MELD remains the best predictor of mortality in outpatients with cirrhosis, severe ascites, and MELD score ≤ 18

Running title: Mortality in cirrhosis with severe ascites

Peter Jepsen[1,2], Hugh Watson[1,3], Stewart Macdonald[4], Hendrik Vilstrup[1], Rajiv Jalan[4]

1. Department of Hepatology and Gastroenterology, Aarhus University Hospital, Aarhus, Denmark.
2. Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark.
3. Infectious Diseases Unit, Sanofi-Aventis R&D, Marcy l'Etoile, France.
4. Liver Failure Group, Institute for Liver and Digestive Health, University College London, London, UK.

Acknowledgement: The authors thank Sanofi-Aventis R&D for providing access to the complete trial database from the satavaptan phase III trials in cirrhotic ascites and for permission to perform the analyses included in this publication.

Word count: 2960.

Tables and Figures: 4 Tables, 1 Figure + 2 Supplemental Figures.

Corresponding author: Peter Jepsen, Department of Hepatology and Gastroenterology, Palle Juul-Jensens Boulevard, Aarhus University Hospital, Aarhus, Denmark; telephone: +45 78450000; fax: +45 78452860; email: pj@clin.au.dk

Authors' contributions: PJ, HW, and RJ conceived and designed the study. PJ conducted the analyses and drafted the manuscript. All authors contributed to the interpretation of the analysis results and revision of the manuscript. All authors approved the final version.

Funding: None.

Conflicts of interest: Rajiv Jalan has research collaborations with Yaqrit. He is the inventor of MNK6105 and MNK6106 which have been patented by UCL and licensed to Mallinckrodt Pharma. He is also the founder of Yaqrit Ltd. (a spin out company from University College London), Thoeris GmBh, Cyberliver Ltd. and Hepyx Ltd.

Summary

Background: The Model for Endstage Liver Disease (MELD) score may put patients with severe ascites at a disadvantage because they often have a poor quality of life and high mortality despite a favorable MELD score.

Aims: To develop a model that is better than the MELD score at predicting 1-year mortality among patients with cirrhosis, severe ascites, and MELD ≤18.

Methods: We used data from a randomized trial (SPARe-1) of patients with cirrhosis and severe ascites to develop a model to predict 1-year mortality. We used stepwise backward elimination and Cox regression to identify the strongest predictors. Performance was assessed with the C index and the Brier score. We examined performance in an external cohort of trial participants with cirrhosis and severe ascites (SPARe-2 participants).

Results: We included 308 patients with a 1-year mortality of 20.4%. The final prediction model (Severe Ascites Mortality score, "SAM score") included four variables: serum bilirubin, serum sodium, history of SBP (yes or no), and history of diabetes (yes or no). No indicators of quality of life were included. After correction for optimism bias, the SAM and MELD scores had nearly identical predictive ability. The external validation cohort included 149 patients whose 1-year mortality was 22.4%. The MELD score performed marginally better in this cohort, partly because the effects of SBP and diabetes on mortality were much smaller in this cohort.

Conclusions: We did not succeed in developing a prediction model that was superior to the MELD score among patients with cirrhosis and severe ascites.

Keywords: Cirrhosis, ascites, prediction model, mortality, clinical epidemiology.

Introduction

In the clinical management of patients with cirrhosis, many decisions rest on predictions of mortality risk. The Model for Endstage Liver Disease score (MELD score) is very widely used to predict mortality, but studies have raised concern that the MELD score puts patients with severe ascites at a disadvantage because they often have a very poor quality of life and high mortality despite a relatively low MELD score.[1-3] This concern has persisted even after serum sodium has been incorporated into the MELD score by the United Network for Organ Sharing.[4, 5] Possibly, there is something special about patients with severe ascites so that the MELD score loses its association with mortality. If so, we might identify those factors and incorporate them in a prediction model that performs better than the MELD score. To that end, we have shown that the physical component score of the SF-36 quality of life scoring system is associated with mortality in this group of cirrhosis patients, and it measures something other than traditional measures of cirrhosis severity.[6] We speculate that this factor could improve upon the MELD score in this patient group with severe ascites and a relatively low MELD score. It is clinically important that we are able to predict short-term mortality in these patients, so we aimed to develop a model that is better than the MELD score at predicting 1-year mortality among patients with cirrhosis, MELD score ≤18, and severe ascites.

Methods

We included patients who participated in the multinational SPARe-1 trial (ClinicalTrials.gov Identifier: NCT00359437),[7] a randomized comparison between satavaptan and placebo on top of diuretic treatment for the treatment of recurrent ascites in patients with cirrhosis. Included patients must receive diuretics and have undergone therapeutic paracentesis in the previous 24 hours with

the removal of ≥4 liters of ascites and on at least one other occasion in the previous 3 months. The 496 included patients were followed for one year.

The trial excluded patients with ascites of cardiac origin or due to peritoneal infection (e.g. tuberculosis) or peritoneal carcinoma, as well as patients with variceal bleeding or spontaneous bacterial peritonitis (SBP) in the 10 days before randomization and patients with a functional transjugular intrahepatic portosystemic shunt (TIPS). Other reasons for exclusion were: serum creatinine >150 μmol/L, serum potassium <3.5 or >5.0 mmol/L, serum sodium >142 mmol/L, serum bilirubin >150 μmol/L, serum magnesium <0.65 mmol/l, INR >3.0, platelets <30,000/mm3, neutrophils <1,000/mm3, systolic arterial pressure <80 mmHg or symptomatic orthostatic hypotension, hepatocellular carcinoma (HCC) exceeding the Milan criteria, use of a potent modifier of the cytochrome P450 3A pathway, hepatitis B virus infection, previous liver transplantation, hepatic encephalopathy West Haven grade 2–4, or gastrointestinal bleeding in the 10 days prior to randomization.

In the trial, satavaptan was not efficacious in preventing the recurrence of ascites, and the trial was stopped early due to a higher mortality in the satavaptan arm.[7] The incidence of other complications was the same in the two treatment arms, and survival after the planned one-year treatment was assessed in all included patients.

We used the trial data for this study, but excluded additional patients: Those with incomplete data on the SF-36 questionnaire, those with missing data on INR, bilirubin, creatinine, sodium, or albumin, and those who were hospitalized or had an infection at the time of randomization. Moreover, we excluded patients who had HCC because their status on the liver transplant waiting list is not based on MELD score alone.

Finally, we calculated the MELD score for all patients, using the current United Network of Organ Sharing method to compute the MELD score. With this method the MELD score is based on INR, serum bilirubin, serum creatinine, and—for some patients—serum sodium.[5] We excluded those whose MELD score was >18 because we were interested in those with a relatively favorable MELD score. We included all the remaining 308 stable outpatients with cirrhosis and a clinical diagnosis of severe ascites and MELD score ≤18. We did not choose a lower MELD score cutoff because that would leave us with cohorts that were too small to give reliable results, so the cutoff at 18 was a compromise.

External validation cohort

Apart from the internal validation procedure described below, we conducted external validation within the trial cohort included in the SPARe-2 trial (ClinicalTrials.gov Identifier: NCT00366795).[7] This trial differed from the SPARe-1 trial in that patients were not receiving diuretics to treat ascites, only recurrent paracentesis. Clinically these patients were sicker, having refractory ascites. Of the 240 trial participants, we excluded those who were hospitalized or infected at the time of randomization, those with HCC, those with a MELD score >18, and those who had missing data on any of the predictors included in the prediction model, or missing MELD score. This left 149 patients for inclusion.

Statistical analysis

We used Cox regression with stepwise backward selection to include the strongest predictors of mortality in the prediction model. We did not consider other outcomes, e.g., cirrhosis complications, because our interest was in liver transplant candidates, and for them the most important outcomes are liver transplantation and death. We combined the stepwise approach with multivariable fractional polynomial transformation of continuous predictors, for which we used the

R package 'mfp'.[8] We considered only first-degree fractional polynomials, and we used a p-value of 0.05 as the limit for preferring a fractional polynomial transformation over the simpler linear association, and a p-value of 0.05 as the limit for retaining a predictor in the prediction model. Higher p-values would lead to more transformed predictors and more predictors in the final model, but not necessarily better prediction outside the development cohort. The predictors in the final model, their transformations (if any), and their hazard ratios formed the prediction model, which we called the SAM (Severe Ascites Mortality) score.

Model performance in the development cohort

We used the C index and the Brier score as the measures of prediction model performance. The C index is a measure of discrimination, i.e. the ability to rank patients correctly according to mortality risk. Specifically, it is the proportion of all possible patient pairs in the cohort that are ranked correctly with respect to mortality risk. It consequently ranges from 0 to 1 with 0.5 being coin-toss model performance. Our interest was in the difference in C index between the SAM score and the MELD score. The confidence interval around that difference in C index was obtained with bootstrap resampling, using 1000 samples.

The Brier score is an overall measure of prediction model performance, the squared difference between predicted and observed mortality risk during the 1-year follow-up. Consequently, a lower Brier score indicates a better model performance.[9]

Correction for optimism bias

A prediction model performs better in the cohort from which it was developed than in other cohorts because it will also model the random variation that is different in other cohorts. This difference in performance is called 'optimism bias'. It can be estimated and corrected for, and for that we used a bootstrap resampling technique described by Harrell on page 114 of his book.[10] Briefly, we took

1000 bootstrap samples of the patients in the development cohort to estimate the optimism bias in the C index (and the Brier score) for the SAM score in the development cohort. For each bootstrap sample we repeated the entire SAM score development procedure and computed the C index (and the Brier score) for the resulting SAM score in the bootstrap sample itself and in the development cohort. The difference between those two C index estimates (and Brier score estimates) is an estimate of the optimism bias; the final optimism bias estimate was estimated as the mean of the 1000 optimism bias estimates.

Model performance in external validation cohort

We computed the difference in C index and Brier score between the SAM score and the MELD score in the external validation cohort of patients included in the SPARe-2 trial. As above, the confidence interval around the C index difference (and Brier score difference) was based on bootstrap resampling, but we did not correct for optimism bias because it is not a concern in external cohorts.

We illustrated the SAM and MELD scores' ability to separate the patients in the development cohort and in the external validation cohort by observed cumulative mortality. This was done by dividing the patients in halves defined by the SAM score and MELD score and using the Kaplan-Meier estimator to compute cumulative 1-year mortality within each half.

Sensitivity analysis

We examined the impact of our choice of p-value limits for inclusion of variables in the SAM score and for preferring fractional polynomial transformation. We did this by repeating our model development procedures with different choices of p-value limits, illustrating their impact on predictors in the SAM score and effect on the differences in C index and Brier score between the SAM and MELD scores. In addition, we repeated all model development procedures forcing the

MELD score itself into the SAM score. This analysis, thus, provided an attempt to improve upon the MELD score rather than replace it.

Patients with a MELD score <15 may be considered "too well to transplant",[11] so a MELD score of 15 is a clinically relevant cutoff . Our development and validation cohorts were too small, in our judgment, to use this cutoff in our primary analyses, but we did repeat our analyses limiting the study to patients with a MELD score <15.

Results

The development cohort included 308 outpatients with severe ascites and MELD ≤18. Their characteristics are shown in Table 1. Their overall 1-year mortality was 20.4% (Figure 1). Table 2 shows the crude individual effects of the candidate predictors. The final prediction model included four variables: serum bilirubin, serum sodium, history of SBP (yes or no), and history of diabetes (yes or no). Neither of the two continuous variables needed fractional polynomial transformation. The final SAM score was computed as 0.0.0167 * bilirubin – 0.1358 * sodium; add 1.0676 if the patient has diabetes and add another 1.2866 if the patient has a history of SBP (Table 3). Notably, the physical component score of the SF-36 was not included in the final prediction model.

The SAM score yielded a C index of 0.722, and the MELD score yielded a C index of 0.626. Consequently, the difference without correction for optimism bias was 0.722 – 0.626 = 0.096. We estimated that the optimism bias was 0.083, meaning that the bias-corrected estimate of the C index in the development cohort was 0.096 – 0.083 = 0.01 (95% CI –0.08 to 0.11) (Table 4, highlighted row).

The SAM score yielded a Brier score of 0.082 vs. 0.092 for the MELD score. The difference was 0.082 – 0.092 = –0.010, suggesting that the SAM score provided better predictions. When we corrected for the optimism bias, however the difference was –0.010 – (–0.015) = 0.006 (95% –

0.004 to 0.016), implying that the MELD score provided marginally better predictions than the SAM score (Table 3, highlighted row).

External validation cohort

The external validation cohort included 149 patients whose overall 1-year mortality was 22.4% (Figure 1), consistent with the assumption that these patients had more severe ascites than patients in the development cohort. In the external validation cohort, the C index was 0.62 for the SAM score and 0.66 for the MELD score. Thus, the difference was 0.62 – 0.66 = –0.04 (95% CI –0.16 to 0.08), indicating that the MELD score was better. The Brier score was 0.104 for the SAM score and 0.098 for the MELD score, for a difference of 0.104 – 0.098 = 0.005 (95% CI –0.005 to 0.015). This difference, too, suggested that the MELD score was superior. Moreover, Figure 1 shows that the separation between patients with the best and the worst prognosis is slightly better with the MELD score than with the SAM score.

Sensitivity analysis

The sensitivity analysis showed that the SAM score would have changed a little if we had chosen different p-value cutoffs, or if we had aimed to improve upon the MELD score rather than replacing it. No version of the SAM score was statistically significantly superior to the MELD score (Table 4).

When we limited the study to patients with a MELD score <15, the development cohort included 202 patients (25 of whom died during the follow-up), and the external validation cohort included 95 patients (14 of whom died). With this smaller development cohort the SAM score included sodium (hazard ratio = 0.85, 95% CI 0.77 to 0.93), history of SBP (hazard ratio = 3.62, 95% CI 1.53 to 8.56), and history of diabetes (hazard ratio = 3.51, 95% CI 1.55 to 7.95), all of which were also identified in the primary analysis but with slightly different weights (compare Table 3). However,

Supplementary Figure 1 shows that this version of the SAM score had little predictive ability in the external validation cohort, thus demonstrating very significant optimism bias.

Discussion

We used data from trial participants with cirrhosis, severe ascites, and a MELD score below 18 to develop a prediction model (the SAM score) that performs better than the MELD score. We failed to do so; whether we developed a *de novo* prediction model or tried to improve upon the MELD score itself, our prediction model did not perform better.

Many prediction models are published and promoted enthusiastically, whereafter no one ever tries to validate them.[12] It is a strong point of our study that we could develop *and* validate the SAM score. Our study provides support for maintaining the MELD score as the preferred clinical tool for ranking patients with cirrhosis and ascites according to their risk of death. We anticipated that we could design a scoring system that was better than MELD because we limited the cohort to a narrow MELD interval and had access to a rich dataset. Here we discuss five possible explanations for the failure to trump the MELD score.

First, whether the MELD score is in fact the optimal scoring system and cannot be improved upon. We should not consider any prediction model as 'optimal' and beyond improvement. Many factors other than the MELD score components impact on mortality. We, too, demonstrated that such factors exist, among them presence of diabetes, but the fact that diabetes increases mortality *on average* does not imply that patients with diabetes *always* die sooner than patients without diabetes. Mortality prediction in cirrhosis is inherently difficult, as any seasoned clinical hepatologist will attest, and we must continue our attempts to improve upon the MELD score.

Second, maybe we did not have data on the strongest predictors of mortality. We speculate that this is a likely explanation. Others have pointed to the importance of considering hepatic

encephalopathy, sarcopenia, frailty, cardiovascular dysfunction and comorbidities alongside the MELD score.[4, 13-17] Our study was not suited to examine the impact of these factors. We did include some patients with hepatic encephalopathy grade 1, but patients with higher grades were excluded from the trials.

Third, the difference in sickness between our development and validation cohorts might explain why the SAM score did not perform better than the MELD score. However, we provided evidence against this explanation: With correction for optimism bias, performance was not noticeably better in the development cohort, either. Thus this is not a credible explanation.

Fourth, due to a small number of outlier patients, the effects of diabetes and SBP on mortality were much stronger in the development cohort than in the validation cohort (cf. Table 2). This explanation requires that some patients in the development cohort had diabetes and/or SBP and a very short survival time. The substantial optimism bias supports this explanation, i.e., that single patients could have exerted an unduly strong influence on our findings. However, we could not identify the patients responsible (Supplementary Figure 2). Thus we find this an unlikely explanation.

Fifth, whether the SAM score is truly better than the MELD score, but the C index fails to recognize this. It is well-established that the C index is not a sensitive measure of predictive ability, i.e., even a factor that has a strong association with mortality may not increase the C index by a large amount.[18] For example, the MELDNa score was introduced on the basis of findings that, among patients with identical MELD scores, those with a low serum sodium had up to twice the mortality risk of those with a serum sodium of 135 mmol/liter, yet the difference in C index between MELDNa and MELD was only 0.015 (0.868 vs. 0.883).[19] For that reason we examined another indicator of predictive ability, the Brier score, but that did not affect our conclusion. We could have

chosen other measures of predictive ability, as there are many to choose from,[9] but there is nothing in our findings to suggest that it would have changed our conclusion, either. Thus we find this an unlikely explanation.

In conclusion, it was as long ago as in 2006 that Biggins and colleagues wrote that "the current goal is to reliably identify a subgroup of patients with severe or complicated ascites that should be compensated with additional MELD score points".[1] Patients in such a subgroup should have a *consistently* higher mortality than other patients; it does not suffice that the characteristic defining the subgroup increases mortality in *some* patients. We were unable to identify those defining characteristic(s), although we had data on many relevant candidate predictors including measures of quality of life. Our preferred explanation to this conclusion is that we did not have data on the strongest prognostic factors, whichever they are. The clinical implication of our findings is that the current practice of using the MELD score to rank patients by mortality risk may continue despite an acknowledgement that we need better tools to predict the mortality of patients with cirrhosis and severe ascites.

References

1. Heuman DM, Abou-Assi SG, Habib A, et al. Persistent ascites and low serum sodium identify patients with cirrhosis and low MELD scores who are at high risk for early death. *Hepatology* 2004; 40: 802-10.
2. Somsouk M, Kornfield R, Vittinghoff E, Inadomi JM, Biggins SW. Moderate ascites identifies patients with low model for end-stage liver disease scores awaiting liver transplantation who have a high mortality risk. *Liver Transpl* 2011; 17: 129-36.
3. Biggins SW, Colquhoun S, Gish RG, Runyon BA. Model for end-stage liver disease (MELD) exception for ascites. *Liver Transpl* 2006; 12: S88-90.

4. Atiemo K, Skaro A, Maddur H, et al. Mortality risk factors among patients with cirrhosis and a low model for end-stage liver disease sodium score (≤15): An analysis of liver transplant allocation policy using aggregated electronic health record data. *Am J Transplant* 2017; 17: 2410-9.
5. United Network for Organ Sharing. Clerical changes for implementation of adding serum sodium to the MELD score. https://optn.transplant.hrsa.gov/media/1575/policynotice_20151101.pdf. Accessed 8 April 2020.
6. Macdonald S, Jepsen P, Alrubaiy L, Watson H, Vilstrup H, Jalan R. Quality of life measures predict mortality in patients with cirrhosis and severe ascites. *Aliment Pharmacol Ther* 2019; 49: 321-30.
7. Wong F, Watson H, Gerbes A, et al. Satavaptan for the management of ascites in cirrhosis: Efficacy and safety across the spectrum of ascites severity. *Gut* 2012; 61: 108-16.
8. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of sas, stata and R programs. *Comput Stat Data Anal* 2006; 50: 3464-85.
9. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 2010; 21: 128-38.
10. Harrell Jr. FE. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. 2nd ed. New York, USA: Springer; 2015.
11. Merion RM. When is a patient too well and when is a patient too sick for a liver transplant? *Liver Transpl* 2004; 10: S69-S73.
12. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. *J Clin Epidemiol* 2008; 61: 1085-94.

13. Lucidi C, Ginanni Corradini S, Abraldes JG, et al. Hepatic encephalopathy expands the predictivity of model for end-stage liver disease in liver transplant setting: Evidence by means of 2 independent cohorts. *Liver Transpl* 2016; 22: 1333-42.
14. Gadiparthi C, Cholankeril G, Yoo ER, Hu M, Wong RJ, Ahmed A. Waitlist outcomes in liver transplant candidates with high MELD and severe hepatic encephalopathy. *Dig Dis Sci* 2018; 63: 1647-53.
15. Montano-Loza AJ, Duarte-Rojo A, Meza-Junco J, et al. Inclusion of sarcopenia within MELD (MELD-sarcopenia) and the prediction of mortality in patients with cirrhosis. *Clin Transl Gastroenterol* 2015; 6: e102.
16. Lai JC, Rahimi RS, Verna EC, et al. Frailty associated with waitlist mortality independent of ascites and hepatic encephalopathy in a multicenter study. *Gastroenterology* 2019; 156: 1675-82.
17. Mazumder NR, Atiemo K, Kappus M, et al. A comprehensive review of outcome predictors in low MELD patients. *Transplantation* 2020; 104: 242-50.
18. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928-35.
19. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. *N Engl J Med* 2008; 359: 1018-26.

Table 1. Characteristics of patients in the development and external validation cohorts. Continuous variables are shown with median and 25th and 75th percentiles, categorical variables with percentages of the total.

	Development (N = 308)	**Validation (N = 149)**
MELD score	13 (10–16)	13 (10–16)
Male	219 (71%)	96 (64%)
Age	59 (52–66)	57 (50–61)
Refractory ascites	182 (59%)	137 (92%)
Alcoholic cirrhosis	208 (68%)	82 (55%)
History of SBP	51 (17%)	21 (14%)
History of variceal bleeding	59 (19%)	29 (19%)
Diabetes	70 (23%)	31 (21%)
Lactulose user	96 (31%)	45 (30%)
NSBB user	145 (47%)	72 (48%)
Peripheral edema	164 (53%)	95 (64%)
Hepatic encephalopathy grade 1[†]	11 (4%)	11 (7%)
CirCom score > 0	71 (23%)	27 (18%)
INR	1.4 (1.2–1.5)	1.4 (1.2–1.5)
Bilirubin	23 (15–33)	21 (13–33)
Creatinine	80 (66–97)	78 (64–96)
Sodium	137 (135–139)	137 (134–140)
White blood cells	5.7 (4.2–7.1)	5.5 (4.2–7.4)
Albumin	34 (31–38)	33 (29–37)
Platelets	145 (107–204)	140 (93–207)
Hemoglobin	119 (105–129)	114 (102–129)
Potassium	4.3 (3.9–4.6)	4.3 (4.0–4.8)
ALT	21 (15–30)	23 (11–35)
Alkaline phosphatase	113 (87–155)	119 (93–179)
Mean arterial pressure	83 (75–90)	82 (73–92)
Mobility impairment (0–100)	27 (5–51)	37 (10–74)
SF-36 Physical component score (0–100)	37 (31–43)	Not measured
SF-36 Mental component score (0–100)	43 (35–53)	Not measured

[†] Patients with hepatic encephalopathy grade 2–4 were excluded.

Table 2. Crude effects of candidate variables on the mortality hazard in the development and validation cohorts. Bold font highlights the factors included in the Severe Ascites Mortality (SAM) score.

	Development	**Validation**
MELD score, per point	1.16 (1.07–1.27)	1.21 (1.08–1.35)
Male, yes vs. no	0.92 (0.53–1.60)	1.04 (0.51–2.09)
Age, per 10 years increase	1.30 (1.01–1.67)	0.93 (0.65–1.34)
Refractory ascites, yes vs. no	1.30 (0.76–2.21)	1.47 (0.35–6.15)
Alcoholic cirrhosis, yes vs. no	0.71 (0.42–1.20)	0.46 (0.23–0.93)
History of SBP, yes vs. no	**3.28 (1.91–5.63)**	**1.31 (0.54–3.16)**
History of variceal bleeding, yes vs. no	2.06 (1.18–3.59)	1.88 (0.90–3.93)
Diabetes, yes vs. no	**2.21 (1.31–3.75)**	**1.23 (0.56–2.71)**
Lactulose user, yes vs. no	1.29 (0.76–2.20)	2.40 (1.22–4.70)
NSBB user, yes vs. no	0.90 (0.54–1.51)	1.32 (0.67–2.60)
Peripheral edema, yes vs. no	1.08 (0.65–1.81)	1.22 (0.59–2.50)
Hepatic encephalopathy grade 1, yes vs. no	0.93 (0.23–3.82)	1.58 (0.56–4.49)
CirCom score > 0, yes vs. no	1.44 (0.83–2.50)	1.15 (0.50–2.64)
INR, per 1.0 increase	1.40 (0.51–3.84)	2.20 (0.68–7.13)
Bilirubin, per 10 µmol/L increase	**1.21 (1.05–1.39)**	**1.24 (1.08–1.42)**
Creatinine, per 10 mmol/L increase	1.06 (0.96–1.18)	0.95 (0.82–1.11)
Sodium, per 5 mmol/L increase	**0.58 (0.41–0.83)**	**0.65 (0.43–0.98)**
White blood cells, per 1*10^9/L increase	1.02 (0.91–1.13)	1.04 (0.92–1.17)
Albumin, per 1 g/L increase	0.96 (0.91–1.01)	0.94 (0.88–0.99)
Platelets, per 50 *10^9/L increase	0.85 (0.71–1.02)	0.85 (0.70–1.05)
Hemoglobin, per 10 g/L increase	0.97 (0.83–1.12)	0.97 (0.82–1.15)
Potassium, per 1 mmol/L increase	1.74 (1.09–2.76)	1.83 (1.10–3.06)
ALT, per 10 U/L increase	1.06 (0.96–1.18)	1.00 (0.90–1.10)
Alkaline phosphatase, per 10 U/L increase	1.00 (0.97–1.03)	1.00 (0.98–1.03)
Mean arterial pressure, per 10 mmHg increase	0.80 (0.64–1.01)	0.92 (0.66–1.27)
Mobility impairment, per 10 points increase	1.03 (0.95–1.13)	0.99 (0.90–1.10)
SF-36 Physical component score, per 10 points increase	0.78 (0.57–1.07)	N/A
SF-36 Mental component score, per 10 points increase	1.13 (0.91–1.40)	N/A

Table 3. Hazard ratios and log(hazard ratios) for the variables in the Severe Ascites Mortality (SAM) score. It follows that the SAM score is computed as 0.0.0167 * bilirubin – 0.1358 * sodium; add 1.0676 if the patient has diabetes and add another 1.2866 if the patient has a history of SBP.

	Hazard ratio	log(hazard ratio)
Bilirubin, per 1 μmol/L increase	1.02 (1.00 to 1.03)	0.0167
Sodium, per 1 mmol/L increase	0.87 (0.81 to 0.94)	–0.1358
History of diabetes, yes vs. no	2.91 (1.68 to 5.04)	1.0676
History of SBP, yes vs. no	3.62 (2.06 to 6.37)	1.2866

Table 4. Sensitivity analysis. A positive ΔC index and a negative ΔBrier score indicate that the Severe Ascites Mortality (SAM) score has better performance than the MELD score. Readers may notice that the results for the development cohort differ across P-value limits although the variables in the model are the same. The explanation is that those results depend on the level of optimism bias, which is estimated from 1000 bootstrap samples and depends on the choice of P-value limits.

P-value limit, variable inclusion	P-value limit, variable transformation	Included variables	ΔC index, development cohort (corrected for optimism bias)	ΔBrier score, development cohort (corrected for optimism bias)	ΔC index, external validation cohort	ΔBrier score, validation cohort
De novo model						
0.01	0.01	Sodium, history of SBP, history of diabetes	0.02 (–0.09 to 0.12)	0.001 (–0.009 to 0.011)	–0.08 (–0.21 to 0.05)	0.011 (0.001 to 0.022)
0.05	**0.05**	**Bilirubin, sodium, history of SBP, history of diabetes**	**0.01 (–0.08 to 0.11)**	**0.006 (–0.004 to 0.016)**	**-0.04 (–0.16 to 0.08)**	**0.005 (–0.005 to 0.015).**
0.10	0.10	Bilirubin(^-0.5), sodium, history of SBP, history of diabetes, creatinine	0.01 (–0.07 to 0.09)	0.008 (–0.002 to 0.017)	–0.06 (–0.16 to 0.04)	0.008 (–0.001 to 0.016)
0.20	0.20	log(bilirubin), sodium, history of SBP, history of diabetes, physical component score, potassium(^3), alkaline phosphatase(^3), age	0.03 (–0.05 to 0.12)	0.007 (–0.004 to 0.018)	N/A (physical component score unavailable)	N/A (physical component score unavailable)
Extension of MELD						
0.01	0.01	MELD, sodium, history of SBP, history of diabetes	0.06 (–0.02 to 0.14)	–0.003 (–0.012 to 0.007)	–0.02 (–0.11 to 0.07)	0.004 (–0.004 to 0.012)
0.05	0.05	MELD, sodium, history of SBP, history of diabetes	0.04 (–0.04 to 0.11)	0.003 (–0.007 to 0.012)	–0.02 (–0.11 to 0.07)	0.004 (–0.004 to 0.012)
0.10	0.10	MELD, sodium, history of SBP, history of diabetes	0.03 (–0.05 to 0.10)	0.006 (–0.003 to 0.016)	–0.02 (–0.111 to 0.07)	0.004 (–0.004 to 0.012)
0.20	0.20	MELD, sodium, history of SBP, history of diabetes, age, physical component score	0.02 (–0.05 to 0.10)	0.008 (–0.002 to 0.019)	N/A (physical component score unavailable)	N/A (physical component score unavailable)

Figure 1. Cumulative mortality in the development cohort (left) and in the external validation cohort (right). Patients in each cohort were divided in halves by Severe Ascites Mortality (SAM) score (black lines) and MELD score (gray lines). The dotted line shows the overall cumulative mortality. The SAM score provides better separation in the development cohort and slightly worse separation in the external validation cohort.