with ESLD.

Finally, some debate continues with regard to patients with intractable complications of portal hypertension such as ascites and hepatic encephalopathy. In our initial evaluation of these complications in conjunction with MELD, it was clear that when the c-statistic was used as the criterion to determine the degree of improvement in the model associated with the addition of these complications, they had only minimal benefit to MELD. However, as shown in the case with hyponatremia (see next section), if these complications occur in a small proportion of patients, addition of them would not increase the model c-statistic materially, because it changes the ranking of only a few patients, even when they may have substantial impact in those few patients. To date, however, we are not aware of data to clearly demonstrate that such is the case, except in the case of hyponatremia. If data strongly supportive of these and other complications add to the prognostic evaluation in patients with ESLD, the transplant community may be faced with a difficult decision whether to re-incorporate these potentially subjective elements back to the organ allocation policy.

Further Refinement and Improvement of MELD

Patients with an increasing MELD score have been thought to have an increased risk of mortality, whereas those with a decreasing MELD score have a lower risk of mortality, even if their MELD scores are identical.69 Thus, it has been proposed that the change in MELD score, that is ΔMELD, may add prognostic information to the MELD score.69 Intuitively, a patient whose MELD is increasing rapidly is more likely to have a worse outcome than those with stable MELD. We conducted a study using the time-dependent analysis of the effect of the current MELD score and ΔMELD (defined as the difference between current MELD and the lowest MELD score measured within 30 days prior to current MELD). Although all of these variables were significant in the univariate phase, ΔMELD was no longer significant in the multivariable analysis, especially when acute increases in MELD in the last few days of life were excluded. This analysis highlighted that the current MELD is the most important predictor of survival, regardless how that MELD was reached.12

MELD could potentially be improved with more accurate indices of liver function and perhaps better ways of assessing renal function. Recently, several studies have shown that the addition of serum sodium can improve the predictive accuracy of the MELD score.70-73 Our study, which is based on a multicenter database, shows that there is a linear relationship between serum sodium and mortality after adjusting for MELD. In addition, serum sodium may be particularly relevant in patients with a low MELD score, e.g., MELD < 20.74 As was alluded to earlier, because of the small proportion affected by hyponatremia, the c-statistics of the model did not change substantially. However, with severe hyponatremia, the risk in mortality increased as much as what would be equivalent to an increase of more than 20 points in MELD. Whereas the impact of serum sodium is quite large, it remains uncertain whether the addition of serum sodium to the MELD score can be used to determine allocation of organs for liver transplantation, especially because of the possible poor post liver transplantation outcome of patients with low serum sodium.75

In conclusion, based on its ability to rank patients with cirrhosis according to their short term mortality, MELD has been recognized as a major contribution to the daily practice of hepatology. Successful implementation of MELD-based liver allocation in the United States has been followed by widespread adoption of the system globally, attesting to its validity. In addition to organ allocation, emerging data support MELD as a useful clinical tool in a wide spectrum of disease severity and variety. These achievements notwithstanding, MELD is by no means a perfect system. Users of MELD must be aware of several features and limitations in its application. In the meantime, efforts for further refinement and validation must continue.

References


44. Srikureja W, Kyulo NL, Runyon BA, Hu KQ. MELD score is a better prognostic model than Child-Turcotte-Pugh score or Discriminant Function score in patients with alcoholic hepatitis. J Hepatol 2005;42:700-706.


RESEARCH ARTICLE

The MELD-Plus: A generalizable prediction risk score in cirrhosis

Uri Kartoun1,2,*, Kathleen E. Corey1,3, Tracey G. Simon1,3, Hui Zheng4, Rahul Aggarwal1,2, Kenney Ng5, Stanley Y. Shaw1,2,*

1 Harvard Medical School, Boston, Massachusetts, United States of America, 2 Center for Systems Biology; Center for Assessment Technology & Continuous Health (CATCH), Massachusetts General Hospital, Boston, Massachusetts, United States of America, 3 Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 4 Center for Biostatistics, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 5 IBM Research, Cambridge, Massachusetts, United States of America

* Current address: IBM Research, Cambridge, Massachusetts, United States of America
* stanley_shaw@hms.harvard.edu

Abstract

Background and aims

Accurate assessment of the risk of mortality following a cirrhosis-related admission can enable health-care providers to identify high-risk patients and modify treatment plans to decrease the risk of mortality.

Methods

We developed a post-discharge mortality prediction model for patients with a cirrhosis-related admission using a population of 314,292 patients who received care either at Massachusetts General Hospital (MGH) or Brigham and Women’s Hospital (BWH) between 1992 and 2010. We extracted 68 variables from the electronic medical records (EMRs), including demographics, laboratory values, diagnosis codes, and medications. We then used a regularized logistic regression to select the most informative variables and created a risk score that comprises the selected variables. To evaluate the potential for generalizability of our score, we applied it on all cirrhosis-related admissions between 2010 and 2015 at an independent EMR data source of more than 18 million patients, pooled from different health-care systems with EMRs. We calculated the areas under the receiver operating characteristic curves (AUROCs) to assess prediction performance.

Results

We identified 4,781 cirrhosis-related admissions at MGH/BWH hospitals, of which 778 resulted in death within 90 days of discharge. Nine variables were the most effective predictors for 90-day mortality, and these included all MELD-Na’s components, as well as albumin, total cholesterol, white blood cell count, age, and length of stay. Applying our nine-variable risk score (denoted as “MELD-Plus”) resulted in an improvement over MELD and MELD-Na scores in several prediction models. On the MGH/BWH 90-day model, MELD-Plus
improved the performance of MELD-Na by 11.4% (0.78 [95% CI, 0.75–0.81] versus 0.70 [95% CI, 0.66–0.73]). In the MGH/BWH approximate 1-year model, MELD-Plus improved the performance of MELD-Na by 8.3% (0.78 [95% CI, 0.76–0.79] versus 0.72 [95% CI, 0.71–0.73]). Performance improvement was similar when the novel MELD-Plus risk score was applied to an independent database; when considering 24,042 cirrhosis-related admissions, MELD-Plus improved the performance of MELD-Na by 16.9% (0.69 [95% CI, 0.69–0.70] versus 0.59 [95% CI, 0.58–0.60]).

Conclusions

We developed a new risk score, MELD-Plus that accurately stratifies the short-term mortality of patients with established cirrhosis, following a hospital admission. Our findings demonstrate that using a small set of easily accessible structured variables can help identify novel predictors of outcomes in cirrhosis patients and improve the performance of widely used traditional risk scores.

Introduction

Cirrhosis-related complications account for 1.1% and 1.8% of all deaths in the United States and Europe, respectively [1, 2]. In addition to increased mortality, individuals with cirrhosis suffer from significantly worse health issues and greater disability compared to those without cirrhosis [3].

Although risk-stratification tools for the prediction of cirrhosis-related mortality are available [4–9], these models are based on small populations and use a limited number of preselected traditional predictors. Improved mortality prediction scores may highlight the clinical variables that contribute to mortality risk, includingmodifiable factors, and guide the allocation of resources to improve cirrhosis care for high-risk patients.

The recent availability of large cohorts of data from electronic medical records (EMRs) allows for the development of improved mortality prediction scores through inclusion of a broader set of clinically applicable, unbiased variables. Not only that, but developing such prediction models allows clinicians to identify the clinical variables that contribute to mortality risk, including modifiable factors. Improving current standard models like the model for end-stage liver disease (MELD) and MELD-Na can guide clinicians in better targeting treatment to improve cirrhosis care and outcomes for high-risk patients.

Cohorts assembled from EMRs represent a powerful resource to study disease complications at the population level. Recent studies have demonstrated the usefulness of EMR analysis to discover or confirm outcome correlations, sub-categories of disease, and adverse drug events [10–14]. The MELD score is based on three commonly used laboratory tests available in the EMRs, and it is the most widely used tool to predict outcomes in patients with cirrhosis [15, 16]. An extended version of MELD, one that incorporates serum sodium levels, the MELD-Na score, has been recently adopted by The Liver and Intestine Transplantation Committee for liver transplant allocation [17]. Although the two scores are simple to calculate and apply in a practical sense, the improved accessibility of a wide variety of variables from EMRs raises the possibility that prediction models could benefit from the inclusion of a broader, unbiased set of clinical variables. Identifying a combination of the most informative variables may improve the prognostic utility beyond that of current risk scores.
The aim of the present study was to develop a risk score to predict mortality following a cirrhosis-related admission. We demonstrated that a score composed of a small set of easily accessible clinical variables improves the prediction performance of both the MELD and MELD-Na scores. We further demonstrated the generalizability of our model through independent validation in a large EMR-based data source.

Methods

Study population

We analyzed a previously defined cohort of 314,292 patients at increased risk for metabolic disease who were admitted to Massachusetts General Hospital (MGH) or Brigham and Women’s Hospital (BWH) between 1992 and 2010 [13]. We identified an admission as cirrhosis-related when the keyword “cirrhosis” was present in the discharge summary of the admission and we observed at least one ICD-9 code (571.2, 571.5, or 571.6 as in [18]) within the 30 days preceding the discharge date, including during the admission. This identification method was validated by a physician (Dr. Kathleen Corey) chart review.

We excluded elective admissions if they included at least one diagnosis or procedure code for liver biopsy, radiofrequency ablation, transarterial chemoembolization, hepatic resection, or liver transplant. We included only patients 18 years of age or older at the time of the admission, and we tracked the records of all patients for 90 days after their discharge. We determined mortality through linkage to the social security master death index.

Prediction modeling

To predict mortality within 90 days, we developed a model that included a large set of structured variables extracted from the EMRs. In addition to variables available during the period of admission, we considered variables available for the period of 12 months preceding the discharge date (see Table 1).

The variables included demographics (e.g., gender, ethnicity, marital status), laboratory measurements (e.g., albumin, sodium), and medications (e.g., anticoagulants, lipid lowering agents). For laboratory variables, we used the most recent values found during admission (when no value was found during admission, we considered the preceding 12 months). Typically, common laboratory measurements were available during the admission (as seen in Table 1). We determined comorbidities from the number of diagnosis codes within the 12 months prior to the discharge date, and we determined medication count by recording the number of prescriptions within the 12 months preceding the discharge date.

Additional variables included body mass index, NAFLD fibrosis score (Eq 1), and the MELD score (Eq 2). Missing values were imputed with the mean of the available data for each variable. We randomly selected two thirds of the admissions to serve as a derivation set, whereas the remaining one third served as a validation set. A complete list of the variables we used is available in S1 Table, and all diagnoses and procedure definitions used in this study are available in S2 Table.

NAFLD Fibrosis Score = \[-1.675 + 0.037 \times \text{Age} + 0.094 \times \text{BMI} +1.13 \times \text{IFG}/\text{Diabetes (yes = 1, no = 0)} +0.99 \times \text{AST}/\text{ALT ratio} - 0.013 \times \text{Platelet} - 0.66 \times \text{Albumin}\]  

MELD Score = \[+6.43 + 9.57 \times \ln(\text{Creatinine}) +3.78 \times \ln(\text{Total Bilirubin}) + 11.2 \times \ln(\text{INR})\]  

(1)

(2)
Table 1. Baseline characteristics. All values extracted during the 12 months preceding discharge date. For laboratory variables, values are the most recent. Comorbidity calculations count the number of diagnosis codes. Prevalence calculations consider admissions with at least one measurement for laboratories and at least one diagnosis code for comorbidities.

<table>
<thead>
<tr>
<th>Variable and category</th>
<th>Cirrhosis-related admissions (n = 4,781)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years); Mean (SD)</td>
<td>60.0 (13.7)</td>
</tr>
<tr>
<td>Gender (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64.4</td>
</tr>
<tr>
<td>Female</td>
<td>35.6</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>77.4</td>
</tr>
<tr>
<td>African American</td>
<td>6.8</td>
</tr>
<tr>
<td>Other</td>
<td>2.3</td>
</tr>
<tr>
<td>Unknown</td>
<td>13.5</td>
</tr>
<tr>
<td>Insurance Type (could be ≥ 1 types per patient) (%)</td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td>5.6</td>
</tr>
<tr>
<td>Medicare</td>
<td>60.0</td>
</tr>
<tr>
<td>Other</td>
<td>98.6</td>
</tr>
<tr>
<td>BMI (kg/m²); Mean (SD)</td>
<td>28.7 (8.2)</td>
</tr>
<tr>
<td>Laboratory values; Mean (SD) / Prevalence (%)</td>
<td></td>
</tr>
<tr>
<td>Sodium (mmol/l)</td>
<td>136.5 (5.8) / 99.7</td>
</tr>
<tr>
<td>eGFR (ml/min/1.73m²)</td>
<td>59.6 (33.7) / 31.9</td>
</tr>
<tr>
<td>WBC (th/cumm)</td>
<td>6.7 (3.8) / 99.8</td>
</tr>
<tr>
<td>Platelets (th/cumm)</td>
<td>141.3 (100.0) / 99.8</td>
</tr>
<tr>
<td>Prothrombin time (INR)</td>
<td>1.5 (0.5) / 91.3</td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
<td>2.9 (0.7) / 98.7</td>
</tr>
<tr>
<td>Total Bilirubin (mg/dl)</td>
<td>2.5 (4.3) / 98.7</td>
</tr>
<tr>
<td>Transaminase SGOT (u/l)</td>
<td>60.2 (76.6) / 98.8</td>
</tr>
<tr>
<td>Transaminase SGPT (u/l)</td>
<td>36.1 (38.2) / 96.7</td>
</tr>
<tr>
<td>GGT (u/l)</td>
<td>249.4 (346.6) / 6.8</td>
</tr>
<tr>
<td>MELD score</td>
<td>14.21 (6.1) / 84.8</td>
</tr>
<tr>
<td>NAFLD Fibrosis score</td>
<td>1.70 (2.1) / 13.7</td>
</tr>
<tr>
<td>Comorbidities; Mean (SD) / Prevalence (%)</td>
<td></td>
</tr>
<tr>
<td>Variceal hemorrhage / Gastrointestinal bleed</td>
<td>0.8 (2.1) / 24.2</td>
</tr>
<tr>
<td>Spontaneous bacterial peritonitis</td>
<td>0.1 (0.6) / 3.4</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>0.7 (5.5) / 4.3</td>
</tr>
<tr>
<td>Hepatorenal syndrome</td>
<td>0.1 (0.5) / 3.2</td>
</tr>
<tr>
<td>Hepatic encephalopathy</td>
<td>1.4 (3.6) / 28.6</td>
</tr>
<tr>
<td>Ascites</td>
<td>2.4 (5.7) / 37.9</td>
</tr>
<tr>
<td>Renal failure</td>
<td>1.8 (8.4) / 13.1</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>0.5 (2.4) / 10.8</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>2.7 (7.9) / 36.7</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3.1 (7.2) / 58.9</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>0.5 (2.1) / 13.4</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>0.3 (1.2) / 12.4</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>0.5 (2.2) / 13.0</td>
</tr>
<tr>
<td>Diabetes</td>
<td>5.9 (10.3) / 57.0</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0186301.t001
To select the most informative variables, we applied feature selection on the derivation set. We used logistic regression with the adaptive least absolute shrinkage and selection operator (LASSO) algorithm [19] because it is considered an efficient algorithm for parsimoniously ranking variables in clinical predictive modeling [20, 21]. We considered all variables that were statistically significantly different from a univariate analysis \((P < 0.05)\) as in [22]. The generalized linear model (GLM) equations used to calculate prediction risk at MGH/BWH and at the independent EMR source are presented in Eqs 3 and 4.

\[
L =
+11.794383 \\
+2.076192 \times \log_{10}(1 + \text{Total Bilirubin}) \\
+2.494291 \times \log_{10}(1 + \text{Creatinine}) \\
-6.049540 \times \log_{10}(1 + \text{Albumin}) \\
+2.525904 \times \log_{10}(1 + \text{INR}) \\
+1.911856 \times \log_{10}(1 + \text{WBC}) \\
+0.015411 \times \text{Length of stay} \\
\text{(number of nights the patient spent in the hospital during the cirrhosis–related admission)} \\
+0.041047 \times \text{Age (years)} \\
-6.625270 \times \log_{10}(1 + \text{Sodium}) \\
-1.445666 \times \log_{10}(1 + \text{Total Cholesterol})
\]

\[
\text{MELD–Plus} = p(90 \text{ day mortality}) = \frac{\exp(L)}{1 + \exp(L)}
\]

To calculate 95% confidence intervals, we applied the bootstrap procedure with 1,000 replicates. We calculated the area under the receiver operating characteristic curves (AUROC) to measure the model’s accuracy in the validation set. Additionally, we evaluated for overfitting by comparing the AUROC in the validation set to an average AUROC value for 100 permutations of randomly selected derivation and validation sets (each including two thirds and one third of the derivation set’s cirrhosis-related admissions, respectively). We compared categorical variables using a chi-squared test, and we compared the differences in the means of continuous variables using a t-test or Wilcoxon rank sum test, as appropriate. We further compared the differences in standard deviations by using an F-test. All statistical tests were two-sided, with Bonferroni corrections for the 68 comparisons, and the adjusted \(P\) value was \(7.4 \times 10^{-4}\) for each comparison. We performed all programming using the R statistical language [23].

Independent validation

We were granted access to a data source of 18,345,793 individuals, pooled from multiple different health-care systems with EMRs (“The IBM Explorys Network”) [24]. The data were standardized and normalized using common ontologies, searchable through a HIPAA-enabled, de-identified cloud-computing platform. Patients were seen in multiple health-care systems between 2010 and 2015, with a combination of data from clinical EMRs, outgoing health-care system bills, and adjudicated payor claims.

We first identified all cirrhosis-related admissions in this database, and then we extracted values for the selected variables of our MGH/BWH 90-day model. We further deployed the
GLM equations on the independent source (Eqs 3 and 4). Missing values were imputed based on
the mean values of the MGH/BWH 4,781 cirrhosis-related admissions: total bilirubin (2.486604493), creatinine (1.375274633), albumin (2.888940902), INR (1.499619615), WBC (6.673836966), sodium (136.5401552), and total cholesterol (133.6246914).

Because death dates were not available for patients in the IBM Explorys Network, we used
the year of death to determine the outcome. We were able then only to use a 1-year estimated
death for this population (e.g., for a patient discharged on October 13, 2010, we could only
determine if the patient died either in 2010 or 2011, or survived after that). To compare the
performance of the IBM Explorys approximate one-year prediction model, we used the origi-
nal 314,292-patient population at MGH/BWH and applied the same approximate one-year
mortality outcome identification method. We calculated AUROCs for MELD, MELD-Na (Eq
5), and for our risk score (Eqs 3 and 4).

\[
\text{MELD-Na} = \text{MELD} + 1.59 \times (135 - \text{Na})
\]

(5)

The institutional review board of Partners HealthCare and IBM approved this study and all
its methods, including the EMR cohort assembly, data extraction, and analyses.

Results

Univariate analysis

We identified a total of 4,781 admissions as cirrhosis-related, of which 778 resulted in death
within 90 days of the discharge date (16.3%). In a sample of 50 randomly selected patients,
64% were admitted primarily for cirrhosis, for instance, due to the presence of ascites or sponta-
naneous bacterial peritonitis, and the rest had the comorbidity of cirrhosis but were admitted
primarily for different reasons such as heart failure or chronic obstructive pulmonary disease
(COPD). Individuals who died within the 90-day period after discharge were older in compari-
son with those who survived (64.1 years versus 59.2 years, \( P = 4.22 \times 10^{-17} \)); however, the two
populations did not differ by gender (65.0% male versus 64.0% female, \( P = 1.0 \)) or ethnicity
(77.0% Caucasian for both, \( P = 1.0 \)).

We calculated event ratios by dividing the values ascertained for the two populations (e.g.,
the mean MELD scores were 18.5 and 13.3 for admissions that resulted in death and survival,
respectively, yielding a ratio of 1.4, \( P = 4.45 \times 10^{-65} \)). Individuals who died within the 90-day
period after discharge had higher ratios of liver-related comorbidities than those who survived,
and these comorbidities included hepatorenal syndrome (ratio = 5.1, \( P = 4.60 \times 10^{-21} \)), hepatocel-
lular carcinoma (ratio = 5.0, \( P = 1.40 \times 10^{-16} \)), and ascites (ratio = 2.1, \( P = 3.55 \times 10^{-24} \)). Laboratory
measurements also significantly differentiated the two populations. For instance, albumin was
lower in those who died within the 90-day period (2.60 g/dl versus 2.95 g/dl, \( P = 5.64 \times 10^{-35} \)),
and the total bilirubin (4.87 mg/dl versus 2.02 mg/dl, \( P = 1.82 \times 10^{-41} \)), INR (1.70 versus 1.46,
\( P = 5.63 \times 10^{-30} \)), and creatinine (1.80 mg/dl versus 1.29 mg/dl, \( P = 1.24 \times 10^{-36} \)) were higher in
those who died within the 90-day period.

No difference was found in the prevalence of COPD, cerebrovascular disease, diabetes, cor-
onary artery disease, peripheral vascular disease, pneumonia, or sleep apnea between the popu-
lations. The complete list of variables, indicating the differences between the surviving and the
deceased populations, is presented in S3 Table.

Logistic regression model

The AUROCs of 0.78 were identical for all three models composed of multiple variables (Fig
1). With generalizability in mind and the potential ease of extraction of commonly available
laboratory values and other trivial variables (e.g., age, length of stay), we decided to follow the model that comprised the 9 readily available clinical variables. To evaluate the contribution of the MELD score to the 90-day mortality prediction, we evaluated the performance of MELD and MELD-Na scores alone. Considering the 4,781 admissions, using the MELD score alone to predict the 90-day mortality resulted in an AUROC value of 0.69. An additional model using the MELD-Na score alone yielded an AUROC value of 0.70.

Each of the MELD-Na components were associated with an increased mortality, including INR (OR, 1.58; 95% CI, 1.30–1.96), creatinine (OR, 1.25; 95% CI, 1.16–1.34), total bilirubin (OR, 1.11; 95% CI, 1.08–1.14), and sodium (OR, 0.97; 95% CI, 0.95–0.99). Other laboratory measurements associated with mortality included WBC (OR, 1.10; 95% CI, 1.07–1.13), total cholesterol (OR, 0.996; 95% CI, 0.993–0.999), and albumin (OR, 0.45; 95% CI, 0.37–0.52). Additional predictors included age at time of the admission (OR, 1.04; 95% CI, 1.03–1.05) and length of stay (OR, 1.02; 95% CI, 1.005–1.03).

Because total cholesterol and hospital length of stay are typically not uniform factors across different hospitals and may vary in different countries, we evaluated an additional model that included only 7 of the 9 variables. This yielded an AUROC of 0.77 and resulted in the following associations with increased mortality: INR (OR, 1.66; 95% CI, 1.38–2.05), creatinine (OR, 1.25; 95% CI, 1.17–1.35), total bilirubin (OR, 1.11; 95% CI, 1.08–1.14), sodium (OR, 0.97; 95% CI, 0.95–0.98), WBC (OR, 1.10; 95% CI, 1.07–1.13), albumin (OR, 0.43; 95% CI, 0.36–0.51), and age (OR, 1.04; 95% CI, 1.03–1.05). We present the GLM equations used to calculate prediction performance at MGH/BWH in Eqs 6.
and 7.

\[ L = \]
\[ +8.53499496 \]
\[ +2.06503238 \times \log_{10}(1 + \text{Total Bilirubin}) \]
\[ +2.59679650 \times \log_{10}(1 + \text{Creatinine}) \]
\[ -6.34990436 \times \log_{10}(1 + \text{Albumin}) \]
\[ +2.99724802 \times \log_{10}(1 + \text{INR}) \]
\[ +1.92811726 \times \log_{10}(1 + \text{WBC}) \]
\[ +0.04070442 \times \text{Age (years)} \]
\[ -6.47834101 \times \log_{10}(1 + \text{Sodium}) \]

\[ \text{MELD-Plus (excluding length of stay and total cholesterol)} = \]
\[ p(90 \text{ day mortality}) = \frac{\exp(L)}{1 + \exp(L)} \]

**Prediction of 90-day mortality after a cirrhosis-related admission**

Using our 9-variable risk score, we divided our population into quintiles and compared the average predicted 90-day mortality with the observed mortality within each quintile. The predicted 90-day mortality derived from a logistic regression model for each admission and indicated the probability that a patient who survived the admission would die within 90 days post discharge. As shown in Fig 2A–2C, the predicted 90-day mortality was strongly correlated with the observed mortality rate throughout the range of risk in both derivation and validation sets (Kendall’s τ = 1.0; \( P = 0.027 \); Pearson correlation \( r = 0.995 \) for the correlation between the average calculated and observed mortality). We provide the logistic regression equations used to calculate the predicted 90-day mortality probabilities in Eqs 3 and 4. The complete list of variables that indicate the differences between the highest-risk quantile and the lowest-risk quantile populations are presented in S4 Table.

**Generalization evaluation**

Applying our 9-variable risk score (the MELD-Plus score) demonstrated an improvement over MELD and MELD-Na scores in all prediction models, as shown in Fig 3. On the MGH/BWH 90-day model, MELD-Plus improved the performance of MELD-Na by 11.4% (0.78 [95% CI, 0.75–0.81] versus 0.70 [95% CI, 0.66–0.73]). On the MGH/BWH approximate 1-year model, MELD-Plus improved the performance of MELD-Na by 8.3% (0.78 [95% CI, 0.76–0.79] versus 0.72 [95% CI, 0.71–0.73]). On the IBM Explorys Network model used for external validation, MELD-Plus improved the performance of MELD-Na by 16.9% (0.69 [95% CI, 0.69–0.70] versus 0.59 [95% CI, 0.58–0.60]).

It is notable that the performance of MELD-Plus on the IBM Explorys data was lower in comparison with both MGH/BWH models (0.69 versus 0.78). Consistent with MELD-Plus, the performance of MELD and MELD-Na were also much lower on the IBM Explorys data in comparison with MGH/BWH. A potential reason for this is that the IBM Explorys Network population was relatively healthier. Patients in the IBM Explorys network had lower severity of liver disease in comparison with the corresponding MGH/BWH 1-year prediction model.
There may be other differences in the data or populations in the independent systems; the Partners HealthCare Research Patient Data Registry collected the MGH/BWH data, whereas dozens of distinct data aggregation mechanisms collected the data for the IBM Explorys Network. Furthermore, the variability in the levels of prediction performance might be influenced by the variability in the data; prediction performance might be higher when there is more variability in the data source (i.e., the population comprising patients with a broad spectrum of
Fig 3. Prediction performance across different cirrhosis populations. (A) MGH/BWH 90-day mortality (4,781 cirrhosis-related admissions). (B) The IBM Explorys Network approximate 1-year mortality (24,042 cirrhosis-related admissions). (C) MGH/BWH approximate 1-year mortality (4,680 cirrhosis-related admissions).

https://doi.org/10.1371/journal.pone.0186301.g003
levels of cirrhosis severity). In the other direction, when the data is more uniform (e.g., most patients have just been diagnosed with cirrhosis for the first time, and only a minority suffers from an advanced cirrhosis), then prediction accuracy is lower. This hypothesis was confirmed because the IBM Explorys network had a statistically significant lower standard deviation of severity of liver disease in comparison with the MGH/BWH 1-year population (STD MELD: 1.8 versus 8.2; \( P < 0.0001 \), STD MELD-Na: 3.6 versus 8.2; \( P < 0.0001 \)).

**Discussion**

In this study, we used accessible EMR variables to develop a highly accurate, predictive model of 90-day post-discharge mortality in individuals with cirrhosis. We identified 9 variables that accurately predicted 90-day mortality with an AUROC of 0.78. Our risk score improved the performance of MELD and MELD-Na scores in multiple, independent patient populations, and this also held true in a large external validation patient cohort. Furthermore, our model’s calculated 90-day mortality risk was highly correlated with the observed mortality rate across all five risk quintiles. In particular, the model’s performance on the highest-risk quintile (the calculated and observed 90-day mortality was 31.6% and 31.2%, respectively) suggests that high-risk patients can be accurately identified. An additional model that included only 7 of the 9 variables and excluded length of stay and total cholesterol yielded an AUROC of 0.77 [95% CI, 0.74–0.80]. Although the 7-variable model demonstrated improved identification ability compared to MELD or MELD-Na, the improved prediction performance achieved by including total cholesterol in MELD-Plus suggests that it may be beneficial for cholesterol labs to be routinely collected in cirrhosis admission order sets.

The MELD score has been used extensively to predict patient outcomes, mortality, and readmission rates in individuals with cirrhosis [25, 26, 4]. Furthermore, although MELD-Na [17] was superior to MELD, the MELD-Plus score yielded improved levels of discrimination consistently in all prediction models, with AUROCs that significantly outperformed the traditional scores. These findings suggest that new types of cirrhosis-related risk indexes utilizing novel risk indicators may improve prognostication in this high-risk population.

MELD-Plus includes all MELD-Na’s components, as well as additional variables (albumin, total cholesterol, WBC, age, and length of stay). It is logical that a prediction model that has all the MELD-Na model variables and additional ones would perform better, as was observed by MELD-Plus. Not only that, but many of the variables have physiological plausibility for inclusion in a prediction model. Decreased albumin correlated with worse outcomes in our model, which may be the result of decreased albumin marking decreased liver function in cirrhosis patients [25, 5]. Increasing age and length of hospital stay helped predict worse outcomes as well as could be expected. Along with that, higher WBC was correlated with a worse prognosis, potentially indicating poorer patient status (e.g., infection) at time of score calculation. Although patients may have multiple WBC measurements during admission, our model is both internally and externally valid because it uses the most recent WBC lab value. We chose the most recent WBC during model development because the last available set of labs is more reflective of the current health of patients than older measurements. Surprisingly, increased total cholesterol predicted a more favorable prognosis. Although unintuitive at first, this aligns with previous reports that claim cholesterol levels become less of a risk factor or even an inverse risk factor for mortality because serious diseases may lower cholesterol soon before death occurs [27].

Although our study describes analyses of retrospective medical databases, the proposed score could be used to identify patients that are at a high-risk of mortality in real time and thus may inform risk-stratification and therapeutic decision-making. In a desirable scenario, our score could be calculated automatically as an integrated component of an EMR system; the
clinician would see a risk score (probability) or a risk quantile (highest, lowest, or in between) associated with the discharged patient, and this could be used to guide outpatient monitoring strategies. With further validation, the MELD-Plus score could also be used longitudinally in outpatients to monitor disease progression and/or responses to therapy.

Our study has limitations. First, it is a retrospective analysis limited to two academic, tertiary-care hospitals. Even though we validated our model on a large external patient cohort, subsequent studies must further assess the validity of our model in the external population and consider different age ranges, coding systems, and data-collection methods. Second, the cirrhosis populations may vary at different centers—for example, alcohol use might significantly vary between patients residing in the Boston area versus patients residing in other states [28]. Furthermore, although MGH and BWH are urban care facilities, the high prevalence of rural populations at the IBM Explorys Network might affect prediction performance. Third, although mortality was recorded, either through linkage to the social security master death index as in the MGH/BWH models or through using EMR or billing/claims in the IBM Explorys model, such death indications may under-represent the true mortality rates. To minimize this potential under-representation, we considered only patients who survived the study follow-up. All patients had EMR data entries (such as laboratory measurements) after the study follow-up, indicating survival, or had a recorded indication of death during the study follow-up, with no EMR data entries found afterward.

Another limitation of MELD-Plus is that it did not specifically consider which procedures patients underwent during the cirrhosis-related admissions. Furthermore, all the patients considered in our models survived the admission, but neither MGH/BWH’s nor IBM’s databases contained information on post-discharge cause of death. To further assess MELD-Plus’s applicability in clinical practice, future analyses should consider subgroups of patients to determine linkages between invasive inpatient procedures and causes of mortality. Regardless of this limitation, however, our MELD-Plus displayed validity in predicting overall mortality, which is clinically applicable, because it provides clinicians with information on populations of patients who need more intense or closer care.

Although we excluded elective admissions for liver biopsy, radiofrequency ablation, transarterial chemoembolization, hepatic resection, or liver transplant, these criteria might exclude patients with early and intermediate hepatocellular carcinoma (HCC), but not patients with advanced HCC who underwent medical treatments only. Liver cancer can lead to early mortality, even in patients with mild liver cirrhosis, and, as such, our exclusion criteria may reduce the applicability of our risk score when applying it to patients with more advanced HCC. Furthermore, because we excluded admissions associated with a liver transplant, mortality risk may decrease after a cirrhosis-related admission if patients successfully underwent a transplant in a preceding admission.

Another limitation of our study is algorithmic. The adaptive LASSO method identified 9 predictors and left out variables that may also be correlated with predicting death. Feature selection algorithms are known to be blind to the clinical importance of variables, and when highly correlated predictors are identified, the algorithm randomly selects one. On the one hand, important variables such as ascites, hepatocellular carcinoma, and diuretic medications were not selected as predictors. On the other hand, the feature selection algorithm assures that a minimal set of covariates produce a high level of prediction accuracy. Furthermore, we conducted our model performance evaluation on a held-out data set not used for training. Although a prediction model’s error usually decreases when more variables are included, this is not always the case. This is true when performance is evaluated on the training set (due to overfitting) but not the case when performance is evaluated on a held-out test dataset, as was used across all our models.
In conclusion, we describe an unbiased and well-validated score to estimate 90-day mortality after a cirrhosis-related admission. This score, comprising a small set of easily available clinical variables extracted from EMRs, improved the MELD and MELD-Na scores in predicting 90-day mortality and approximate 1-year mortality. In addition, we identified high-risk patients with great accuracy. MELD-Plus’s strong performance demonstrates potential for it to replace current standard models, allowing for greater accuracy in the identification of high-risk cirrhosis patients.

Supporting information

S1 Table. Summary of variables.
(Document)

S2 Table. Billing codes used to define conditions.
(Document)

S3 Table. Comparison of variables in patients who died vs. survived 90-days after discharge.
(Document)

S4 Table. Comparison between the highest-risk (1st) and the lowest-risk (5th) quintiles.
(Document)

Acknowledgments

We acknowledge Chin Hur MD MPH (Director, GI Health Outcomes Research, GI Unit, Massachusetts General Hospital / Harvard Medical School) for his critical review of this manuscript.

Author Contributions

Conceptualization: Uri Kartoun, Kathleen E. Corey, Stanley Y. Shaw.

Data curation: Uri Kartoun, Rahul Aggarwal.

Formal analysis: Uri Kartoun, Rahul Aggarwal, Kenney Ng, Stanley Y. Shaw.

Funding acquisition: Kathleen E. Corey, Stanley Y. Shaw.

Investigation: Stanley Y. Shaw.

Methodology: Uri Kartoun, Kathleen E. Corey, Kenney Ng, Stanley Y. Shaw.

Resources: Stanley Y. Shaw.

Software: Uri Kartoun, Kenney Ng.

Supervision: Kathleen E. Corey, Stanley Y. Shaw.

Validation: Uri Kartoun, Rahul Aggarwal, Kenney Ng.

Writing – original draft: Uri Kartoun.

Writing – review & editing: Kathleen E. Corey, Tracey G. Simon, Hui Zheng, Rahul Aggarwal, Kenney Ng, Stanley Y. Shaw.

References


