Improving survival prediction models for liver transplantation candidates

ISBN: 978-94-6419-485-2
Printing: Gildeprint
Cover: Cover design by Anke @ Persoonlijk Proefschrift

Financial support for the publication of this thesis was kindly provided by the LUMC biostatistics department, the Dutch Transplant Society, Chipsoft and Chiesi.

Improving survival prediction models for liver transplantation candidates

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties,
te verdedigen op woensdag 29 juni 2022
klokke 15.00 uur

door

Ben Felix Jacob Goudsmit

Promotores:

prof. dr. B. van Hoek
prof. dr. H. Putter

Copromotor:

dr. A.E. Braat

Leden promotiecommissie:

prof. dr. H.J. Metselaar	Erasmus Universiteit Rotterdam
prof. dr. J. Pirenne	Katholieke Universiteit Leuven
prof. dr. I.P.J. Alwayn	
prof. dr. M. Fiocco	
dr. M.J. Coenraad	

Contents

Chapter 1

General introduction

"*Zij zag, zij zag, wat niemand zag.*

Maar ach, voor haar kwam toch een dag

Dat zij het niet precies kon zien

Heeft u dat ook misschien"

— Heinz Hermann Polzer

Research in context

This thesis focuses on survival prediction models in liver transplantation (LT). Predicting survival is important because it is used to prioritize patients in need of transplantation. Many patients are not transplanted in time, due to the shortage of donor liver grafts.[1,2] Optimizing survival prediction models is therefore a matter of life and death.

A short history of liver transplantation

The shortage of donor liver grafts and the subsequent need for survival prediction exist because of the success of LT as treatment. Thomas Starzl was the first to perform a LT in 1963, trying to save a severely ill child.[3] In this first and later attempts, patients often died shortly after transplantation. Only a laboratory pig managed to survive many years without immunosuppression and therefore became his favorite transplant mascot.[4] Performing LT remained an experimental treatment until further improvements in immunosuppression, operation and preservation techniques, diagnosis of liver diseases, and postoperative management had been made.[5,6] In the late 70s the early LT programs started, in 1983 LT was declared not experimental anymore, and only in the 90s many other LT centers started. Five-year post-transplant survival probabilities increased from 21% to 71% in 25 years.[7]

Through these improvements, an increasing diversity and number of patients were eligible and could be treated.[5] This created shortages of donor livers, as the number of patients in need of LT ('LT candidates') outnumbered the available donor livers. Even now, despite the development of donation after circulatory death, living donor liver transplantation, and machine perfusion for marginal organs, the imbalance between available donor organs and number of candidates persists, with an average waiting time for LT of five months in the Eurotransplant region and eight months in the United Network for Organ Sharing (UNOS) regions.[1,2] As a result, nowadays many patients are still not transplanted in time and die on the waiting list. For the Eurotransplant region in 2020, 25.3% (374/1481) of the LT candidates died on the waiting list.[1] In the US in 2019, LT came too late for 18.3% (2405/13,093) of the waiting patients.[2] Indeed, the mortality of patients on the liver waiting list is highest compared to all other transplantable organs. Therefore, the

development and improvement of survival prediction for the LT waiting list is most important.

The start of liver allocation

Because of the plethora of indications and increasing number of patients listed for LT, donated liver grafts needed to be distributed in a systematic and just way. Thus, the field of liver allocation came into existence, where allocation can be defined as "the process of giving someone their part of a total amount of something to use in a particular way."[8] Initially, liver grafts were assigned to patients without applying uniform rules, because there were only a small number of patients and grafts involved. Then, following the field of kidney allocation, LT was offered based on waiting time, i.e., first come first served.[5] With an increasing number of patients involved, waiting times increased and therefore governmental regulation was initiated, which gave rise to organ procurement organizations (OPO). Importantly, it was shown that waiting time did not correlate well with waiting list mortality.[9] Thus, consensus was reached that the most relevant consideration was not how long a patient had waited, but what the risk of death was while waiting. In other words, the principle of transplanting the sickest first was employed.[10] As a result, the OPO's effectuated policies that sought to incorporate measures of disease severity into allocation. The rationale was that prioritizing patients with the highest expected mortality without LT would reduce deaths among waiting patients.

Survival prediction

Because liver allocation aims to prioritize patients who will die soonest without transplantation, it is important to understand survival prediction models.

Clinical information of a patient can relate to the true patient state that is either currently present (diagnosis) or will be present in the future (prognosis).[11] In the setting of survival prediction, future survival (prognosis) of a patient is estimated. Typically, only some patients will experience the event of interest (death) within the studied time. Therefore, survival times will be unknown for many patients, which is known as censoring of outcome.[12] Right censoring can occur because (1) the

patient did not die yet, (2) because the patient was lost during study follow-up, or (3) another event took place which disabled further follow-up (e.g., transplantation prevented further follow-up for death on the waiting list). The fact that patients can be censored makes survival analysis difficult, as the aim is to use all the available information and to not discard follow-up times of patients without the event. In the setting of LT and this thesis, survival analysis will mostly be done for LT candidates on the waiting list. The survival probability on the waiting list is the chance that a patient survives from a specified time of origin (e.g., first registration) until a future time point (e.g., 90 days). Post-transplant survival will also be used in chapter 6 of this thesis and is defined as the probability of survival from the moment of transplantation to the earliest of post-transplant death, loss to follow-up, or end of study.

When predicting survival, it is important to consider that patient characteristics can affect survival and that these characteristics can be differently distributed within the studied population. For example, older patients might more frequently have more severe disease. The Cox proportional hazards model is used to study the impact of one variable, adjusted for the impact of other variables, and to estimate the effect size of each variable.[13] For the above mentioned example, a Cox model could show that for two patients with the disease severity, a higher age would increase the risk of death. The probability that a patient on the waiting list dies at a given moment is called the hazard $h(t)$:

$$h(t) = h_0(t) \times e^{b_1 x_1 + b_2 x_2 + ...}$$

This formula shows that the hazard depends on the chosen time t, which seems likely for the waiting list. It also depends on chosen predictors $(x_1, x_2, ..)$ that have a certain impact, which is expressed through the size of the coefficients $(b_1, b_2, ..)$. The baseline hazard h_0 is the hazard if all predictors $(x_1, x_2, ..)$ were set to zero. For example, if age were used to predict survival, h_0 would be the instantaneous risk of death at age 0.

Measuring model prediction performance

The outcome of analysis is often binary (death or alive) and therefore predictions of this outcome can be expressed as absolute risks (e.g.,

60% chance of dying in the next five years). Because the aim is to best estimate the true future patient state, it is important to assess performance of a prediction model. The first essential measure is discrimination, which assesses whether the model can discriminate between patients based on their risks of the outcome (e.g., death). Patients with the event should have higher risk estimates than patients without the event. For allocation purposes, this means that patients with a higher risk of death will be ranked above patients with lower risks. Therefore, good model discrimination is essential for allocation. Discrimination is often expressed through the concordance statistic (or c-index). Imagine that the prediction model is offered information on two patients from the population and that it must decide which patient has a higher risk of death than the other. The percentage of correct decisions by the model corresponds to the c-index. A c-index of 0.5 means that the model is as good as flipping a coin. A c-index of 1 means that a model perfectly ranks patients, which in practice is not possible. In real cohort data, a c-index above 0.8 is considered excellent.

The next essential measure of model performance is calibration, which measures model accuracy. In other words, calibration tells how well the predicted risks match the observed risks in the studied population. Discrimination indicates which patient has a higher risk, but it tells nothing about the value of that risk (e.g., 10% or 90% chance of death). Calibration assesses the absolute risks and therefore is essential for research and communication with patients.[14] For example, clinicians and patients may make decisions based on the expected risk for an event (e.g., decide to transplant a patient based on an expected high risk of death without treatment). Strong over- or underestimation of these risks are unacceptable for clinical practice.[15] Thus, survival prediction models aim to estimate future survival, which is essential for allocation of scarce liver grafts.

The MELD score

Several survival prediction models have previously been applied in liver allocation. Perhaps the most noteworthy is the Child-Turcotte-Pugh (CTP) score, which was the first widespread model used to reflect patient disease severity for the LT waiting list.[16] Although this score was a well-established predictor of mortality in cirrhotic patients, it failed in

effective sickest-first allocation for several reasons. One important limitation was that patients were categorized in only three groups. These groups still encompassed many patients with varying disease severity and therefore risk stratification was not precise. Also, within each group, waiting time was still used as main prioritizing principle. Another important limitation of the CTP score was the subjective grading of ascites and encephalopathy, which could lead to inter-observer variability when scoring disease severity.[17] To reach a more reproducible and objective representation of disease, the Model for End-stage Liver Disease (MELD) was considered.

The MELD score was developed in the Mayo Clinic to predict early survival after transjugular intrahepatic portosystemic shunt (TIPS) placement in 231 cirrhotic patients.[18] These patients received TIPS to prevent variceal rebleeding and to treat refractory ascites. The seminal study by Malinchoc et al. found that three blood measurements best predicted survival, i.e., serum bilirubin, creatinine, and the international normalized ratio for prothrombin time (INR). After validation in 71 Dutch patients, the original MELD equation was proposed:

$$MELD = 9.57 \times ln(creatinine) + 3.78 \times ln(bilirubin) + 11.2 \times ln(INR) + 6.43 \times (causeofcirrhosis)$$

Because survival after TIPS mainly depended on the severity of liver disease, it was hypothesized that MELD was also suitable for survival prediction in patients without TIPS. Thus, retrospective validation was done to investigate whether MELD could also be applied to different etiologies and severities of liver disease.[19] It was found that the cause of cirrhosis could be excluded from the equation without lowering predictive performance. Thus, the final form became:

$$MELD = 9.57 \times ln(creatinine) + 3.78 \times ln(bilirubin) + 11.2 \times ln(INR) + 6.43$$

Then, because it was considered as potential allocation model, MELD was prospectively evaluated by ranking patients on the waiting list.[20] In 2002, MELD was applied as the basis of liver allocation in the United States (US). In 2006, the Eurotransplant region followed. In 2016, the US progressed to a newer form of MELD: the MELD sodium (MELD-Na) score.[21] However, MELD has remained unchanged as the main driver for liver allocation in the Eurotransplant region.

Justifying why research is needed: central argument

In the Eurotransplant region, the fundamental model that predicts survival for liver allocation has remained unchanged since 2006. It is therefore easy to see the gap this thesis aims to fill. The current LT allocation system prioritizes the sickest patients, based on estimated survival on the waiting list. Thus, the primary goal of this thesis is to improve models that predict survival in LT candidates. These improved models could help to distribute liver grafts in the best way possible. To understand the possible improvements, some current problems are outlined below.

Thesis layout: research questions and addressed problems

Considering sodium

The MELD score uses three blood measurements, i.e., serum bilirubin, creatinine, and the INR. In cirrhotic patients, hyponatremia indicates worse survival.[22,23] This is because the kidneys fail to compensate lowered splanchnic blood pressure caused by cirrhosis-induced vasodilatation. As such, hyponatremia does not represent kidney failure per se, but it indicates a decompensation of regulating systems in a setting of portal hypertension. It was shown that sodium (Na) affected death risk corrected for MELD.[24–26] Therefore, the addition of sodium (Na) to MELD (MELD-Na) was investigated in the US. Important advantages of serum sodium were that it was readily available and could be measured reliably and objectively. Results showed that MELD-Na better predicted survival on the waiting list and therefore possibly enabled better sickest-first LT allocation.[21] In 2016, the US implemented MELD-Na as the basis of LT allocation. Subsequent evaluation indeed showed a reduction in waiting list mortality.[27] Thus, as a starting point for this thesis, it was suggested to validate MELD-Na for the Eurotransplant region. Therefore, in **Chapter 2**, we hypothesized that MELD-Na would also improve LT candidate survival prediction in the Eurotransplant region.

Updating coefficients

After validating MELD-Na, the author questioned whether the current forms of MELD and MELD-Na were suitable for the Eurotransplant region. This question arose because a model best represents the population it is fit in. As the MELD score was fitted 20 years ago in 231 US patients,[18] it was assumed to be a bad representation of the current Eurotransplant population. Therefore, in **Chapter 3**, the second question posed in this thesis was whether refitting MELD for the Eurotransplant region would improve survival prediction. To understand what refitting means and how improvements could be made, consider the abovementioned MELD equation. It shows three parameters (bilirubin, creatinine, and INR) and their coefficient (relative weight). The values of these coefficients represent the relations of the variables to survival in the studied population. However, MELD's coefficients were set in a small (n=231) and likely unrepresentative population.[18] Population changes (most notably disease incidence) have decreased MELD's predictive power over the years.[28] Even in the US, updating MELD showed an improvement in survival prediction.[29,30] We hypothesized that in a different population, like the Eurotransplant region, MELD variables relation to survival would be different. It is remarkable that the Eurotransplant region has been using a MELD equation that is 20 years old. Thousands of patients have been prioritized based on an unadjusted and possibly suboptimal model.

Past and current disease

When refitting MELD, the question arose why only one baseline measurement was used to fit the model. The third problem addressed in this thesis therefore was that current survival prediction for LT prioritization is based on one moment in time, i.e., the last available measurement of MELD. This is not the moment the model was fit for. Previous measurements are also ignored. In the Eurotransplant region at the end of 2020, patients on the LT waiting list spent a median time of 10 months waiting (*Eurotransplant public statistics library: 3085P_All ET*). During this time, disease develops and expected survival changes. This past course of disease encompasses valuable information for the future survival probability and thus patient priority for LT. Although it is currently ignored by MELD, in clinical practice it would be consid-

ered undesirable to ignore previous disease information. In **Chapter 4** we hypothesized that using both previous and current disease severity to predict survival would be an improvement. The idea was to better mimic clinical survival prediction, as an experienced clinician would consider both previous and current disease development to estimate patient prognosis. To achieve this, joint models (JMs) were applied.[31] The JM combines longitudinal and survival analysis. This way, complex questions can be answered, such as: what is the effect of a change in MELD over time on future patient survival? Importantly, JMs yield predictions that are dynamic (updated based on accumulating evidence) and personalized (for the population average and individual). However, JMs were never applied on a large scale in medicine, also not in the field of LT.

Acute-on-chronic liver disease

Most patients on the LT waiting list have chronic liver disease, which gradually worsens in severity. However, some patients can develop acute-on-chronic liver failure (ACLF). ACLF is a syndrome characterized by three major features: systemic inflammation, relationship with precipitating events (e.g., infections or alcoholic hepatitis), and an association with single- or multi-organ failure.[32] As a result, in ACLF mortality is high and a proportion of these patients urgently needs LT for treatment. However, the MELD score underestimates mortality in ACLF, because it 'only' measures liver and kidney failure.[33] In **Chapter 5**, JMs were applied to model disease and survival in patients with ACLF. We proposed that the dynamic JMs would be valuable in ACLF, because ACLF disease severity and mortality change rapidly over time.[34] The JM can consider both the value of disease severity and its rate of change at each moment in time. Analogous to speed, one can measure a value (e.g., 15 m/s) and a rate of change (e.g., an acceleration of 5 m/s^2). The rate of change indicates worsening, stable, or improving disease severity, which is valuable prognostic information.

Benefit of transplantation: life years gained

The LT waiting list prioritizes the sickest patients to receive transplantation offers first. It is based on the principle of urgency, to prevent

deaths on the waiting list. However, this ignores post-transplant outcomes, which, in an extreme example, could result in transplanting patients who die the next day. Also, patients could be transplanted even though it does more harm than good.[35] An alternative principle of allocation could be based on survival benefit, or the life years gained from transplantation.[36] Survival benefit is calculated by comparing the estimated survival with and without LT. For the clinician, it is intuitive to weigh the possible consequences of (not) treating a patient. This is especially true for LT, because it is a treatment with inherent increased risk of death due to surgery and e.g. infections due to post-transplant immunosuppression.[37] Also, there are far more patients in need of transplantation than there are available liver grafts,[1,2] which further necessitates the prevention of futile LT.[38,39] Thus, it is relevant to investigate whether and to what extend patients gain life years from LT. In **Chapter 6**, survival benefit of US LT candidates is estimated. A survival benefit comparison is made between patients with and without hepatocellular carcinoma (HCC). This is done because survival with and without LT is different between (non-)HCC patients.[40,41] Allocation also differs between HCC and non-HCC-patients, because MELD(-Na) fails to adequately predict survival in patients with HCC. To compensate this inadequacy, an alternative system of artificial exception points was devised.[42] Although the aim of the exception point system was to equalize LT access, in practice HCC patients have gained too much LT access.[43–45] In the Eurotransplant exception system, eligible HCC patients receive an initial MELD score that equals 10% (MELD 20) or 15% (MELD 22) 90-day mortality, depending on the country of listing. The initial MELD score is then increased with 10% mortality every 90 days, which intends to mimic tumor progression. It is however evident that exception points fail to represent patient characteristics and are arbitrary. Although the aim of the exception point system was to equalize LT access, in practice HCC patients gained too much LT access.[43–45] Survival benefit, based on actual patient characteristics, could therefore serve as equalizing principle for survival prediction and allocation.

In **Part IV**, this thesis will be summarized, discussed, and provided with future perspectives. The appendix provides two supplementary chapters. Lastly, a **summary in Dutch** will be given.

References

1. Eurotransplant International Foundation. Annual Report 2020.; 2020.
2. Kwong AJ, Kim WR, Lake JR, et al. OPTN/SRTR 2019 Annual Data Report: Liver. Am J Transplant. 2021;21(S2):208-315. doi: 10.1111/ajt.16494
3. Starzl TE, Marchioro TL, Von Kaulla KN, Hermann G, Brittain RS, Wadell WR. Transplantation of the liver. Surg Gynecol Obs. 1963;117(2):659-676. doi:10.1097/00000658-197808000-00001
4. Starzl TE. The Puzzle People: Memoirs of a Transplant Surgeon. University of Pittsburgh Press; 1992.
5. Merion RM, Sharma P, Mathur AK, Schaubel DE. Evidence-based development of liver allocation: A review. Transpl Int. 2011;24(10):965-972. doi:10.1111/j.1432-2277.2011.01274.x
6. Zarrinpar A, Busuttil RW. Liver transplantation: Past, present and future. Nat Rev Gastroenterol Hepatol. 2013;10(7):434-440. doi:10.1038/nrgastro.2013.88
7. Dutkowski P, De Rougemont O, Müllhaupt B, Clavien PA. Current and Future Trends in Liver Transplantation in Europe. Gastroenterology. 2010;138(3):802-809.e4. doi:10.1053/j.gastro.2010.01.030
8. Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/
9. Freeman RB, Edwards EB. Liver transplant waiting time does not correlate with waiting list mortality: Implications for liver allocation policy. Liver Transplant. 2000;6(5):543-552. doi:10.1053/jlts.2000.9744
10. Organ Procurement and Transplantation Network–HRSA. Final Rule with Comment Period.; 1998.
11. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2
12. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis Part I: Basic concepts and first analyses. Br J Cancer. 2003;89(2):232-238. doi:10.1038/sj.bjc.6601118
13. Cox DR. Regression Models and Life-Tables. J ofthe R Stat Soc. 1972;34(2):187-220. doi:10.1016/0006-8993(85)91540-9

14. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: From utopia to empirical data. J Clin Epidemiol. 2016;74:167-176. doi:10.1016/j.jclinepi.2015.12.005
15. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7
16. Pugh R, Murray-Lyon I, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg. 1973;60(8):646-649.
17. Durand F, Valla D. Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. J Hepatol. 2005;42(SUPPL. 1):100-107. doi:10.1016/j.jhep.2004.11.015
18. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology. 2000;31(4):864-871. doi:10.1053/he.2000.5852
19. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. Hepatology. 2001;33(2):464-470. doi:10.1053/jhep.2001.22172
20. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology. 2003;124(1):91-96. doi:10.1053/gast.2003.50016
21. 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(10):1018-1026. doi:10.1007/s11250-017-1262-3
22. Ginés P, Berl T, Bernardi M, et al. Hyponatremia in cirrhosis: From pathogenesis to treatment. Hepatology. 1998;28(3):851-864. doi:10.1002/hep.510280337
23. Angeli P, Wong F, Watson H, et al. Hyponatremia in cirrhosis: Results of a patient population survey. Hepatology. 2006;44(6):1535-1542. doi:10.1002/hep.21412
24. Biggins SW, Rodriguez HJ, Bacchetti P, Bass NM, Roberts JP, Terrault NA. Serum sodium predicts mortality in patients listed for liver transplantation. Hepatology. 2005;41(1):32-39. doi:10.1002/hep.20517
25. Biggins SW, Kim WR, Terrault NA, et al. Evidence-Based Incorporation of Serum Sodium Concentration Into MELD. Gastroen-

terology. 2006;130(6):1652-1660. doi:10.1053/j.gastro.2006.02.010
26. Ruf AE, Kremers WK, Chavez LL, Descalzi VI, Podesta LG, Villamil FG. Addition of serum sodium into the MELD score predicts waiting list mortality better than MELD alone. Liver Transplant. 2005;11(3):336-343. doi:10.1002/lt.20329
27. Nagai S, Chau LC, Schilke RE, et al. Effects of Allocating Livers for Transplantation Based on Model for End-Stage Liver Disease-Sodium Scores on Patient Outcomes. Gastroenterology. 2018;155(October):1451-1482. doi:10.1053/j.gastro.2018.07.025
28. Godfrey EL, Malik TH, Lai JC, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. Am J Transplant. 2019;19(12):3299-3307. doi:10.1111/ajt.15559
29. Leise MD, Kim WR, Kremers WK, Larson JJ, Benson JT, Therneau TM. A revised model for end-stage liver disease optimizes prediction of mortality among patients awaiting liver transplantation. Gastroenterology. 2011;140(7):1952-1960. doi:10.1053/j.gastro.2011.02.017
30. Sharma P, Schaubel DE, Sima CS, Merion RM, Lok ASF. Reweighting the Model for End-Stage Liver Disease Score Components. Gastroenterology. 2008;135(5):1575-1581. doi:10.1053/j.gastro.2008.08.004
31. Rizopoulos R. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. 1st ed. Chapman and Hall/CRC; 2012.
32. Arroyo V, Moreau R, Jalan R. Acute-on-Chronic Liver Failure. N Engl J Med. 2020;(382):2137-2145. doi:10.1056/NEJMra1914900
33. Hernaez R, Liu Y, Kramer JR, Rana A, El-Serag HB, Kanwal F. Model for end-stage liver disease-sodium underestimates 90-day mortality risk in patients with acute-on-chronic liver failure. J Hepatol. Published online 2020:1-9. doi:10.1016/j.jhep.2020.06.005
34. Goudsmit BFJ, Tushuizen ME, Putter H, Braat AE, van Hoek B. The role of the model for end-stage liver disease-sodium score and joint models for 90-day mortality prediction in patients with acute-on-chronic liver failure. J Hepatol. 2021;74(2):475-476. doi:10.1016/j.jhep.2020.08.032
35. Cillo U, Vitale A, Polacco M, Fasolo E. Liver transplantation for hepatocellular carcinoma through the lens of transplant benefit. Hepatology. 2017;65(5):1741-1748. doi:10.1002/hep.28998

36. Merion RM, Schaubel DE, Dykstra DM, Freeman RB, Port FK, Wolfe RA. The survival benefit of liver transplantation. Am J Transplant. 2005;5(2):307-313. doi: 10.1111/j.1600-6143.2004.00703.x
37. Martin P, Dimartini A, Feng S, Brown R, Fallon M. Evaluation for liver transplantation in adults: 2013 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation. Hepatology. 2014;59(3):1144-1165. doi:10.1002/hep.26972
38. Petrowsky H, Rana A, Kaldas FM, et al. Liver transplantation in highest acuity recipients: Identifying factors to avoid futility. Ann Surg. 2014;259(6):1186-1194. doi:10.1097/SLA.0000000000000265
39. Linecker M, Krones T, Berg T, et al. Potentially inappropriate liver transplantation in the era of the "sickest first" policy – A search for the upper limits. J Hepatol. 2018;68(4):798-813. doi: 10.1016/j.jhep.2017.11.008
40. Vitale A, Volk ML, De Feo TM, et al. A method for establishing allocation equity among patients with and without hepatocellular carcinoma on a common liver transplant waiting list. J Hepatol. 2014;60(2):290-297. doi:10.1016/j.jhep.2013.10.010
41. Berry K, Ioannou GN. Comparison of Liver Transplant-Related Survival Benefit in Patients with Versus Without Hepatocellular Carcinoma in the United States. Gastroenterology. 2015;149(3):669-680. doi:10.1053/j.gastro.2015.05.025
42. Freeman RB, Gish RG, Harper A, et al. Model for End-Stage Liver Disease (MELD) Exception Guidelines: Results and Recommendations From the MELD Exception Study Group and Conference (MESSAGE) for the Approval of Patients Who Need Liver Transplantation With Diseases Not Considered by the Standar. Liver Transplant. 2007;13(5):767-768. doi:10.1002/lt
43. Massie AB, Caffo B, Gentry SE, et al. MELD exceptions and rates of waiting list outcomes. Am J Transplant. 2011;11(11):2362-2371. doi:10.1111/j.1600-6143.2011.03735.x
44. Washburn K, Edwards E, Harper A, Freeman RB. Hepatocellular Carcinoma Patients Are Advantaged in the Current Liver Transplant Allocation System. Am J Transplant. 2010;10(7):1652-1657. doi:10.1111/j.1600-6143.2010.03127.x
45. Northup PG, Intagliata NM, Shah NL, Pelletier SJ, Berg CL,

Argo CK. Excess mortality on the liver transplant waiting list: Unintended policy consequences and model for End-Stage Liver Disease (MELD) inflation. Hepatology. 2015;61(1):285-291. doi: 10.1002/hep.27283

Part I: Forms of MELD

"All models are wrong, but some are useful."

— George Box

Chapter 2

Validation of the Model for End-stage Liver Disease sodium (MELD-Na) score in the Eurotransplant region

Goudsmit BFJ, Putter H, Tushuizen ME, de Boer J, Vogelaar S, Alwayn IPJ, et al. *American Journal of Transplantation.* 2020; doi:10.1111/ajt.16142.

Abstract

Background & Aims: The MELD score is used in the Eurotransplant (ET) region to allocate liver grafts. Hyponatremia in cirrhotic patients is an important predictor of death but is not incorporated in MELD. This study investigated the performance of the MELD-Na score for the ET region.

Methods:All adult patients with chronic liver disease on the ET liver transplantation waiting list (WL) allocated through lab MELD scores were included. The MELD-corrected effect of serum sodium (Na) concentration at listing on the 90-day WL mortality was calculated using Cox regression. The MELD-Na performance was assessed with c-indices, calibration per decile and Brier scores. The reclassification from MELD to MELD-Na score was calculated to estimate the impact of MELD-Na-based allocation in the ET region.

Results: For the 5223 included patients, the risk of 90-day WL death was 2.9 times higher for hyponatremic patients. The MELD-Na had a significantly higher c-index of 0.847 (SE 0.007) and more accurate 90-day mortality prediction compared to MELD (Brier score of 0.059 versus 0.061). It was estimated that using MELD-Na would reduce WL mortality by 4.9%.

Conclusion: The MELD-Na score yielded improved prediction of 90-day WL mortality in the ET region and using MELD-Na for liver allocation will very likely reduce WL mortality.

Introduction

Liver transplantation (LT) is the treatment of choice for end-stage liver disease. However, the number of patients in need of LT exceeds the number of available donor grafts.[1] Over the past years the prevalence and disease load of end-stage liver disease has been increasing[2-4] and is estimated to triple in the next 10 years.[5] Therefore, the limited supply of donated livers should be carefully distributed.

For optimal matching and use of donor livers in the Eurotransplant (ET) region, patients are placed on a waiting list (WL) for LT. Since 2006, the Model for End-stage Liver Disease (MELD) score has been used to rank and prioritize LT candidates in the Eurotransplant region.[6] The MELD score estimates disease severity in LT candidates based on serum creatinine, bilirubin, and the International Normalized Ratio (INR) of the prothrombin time.[7] Additionally, a high urgency (HU), i.e. United Network for Organ Sharing (UNOS) status 1, and exception point system are used for those patients in which MELD does not adequately reflect disease severity.[6]

To improve the survival prediction and allocation by the MELD score, the addition of the serum sodium (Na) concentration was proposed, as hyponatremia is an independent prognostic factor in patients with cirrhosis.[8-12] In cirrhosis, portal hypertension leads to systemic vasodilatation, secondary neurohormonal compensation and less renal excretion of solute-free water.[13,14] The severity of portal hypertension is inversely related to the serum Na concentration.[15,16] Clinically, Na levels influence the outcomes of LT candidates before and possibly even after LT.[17-20] Interestingly, in the UNOS regions, MELD-Na has been used for liver graft allocation since 2016.[21]

After the introduction of MELD-Na in the United States (US), recent evaluation showed a decline in WL mortality.[22] However, the populations of the US and Eurotransplant differ.[1,23] Recently, it was shown that differences in population characteristics influenced the predictive power of MELD and MELD-Na.[24] Therefore, MELD-Na-based allocation needs to be investigated in Eurotransplant before implementation. We hypothesized that the serum sodium levels at listing were similar between the Eurotransplant and US regions. If so, MELD-Na-based allocation could also lead to a reduction in WL mortality in the Eurotransplant region. Therefore, our aim was to validate the UNOS

MELD-Na score for the Eurotransplant region. For this, the prediction of 90-day WL mortality by the MELD-Na score was investigated in the Eurotransplant population. In addition, the potential effect of MELD-Na-based liver allocation on the Eurotransplant waiting list mortality was estimated.

Methods

Study design and population

The TRIPOD statement was used to report this study.[25] Data was retrospectively gathered from the Eurotransplant Network Information System (ENIS) and the Eurotransplant Liver Follow-up Registry (ELFR). All patients with chronic liver disease, at least 18 years old, and registered on the Eurotransplant waiting list for a first LT between January 1st 2007 and December 31st 2018 were included. Patients not allocated based on lab MELD, with HU status (i.e. UNOS status 1) or (non-)standard exception ((N)SE) points, listings for multiple organs (other than combined liver-kidney), grafts from outside Eurotransplant, or missing data at listing were excluded. The HU status is granted for acute liver failure. Exception points are given when lab MELD does not reflect disease severity or risk of dying on the waiting list (e.g. with HCC, hepatopulmonary syndrome, etc.). A detailed description of the Eurotransplant adult liver allocation is available elsewhere.[6] Patients were followed from first active listing to death, first delisting, or until 90 days. Reasons for delisting and censoring were transplantation, HU-status, (N)SE-points, and removal due to clinical condition (improvement or decline without 90-day death) or other reasons. The outcome for the prediction models was death within 90 days of listing. Removal within 90 days, due to being too sick for transplantation and subsequent death within 90 days, was also counted as 90-day mortality. Patients with a serum sodium above 150 mmol/L were excluded from the analysis, as the effects of hyponatremia were studied. The MELD score and serum Na level (mmol/L) at listing were used as predictors for the multivariate models. The sample size was set by the retrospective design of the study.

Statistical analysis

For the complete-case analysis, continuous variables were reported as mean (SD) or median (IQR). Categorical variables were reported as counts (percentage). To investigate possible selection bias, complete cases were compared to eligible patient with missing Na at listing. The MELD score was calculated according to Wiesner et al.[26] Cumulative incidence plots, accounting for the competing risks of transplantation, removal and death, were plotted for the <=130, 131-134 and >=135 mmol/L sodium levels at listing. For these groups, 90-day Kaplan-Meier survival curves were also plotted. A multivariate Cox proportional hazards (PH) regression analyzed the relation between the MELD score, Na, and 90-day mortality. The PH assumptions were checked through Schoenfeld residuals methods. A generalized additive model (GAM) with smoothing splines and fitted Cox models were used to assess the linearity of the MELD-corrected effect of Na on 90-day mortality. The upper and lower Na limits were set between 125 and 140 mmol/L, in accordance to UNOS MELD-Na.[9] Within this range, PH models adjusted for MELD and Na assessed the interaction between the predictors and calculated the hazard ratio (HR) for 90-day mortality per unit increase in MELD or Na. Then, the MELD-Na score was calculated using the standard formula.[9] Concordance statistics (c-index) were used as a measurement of discrimination between death and survival. An analysis of c-index development over the years 2007-2018 was done to assess a possible decline in c-index value for MELD and MELD-Na.[24] For the MELD-Na, a calibration plot was made of the observed and expected risk estimate per decile, with detailed risks attached in a supplementary table. As a measure of prediction error reduction, Brier scores of MELD and MELD-Na were calculated. A heatmap was constructed of the gained MELD-Na points at listing and of the differences in predicted 90-day death risk between MELD and MELD-Na scores. Interactive versions of these heatmaps were published as online supplement using the R plotly package.[27] The reclassification rate from MELD to MELD-Na score at listing was calculated. To make comparison with UNOS data possible,[9] the reclassification per MELD and MELD-Na stratum was also calculated (supplement 3). All statistical analyses were performed using SPSS v25.0 (IBM Corp, Armonk, NY) and R v3.6.1(R Foundation for Statistical Computing, Vienna, Austria).

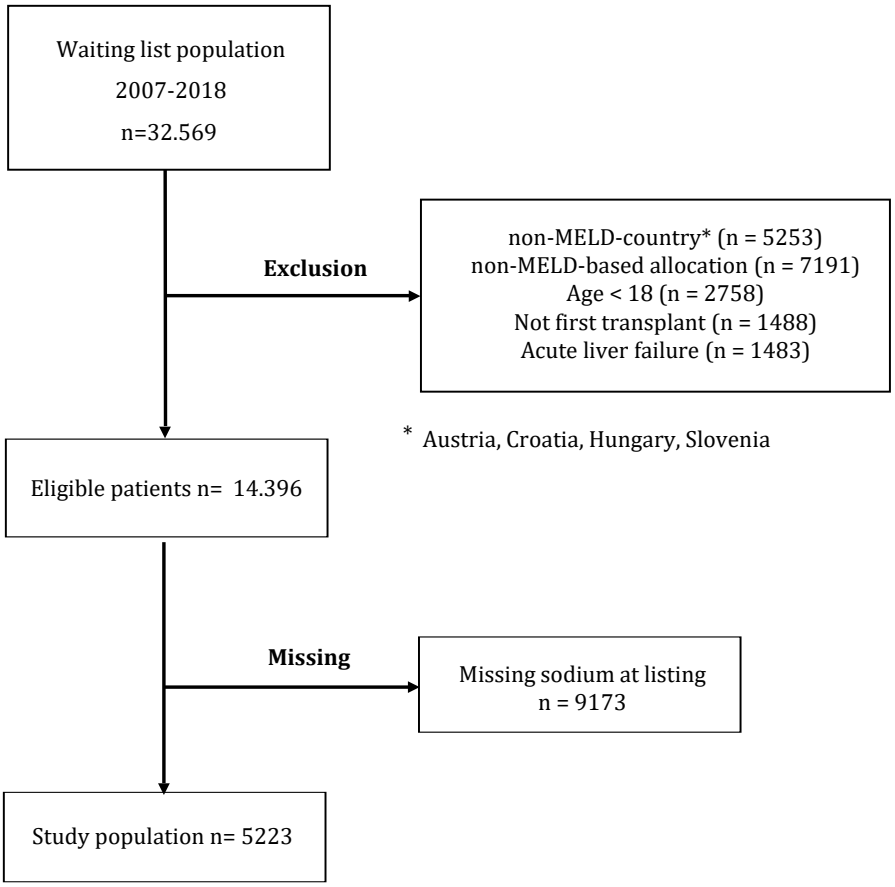


Figure 2.1: The flowchart of in- and exclusion for this study

Results

Study population

For this study, 14.396 patients were eligible. After excluding patients with missing serum Na at listing, 5223 patients were included. See Figure 2.1. The baseline characteristics of included patients at first active listing are shown in Table 2.1.

The median lab MELD score was 16 (IQR 11-21) and the median sodium concentration was 137 (IQR 134-140) mmol/L. Hyponatremia of <135, <130, and <125 mmol/L was found in respectively 28.5%, 8.8%, and

Table 2.1: Demographics of the patients at first active listing

Characteristics	(n=5223)
Age at listing	56 (49-62)
Sex (Male)	3565 (68.3)
Height (cm)	174 (167-180)
Weight (kg)	78 (67-90)
ABO	
A	2201 (42.1)
O	2081 (39.8)
B	702 (13.4)
AB	239 (4.6)
Lab-MELD at listing	16 (11-21)
MELD parameters	
Bilirubine	2.75 (1.31-6.40)
Creatinin	1.0 (1.00-1.27)
INR	1.39 (1.20-1.70)
Serum sodium at listing	137 (134-140)
Grouped sodium	
<125	136 (2.6)
<130	460 (8.8)
<135	1489 (28.5)
>=135	3734 (71.5)
MELD-Na at listing	18 (13-24)
Disease	
Alcoholic cirrhosis	1873 (35.9)
Non-cholestatic cirrhosis	1510 (28.9)
Cholestatic cirrhosis	773 (14.8)
HCC and cirrhosis	709 (13.6)
Other	355 (6.8)
Waiting list outcome (90 days)	
Still on the waiting list	2306 (44.2)
Transplanted	1114 (21.3)
Removed clinical condition	812 (15.6)
Removed other	380 (7.3)
Deceased after removal, within 90d	448 (8.6)
Deceased while listed	147 (2.8)

Note:
Median (25th-75th percentile)

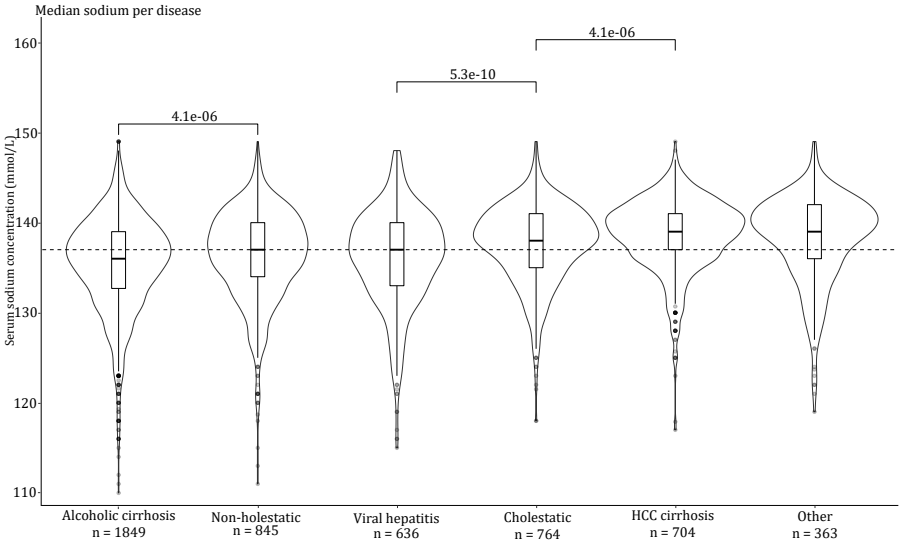


Figure 2.2: Violin plots with embedded box plots of the median serum sodium (Na) levels at listing, for the most frequent causes of liver disease. The dotted line represents the median Na of 137 mmol/L for the whole cohort. For the significant differences between Na levels, P values for pairwise comparisons are shown

2.6% of the patients. Patients with alcohol-induced cirrhosis (ALD) had the lowest median Na levels, see Figure 2.2.For the assessment of selection bias, an analysis of all eligible patients (Na present versus absent) was added (supplement 1). Compared to the included patients, eligible patients with missing serum Na were more often female (31.9% vs 35.5%) and had higher rates of alcohol- or virus-induced liver cirrhosis (respectively 35.9% vs 41.0% and 12.4% vs 15.3%, $p<0.001$). MELD scores were comparable, but excluded patients had significantly higher creatinine levels at listing (1.36 vs 1.42 mg/dL $p<0.001$).

Competing risk analysis showed that 90-day mortality and transplantation rates increased as sodium levels decreased, see Figure 2.3. Na<130, 130-134 and >=135 patients had 90-day death risks of respectively 27%, 18% and 8%. The 90-day transplant rates were respectively 33%, 27% and 18.0%. The grouped Na levels showed diverging survival curves, i.e. at lower Na levels the mortality risk increased at a higher rate (supplement 2). The 90-day death HRs for Na <130 and Na 130-134 compared to Na >=135 patients were 4.72 (95%CI 3.81-5.83), and 2.72

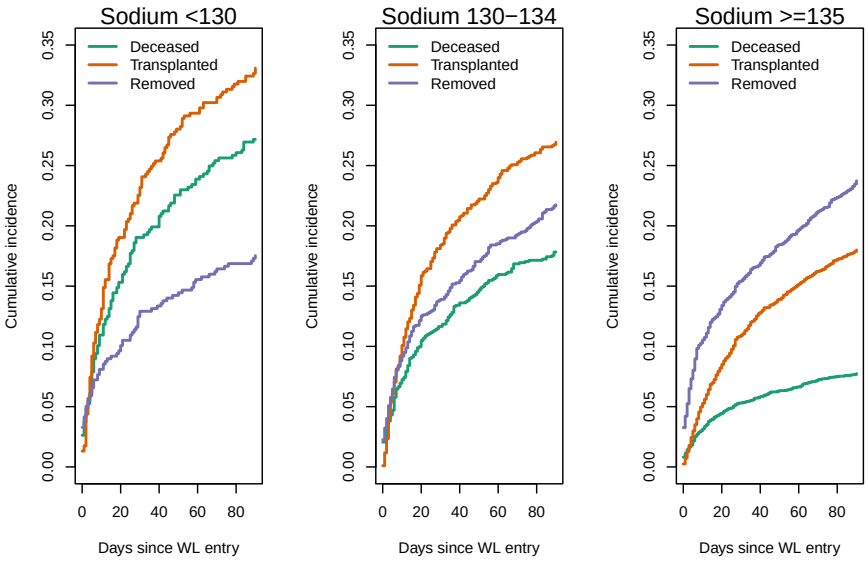


Figure 2.3: Cumulative incidence plots for 90-day WL outcomes, with competing risks of death, transplantation and removal due to clinical condition or censoring for NSE or HU status during waiting. Hyponatriemic patients show increased rates of mortality (27%) and transplantation (33%) compared to normonatriemic patients (respectively 8% and 18%) patients. For an explanation of the NSE and HU status, see Jochmans et al.

(95%CI 2.26-3.28), respectively.

MELD-Na performance

Per MELD point increase, the 90-day mortality risk increased by 17% (HR 1.17; 95%CI 1.16 – 1.18; p<0.001), c-index 0.832 (SE 0.008). The GAM with splines of the MELD-corrected effect of Na level on 90-day mortality showed approximate linearity in the 125-140 mmol/L range, see Figure 2.4. Within this interval, the risk of 90-day death increased by threefold (HR 2.9; 95%CI 2.30-3.53; p<0.001). In the MELD-Na model, each gained MELD and lowered Na point increased 90-day mortality risk by respectively 16% (HR 1.16; 95%CI 1.15 – 1.17; p<0.001), and 8% (HR 0.92; 95%CI 0.90 – 0.94; p<0.001), c-index 0.847 (SE 0.007). For each year of the study period, the c-index of MELD

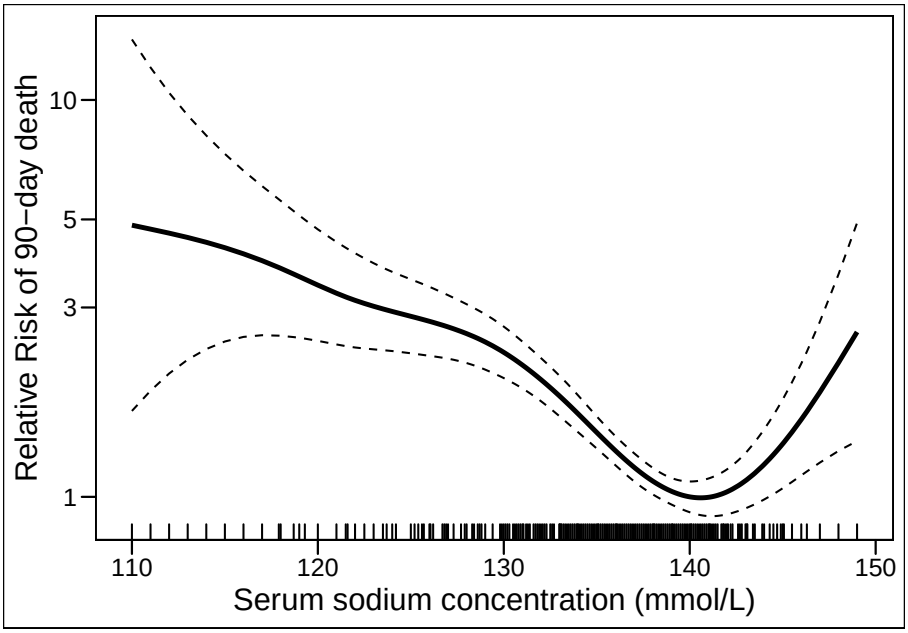


Figure 2.4: Generalized additive Cox model with spline showing the effect of serum sodium at listing on 90-day mortality, corrected for the MELD score.

and MELD-Na was plotted, see Figure 2.5. Between 2007-2018, the c-index of MELD and MELD-Na decreased significantly, respectively from 0.866 to 0.810 and 0.946 to 0.828 (Table 2.2). In this period, the MELD, age and distribution of liver disease changed significantly (supplement 4). Alcohol-induced liver disease, HCC, primary biliary cirrhosis (PBC) and non-alcoholic steatohepatitis (NASH) cirrhosis increased and primary sclerosing cholangitis (PSC), hepatitis-C (HCV), hepatitis-B (HBV) and other causes decreased.

The MELD-Na calibration plot showed a well calibrated model for 90% of the predicted risks in the population, with an overestimation for the highest 10% (504 patients) predicted risks (Figure 2.6 and supplement 6). The prediction error of 90-day death was lower for MELD-Na than for MELD, with Brier scores of respectively 0.059 (34% prediction error reduction), and 0.061 (32% reduction).

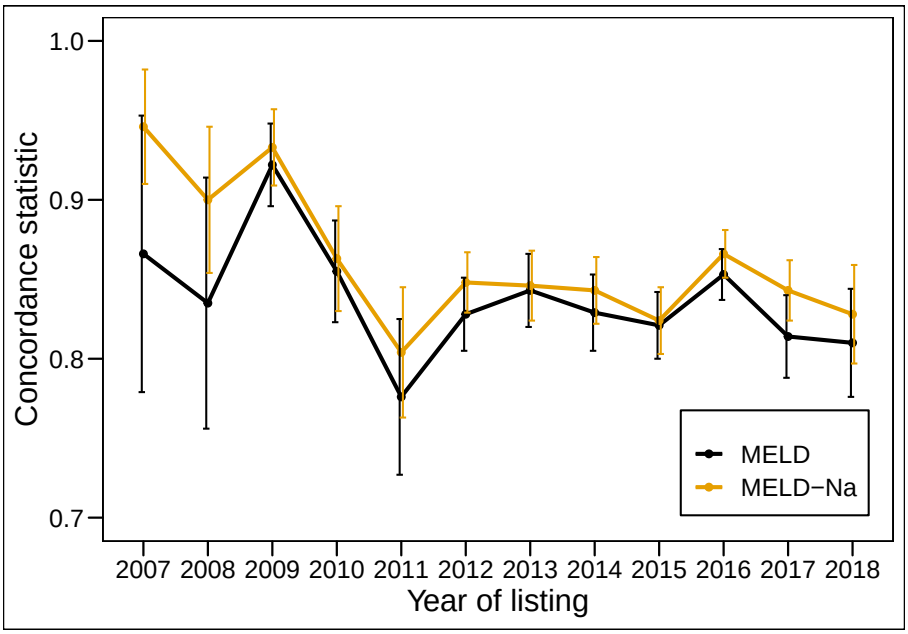


Figure 2.5: The concordance statistics (c-indices) for 90-day mortality of MELD and MELD-Na between 2007 and 2018.

Table 2.2: The 90-day mortality concordance statistics of MELD and MELD-Na

Year	MELD	SE	MELD-Na	SE
2007	0.866	0.087	0.946	0.036
2008	0.835	0.079	0.900	0.046
2009	0.922	0.026	0.933	0.024
2010	0.855	0.032	0.863	0.033
2011	0.776	0.049	0.804	0.041
2012	0.828	0.023	0.848	0.019
2013	0.843	0.023	0.846	0.022
2014	0.829	0.024	0.843	0.021
2015	0.821	0.021	0.824	0.021
2016	0.853	0.016	0.866	0.015
2017	0.814	0.026	0.843	0.019
2018	0.810	0.034	0.828	0.031

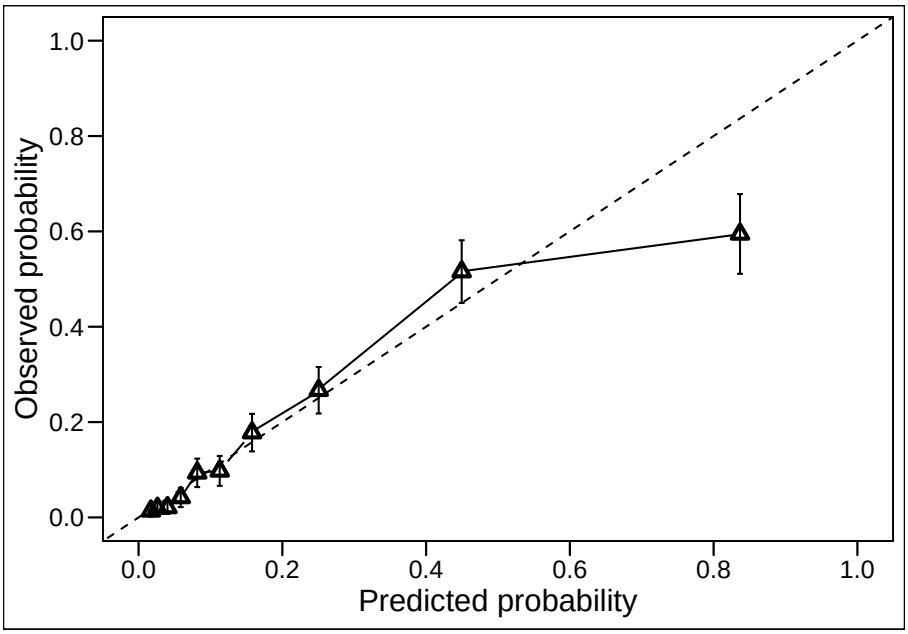
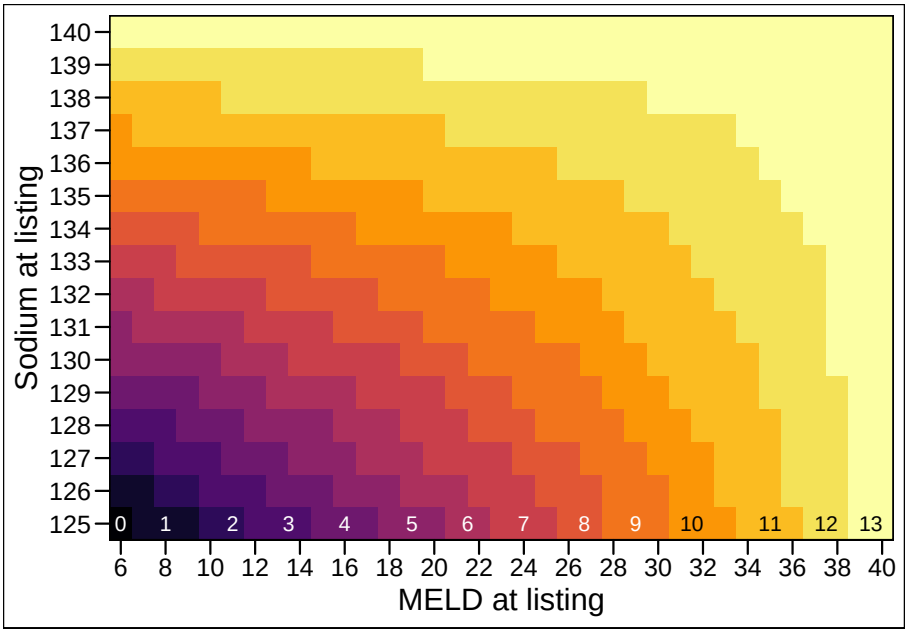


Figure 2.6: Calibration plot of the MELD-Na model showing the predicted and observed risks of death per decile (10%) of the patient population. The diagonal line represents a perfect calibration.

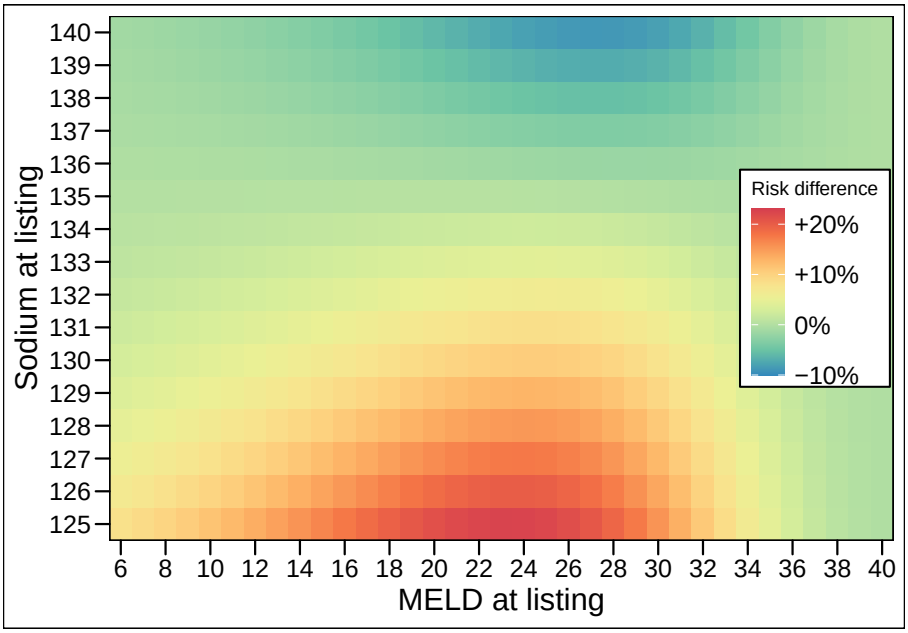
Impact on the waiting list

On the WL, implementation of the MELD-Na score would lead to competition for transplantation between hyponatremic and high-MELD patients. The constructed heatmap of risk differences showed that compared to MELD, approximately 20% of the patients gained significant predicted 90-day mortality risks according to MELD-Na (red area). The largest increase (+22.5%) was found for MELD 23 Na 125 patients. Approximately 19% of the patients had significantly lower predicted risks with MELD-Na compared to MELD (blue area), of which the largest decrease (-8.72%) was estimated for MELD 27 Na 140 patients, see Figure 2.7. Thus, the patients in the red area (19%) are prioritized most by MELD-Na. On the other hand, the lowest 20% of predicted risks (blue area) would have a reduced chance of transplantation compared to MELD allocation. The interactive heatmaps allow specific assessment of the gained risks and MELD-Na points for individual patients (online supplement https://plot.ly/~Liver_Research/3/ and https://plot.ly/~Liver_Research/5/). In total, 3384 (64.9%) patients gained an average of 1.94 MELD-Na points at listing. The highest reclassification rates, i.e. lowest percentage on the diagonal, were seen between MELD 12 to 30 (figure 8 and https://plot.ly/~Liver_Research/7/ and https://plot.ly/~Liver_Research/18/). On average, MELD 23 patients gained the most, i.e. an average of 2.73, MELD-Na points. From 19 points and above, the frequency of MELD-Na scores at listing was significantly higher than MELD scores, with the exception of MELD 40 (online supplement https://plot.ly/~Liver_Research/11/).

To make comparison to the UNOS data possible, we calculated the stratified MELD reclassification rates and estimated WL mortality reduction (supplement 3). Stratification of scores in accordance to Kim et al.[9] showed a reclassification rate of 26.3% (156 / 593) in the deceased patients. This led to an estimated 4.9% reduction in 90-day waiting list mortality. The analysis of disease-specific prioritization in the deceased patients showed that patients with HCC and hepatitis B had the highest chance of reclassification to a higher MELD-Na stratum, 36% and 30% respectively (supplement 3). However, patients with (post)alcoholic cirrhosis had the highest increase in mean MELD-Na compared to MELD. This illustrated that the strata chosen by Kim et al. could enable stage migration bias (supplement 3 and 5). Therefore, we believe that the total number of reclassified pa-



(a) Gained MELD-Na points for each combination of MELD and serum sodium level at listing



(b) Difference in predicted death probability between MELD and MELD-Na

Figure 2.7: Point and risk differences between MELD and MELD-Na.

tients and the distribution of the gained MELD-Na points are more useful information when estimating the possible impact of MELD-Na-based allocation (Figure 2.8 and https://plot.ly/~Liver_Research/7/ and https://plot.ly/~Liver_Research/18/).

Discussion

This cohort analysis validated the UNOS MELD-Na score for the Eurotransplant region and provided the first examination of the extent of hyponatremia among LT candidates in this region. It was shown that the mortality hazards for mild and severely hyponatremic patients continued to increase during waiting for LT. The precise relation between the sodium concentration at listing and the 90-day WL mortality was calculated. Our analysis showed that MELD-Na had better prognostic abilities than MELD for the prediction of 90-day WL mortality, even though both MELD and MELD-Na declined the past years. Therefore, the use of the MELD-Na score could improve the allocation of donor livers in the Eurotransplant region.

MELD-Na prediction performance

Accounting for serum sodium is relevant for the Eurotransplant population, as the prevalence of hyponatremia was similar,[9,22] or even higher compared to another large study.[28] The severity of hyponatremia was associated with a continuous increase in the risk of death on the WL, as shown by the cumulative incidence plots and diverging survival curves (Figure 2.3 and supplement 2). Compared to MELD, MELD-Na showed better discrimination between death and survival at 90-days, with a c-index of respectively 0.832 and 0.847. The c-index of MELD-Na was higher than found by some[29] and comparable to that found by other investigators.[9,30] Although the improvement in c-index by using MELD-Na was modest, it represented an important improvement in mortality prediction by considering hyponatremia as an independent risk factor of 90-day mortality. As the sickest candidates on the waiting list are prioritized, the increased discrimination would improve allocation.

Although MELD-Na performed better than MELD, both models showed significantly declining c-indices between 2007-2018 (Figure 2.5, Table

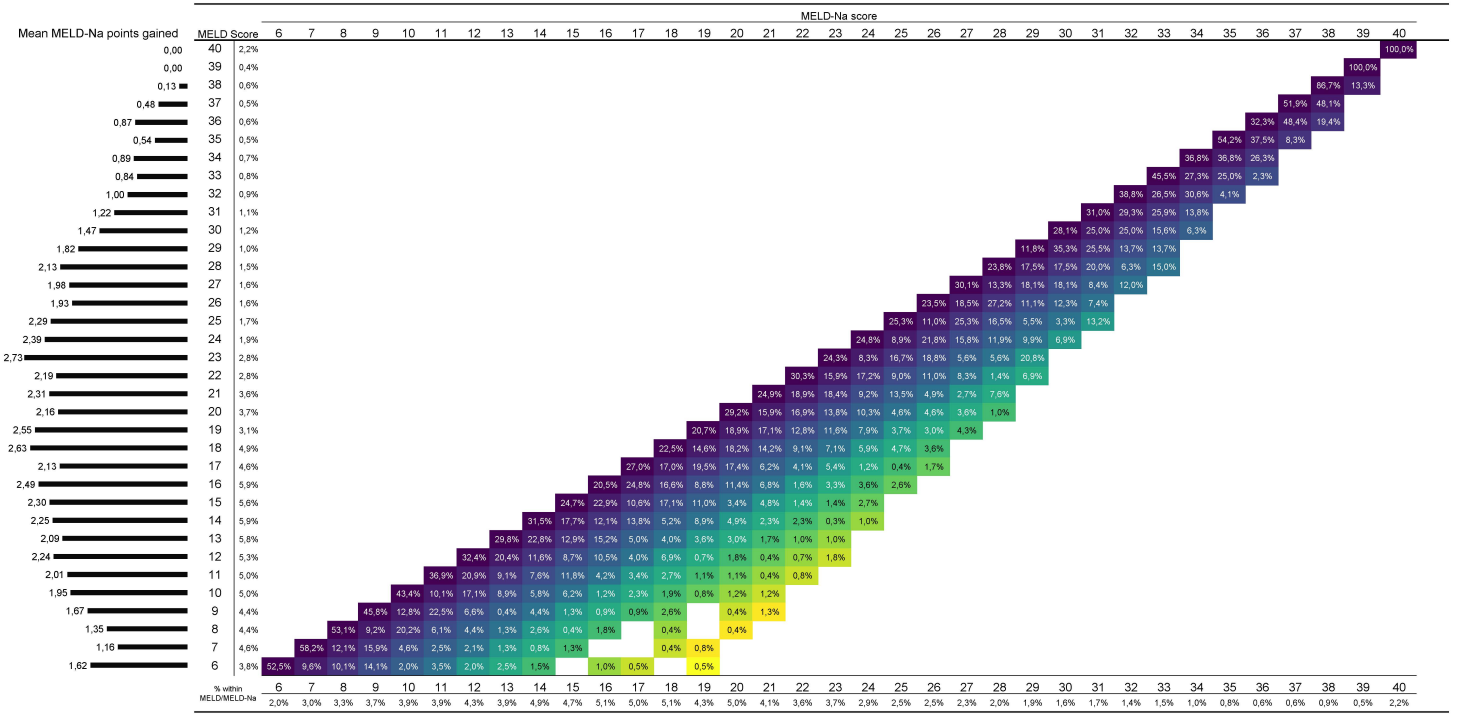


Figure 2.8: Reclassification from MELD (y-axis with percentage of patients with that score) to MELD-Na (x-axis with percentage of patients with that score). The diagonal shows which patients remain in the same stratum, that is, not reclassified, and which patients are reclassified to a higher MELD-Na score (percentages in the tiles). A lighter color indicates a higher difference between MELD and MELD-Na scores. The histogram on the left shows for each MELD score the average gain in MELD-Na points.

2.2). It is possible that the exceptionally high MELD-Na c-indices in the years 2007-2009 were due to population sampling, which would also make the decrease in c-index over the years seem excessive. In this period, average age and MELD at listing increased significantly. Most importantly, the distribution of causes of liver disease significantly changed (supplement 4). Compared to the US, the Eurotransplant population comprised more patients with ALD and HCC and less with HCV and NASH.[24] Godfrey et al. first showed declining c-indices over the years for MELD and MELD-Na, which they attributed to the decrease in HCV and increase of NASH and ALD. Despite the different distribution of causes of liver disease compared to the US, a similar change over time was seen. This could explain the initially higher but similarly declining c-indices of MELD and MELD-Na. Policy makers should consider this decline when evaluating a possible shift from MELD to MELD-Na. Still, MELD-Na would be a significant improvement because of the increasing prevalence of hyponatremia, its effect on 90-day mortality and the significantly higher c-indices of MELD-Na.

The MELD-Na showed good calibration, with overestimation of risks only in the top 10% of the patients. Both MELD and MELD-Na overestimated the highest predicted risks (supplement 6), as also shown by others.[9] However, MELD-Na showed a higher reduction in the prediction error of 90-day death compared to MELD, as calculated with Brier scores. Thus, MELD-Na was a more accurate predictor of 90-day WL death than MELD alone.

Effect of MELD-Na

Since we validated the UNOS MELD-Na score, we used the Na 125-140 mmol/L interval to fit our model. In this interval we showed a 1.5 higher increase in 90-day mortality risk per Na unit as compared to the UNOS regions.[9] Therefore, a greater reduction in WL mortality could be achieved through MELD-Na-based allocation. In the US, introduction of MELD-Na-based allocation reduced (HR 0.738) 90-day waiting list mortality for almost all MELD scores.[22] However, the number of transplants was higher in the studied MELD-Na period, which also could have reduced WL mortality. Still, Nagai et al. showed that the intended recognition of hyponatremia was achieved, as the WL mortality hazards of mild and severe hyponatremia decreased with respectively 27.9% and 48.3%.[22] In the US, it was shown

that in MELD<12 patients hyponatremia was not associated with LT survival benefit.[20] Thus, UNOS MELD-Na is only used to allocate liver grafts in MELD>11 patients. In our population, very few (2.8%) MELD<12 patients had severe hyponatremia. Although these patients would gain transplant chances through MELD-Na allocation, others would be prioritized more often. Our data also showed that the frequency of MELD-Na >18 scores increased significantly (https://plot.ly/~Liver_Research/11/). This would reduce transplant chances for patients listed with exception points, e.g. HCC patients, as these patients initially receive 20 points at listing in Eurotransplant.[6] Although the reduced advantage of (N)SE points is warranted according to some,[31,32] (N)SE point policy did not change after MELD-Na implementation in the UNOS regions (personal SRTR communication). Still, many patients are listed with exception points, both in Eurotransplant and in the US. Therefore, the distribution of gained MELD-Na points, survival benefit and influence on exception points of Eurotransplant LT candidates should be considered before implementation of MELD-Na-based allocation. A simulation of MELD-Na-based allocation would give the most accurate estimates of the effect on WL mortality.

Limitations

This study has several limitations. First, only one measurement, i.e. at first listing, of the MELD and sodium was used to study the effect on 90-day mortality. Since the disease state of the patient is a dynamic process, a time-dependent analysis with more datapoints might have been a better representation of the true risk posed by hyponatremia. Indeed, we showed that the effect of hyponatremia increased with time (Figure 2.3 and supplement 3). Also, serum sodium levels in the MELD-Na model were bound between 125-140 mmol/L. The fitted Cox model between these borders had a excellent c-index, but the relationship between serum sodium level and mortality was slightly different for the Eurotransplant region compared to the UNOS regions.[9] However, the goal was to validate the MELD-Na as used in the UNOS regions for the Eurotransplant region, and this goal was achieved. Still, refitting of the MELD parameters for the Eurotransplant population could be a valuable, especially regarding the decline in c-index between 2007-2018. Second, sodium data at first listing was missing for many eligible patients (supplement 1). This could have caused selection bias, pos-

sibly making the results less generalizable. However, analysis of the differences between the patients with and without registered sodium at listing showed that there was no reason to suspect selection bias. In the missing Na group, a significantly higher prevalence of alcoholic cirrhosis and virus-induced hepatitis was seen (supplement 4). Also, patients in the group with missing Na had a significantly higher serum creatinine. Thus, the prevalence of hyponatremia in those eligible patients could very well be even higher than found in the current cohort. Moreover, even though some data was missing, the number of patients included in this study sufficed to evaluate and estimate the improvements of MELD-Na with great statistical precision. Thus, the results of this study should be an incentive for the mandatory collection of sodium values across the Eurotransplant region.

Conclusion

In conclusion, this study showed that the MELD-Na gave better 90-day mortality prediction than MELD for LT candidates on the Eurotransplant waiting list. As stated before, "the MELD-based allocation system will and also must evolve."[26] The recognition of the independent prognostic impact of hyponatremia should lead to a more effective allocation. Thus, in the Eurotransplant region the MELD should be replaced by the MELD-Na as the basis allocation of donor livers.

References

1. Eurotransplant. Annual Report 2017.; 2018. www.eurotransplant.org.
2. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384(9945):766-781. doi: 10.1016/S0140-6736(14)60460-8
3. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. J Hepatol. 2019;70(1):151-171. doi:10.1016/j.jhep.2018.09.014
4. Pimpin L, Cortez-Pinto H, Negro F, et al. Burden of liver disease in Europe: Epidemiology and analysis of risk factors to identify prevention policies. J Hepatol. 2018;69(3):718-735.
5. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123-133. doi:10.1002/hep.29466
6. Jochmans I, Van Rosmalen M, Pirenne J, Samuel U. Adult Liver Allocation in Eurotransplant. Transplantation. 2017;101(7):1542-1550. doi:10.1097/TP.0000000000001631
7. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology. 2000;31(4):864-871. doi:10.1053/he.2000.5852
8. Biggins SW, Rodriguez HJ, Bacchetti P, Bass NM, Roberts JP, Terrault NA. Serum sodium predicts mortality in patients listed for liver transplantation. Hepatology. 2005;41(1):32-39. doi:10.1002/hep.20517
9. 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(10):1018-1026. doi:10.1007/s11250-017-1262-3
10. Ruf AE, Kremers WK, Chavez LL, Descalzi VI, Podesta LG, Villamil FG. Addition of serum sodium into the MELD score predicts waiting list mortality better than MELD alone. Liver Transplant. 2005;11(3):336-343. doi:10.1002/lt.20329
11. 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(4):802-810. doi:10.1002/hep.1840400409
12. Biggins SW, Kim WR, Terrault NA, et al. Evidence-Based Incorporation of Serum Sodium Concentration Into MELD. Gastroenterology. 2006;130(6):1652-1660. doi:10.1053/j.gastro.2006.02.010
13. Ginès P, Guevara M. Hyponatremia in cirrhosis: Pathogenesis, clinical significance, and management. Hepatology. 2008;48(3):1002-1010. doi:10.1002/hep.22418
14. John S, Thuluvath PJ. Hyponatremia in cirrhosis: Pathophysiology and management. World J Gastroenterol. 2015;21(11):3197-3205. doi:10.3748/wjg.v21.i11.3197
15. Wong F, Blei AT, Blendis LM, Thuluvath PJ. A vasopressin receptor antagonist (VPA-985) improves serum sodium concentration in patients with hyponatremia: A multicenter, randomized, placebo-controlled trial. Hepatology. 2003;37(1):182-191. doi:10.1053/jhep.2003.50021
16. Wong F, Sniderman K, Liu P, Allidina Y, Sherman M, Blendis L. Transjugular intrahepatic portosystemic stent shunt: Effects on hemodynamics and sodium homeostasis in cirrhosis and refractory ascites. Ann Intern Med. 1995;122(11):816-822. doi:10.7326/0003-4819-122-11-199506010-00002
17. Londoño MC, Guevara M, Rimola A, et al. Hyponatremia Impairs Early Posttransplantation Outcome in Patients With Cirrhosis Undergoing Liver Transplantation. Gastroenterology. 2006;130(4):1135-1143. doi:10.1053/j.gastro.2006.02.017
18. Dawwas MF, Lewsey JD, Neuberger JM, Gimson AE. The Impact of Serum Sodium Concentration on Mortality After Liver Transplantation: A Cohort Multicenter Study. Liver Transplant. 2007;13(5):767-768. doi:10.1002/lt
19. Leise MD, Yun BC, Larson JJ, et al. Effect of the Pretransplant Serum Sodium Concentration on Outcomes Following Liver Transplantation. Liver Transplant. 2014;14(20):687-697. doi:10.1002/lt
20. Sharma P, Schaubel DE, Goodrich NP, Merion RM. Serum Sodium and Survival Benefit of Liver Transplantation. Liver Transplant. 2015;21:308-313. doi:10.1002/lt.
21. Allocation of livers and liver- intestines. Organ Procurement and Transplantation Network Liver and Intestine Committee.
22. Nagai S, Chau LC, Schilke RE, et al. Effects of Allocating Livers for Transplantation Based on Model for End-Stage Liver

Disease-Sodium Scores on Patient Outcomes. Gastroenterology. 2018;155(October):1451-1482. doi:10.1053/j.gastro.2018.07.025

23. 2017 Annual Data Report. Scientific Registry of Transplant Recipients. http://srtr.transplant.hrsa.gov/annual_reports/Default.aspx. Published 2017. Accessed September 5, 2019.
24. Godfrey EL, Malik TH, Lai JC, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. Am J Transplant. 2019;19(12):3299-3307. doi:10.1111/ajt.15559
25. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73. doi:10.7326/M14-0698
26. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology. 2003;124(1):91-96. doi:10.1053/gast.2003.50016
27. Sievert C, Parmer C, Hocking T, et al. Create Interactive Web Graphics via "plotly.js." 2019. https://plot.ly/r.
28. Angeli P, Wong F, Watson H, et al. Hyponatremia in cirrhosis: Results of a patient population survey. Hepatology. 2006;44(6):1535-1542. doi:10.1002/hep.21412
29. Biselli M, Gitto S, Gramenzi A, et al. Six Score Systems to Evaluate Candidates with Advanced Cirrhosis for Orthotopic Liver Transplant: Which Is the Winner? Liver Transplant. 2010;16:964-973. doi:10.1002/lt.
30. Cejas N, Villamil F, Lendoire J, et al. Improved Waiting-List Outcomes in Argentina After the Adoption of a Model for End-Stage Liver Disease–Based Liver Allocation Policy. Liver Transplant. 2013;13(19):711-720. doi:10.1002/lt
31. Northup PG, Intagliata NM, Shah NL, Pelletier SJ, Berg CL, Argo CK. Excess mortality on the liver transplant waiting list: Unintended policy consequences and model for End-Stage Liver Disease (MELD) inflation. Hepatology. 2015;61(1):285-291. doi:10.1002/hep.27283
32. Umgelter A, Hapfelmeier A, Kopp W, van Rosmalen M, Rogiers X, Guba M. Disparities in Eurotransplant liver transplantation wait-list outcome between patients with and without model for end-stage liver disease exceptions. Liver Transplant. 2017;23(10):1256-1265. doi:10.1002/lt.24805

Chapter 3

Refitting the Model for End-stage Liver Disease for the Eurotransplant region

Goudsmit BFJ, Putter H, Tushuizen ME, et al. Refitting the Model for End-Stage Liver Disease for the Eurotransplant Region. *Hepatology*. 2021;74(1):351-363. doi: 10.1002/hep.31677.

Abstract

Background & Aims: The United Network for Organ Sharing's Model for End- Stage Liver Disease (UNOS-MELD) score is the basis of liver allocation in the Eurotransplant region. It was constructed 20 years ago in a small US cohort and has remained unchanged ever since. The best boundaries and coefficients were never calculated for any region outside the United States. Therefore, this study refits the MELD (reMELD) for the Eurotransplant region.

Methods: All adult patients listed for a first LT between 01.01.2007-31.12.2018 were included. Data was randomly split in a training (70%) and validation (30%) set. In the training data, generalized additive models (GAMs) with splines were plotted for each MELD parameter. The lower and upper bound combinations with the maximum log-likelihood were chosen for the final models. The refit models were tested in the validation data with c-indices and Brier scores. Through likelihood ratio tests the refit models were compared to UNOS-MELD. The correlation between scores and survival of prioritized patients was calculated.

Results: A total of 6,684 patients were included. Based on training data, refit parameters were capped at creatinine 0.7-2.5 (mg/dL), bilirubin 0.3-27 (mg/dL), INR 0.1-2.6 and sodium 120-139 (mmol/L). ReMELD and reMELD-Na showed c-indices of 0.866 and 0.869 respectively. ReMELD-Na prioritized patients with 1.6 times higher 90-day mortality probabilities as compared to UNOS-MELD.

Conclusion: Refitting MELD resulted in new lower and upper bounds for each parameter. The predictive power of reMELD-Na was significantly higher than UNOS-MELD. Refit MELD prioritized patients with higher 90-day mortality rates. Thus, reMELD(-Na) should replace UNOS-MELD for liver graft allocation in the ET region.

Introduction

The number of patients in need of a liver transplantation (LT) in the Eurotransplant region exceeds the available donor grafts.[1] Therefore, patients with end-stage liver disease are placed on a waiting list (WL) which prioritizes the patients with the most severe liver disease, i.e. most in need of transplantation. The Model of End-stage Liver Disease (MELD) estimates disease severity in LT candidates, based on three parameters: serum creatinine, bilirubin and the international normalized ratio (INR) for prothrombin time.[2] The MELD was weighed, i.e. the relative importance of each parameter, based on a cohort from 1991-1995.[3] For clinical use, the lower boundaries for the parameters were set to one, to prevent negative MELD scores after natural logarithm (ln) transformation. Creatinine levels were maximized to four for regular patients and set to four for patients who received dialysis. According to some of the proposers of MELD, these boundaries were "based entirely on the clinical intuition of the policy-making body when the MELD score was implemented."[4] Others also noted that "arbitrary changes not based on mortality risk evidence were incorporated into the form of MELD" and that these lower and upper limits were "set without any particular objective rationale."[5]

On another continent and almost 20 years later, the original UNOS-MELD equation is still being used for the allocation of liver grafts in the Eurotransplant region and elsewhere. Due to changing population characteristics, the predictive power of UNOS MELD has declined significantly in the last years.[6] However, an update of the MELD coefficients in UNOS data showed that performance could still be further improved.[4] As the Eurotransplant population differs from the original MELD cohort,[3,7] improvement of the Eurotransplant liver allocation is very well possible by refitting MELD to the Eurotransplant population. Refitting is the reweighing of predictors and establishment of lower and upper bounds of each parameter, based on the best fit to the current data. It was hypothesized that the UNOS-MELD is not optimally fit for the Eurotransplant patients, as it was fit on the UNOS population. This could diminish MELDs predictive power and discrimination ability between survival and death. It is the optimization of this discrimination that gives the most effective sickest-first allocation.

Therefore, this study constructs a refit MELD score for the Eurotrans-

plant region, by reweighing the MELD coefficients and re-evaluating the boundaries for the three parameters based on recent Eurotransplant data. The refitting methods presented here could be used to improve prediction models for any region. Also, the added value of the serum sodium (Na) levels at listing in an Eurotransplant refit MELD-Na score will be evaluated. The performance of the constructed refit Eurotransplant models will be compared to the currently used UNOS-MELD.

Methods

Patient data

The TRIPOD statement was used to report the development of the multivariate prediction models in this study.[8] Data was requested from the Eurotransplant Database. All adult patients actively listed for a first liver transplantation between January 1ˆst, 2007 - December 31ˆst, 2018 were included. The starting point of inclusion was chosen after the start of MELD-based allocation in 2006. Patients were excluded if they received (non)standard exception points (NSE), a high urgency (HU) status (i.e. UNOS status 1), living donor grafts or multi-organ transplantations (other than kidney).[9] Patient data was collected from the date of active listing until delisting or the end of 90-day follow-up. Reasons for delisting were death, transplantation, removal because of clinical condition or other reasons. The primary outcome was death within 90 days of first active listing. The predictors used for the multivariate models were both the bound and continuous levels of serum creatinine, bilirubin, INR and sodium at first active listing. For the survival analysis, patients were censored at transplantation, removal from the list, end of follow-up at 31.12.2018 or after receiving NSE points or a HU status during active waiting. The sample size for this study was set by the retrospective design. Missing data (in $<0.01\%$) was not imputed.

Statistical methods

The data was randomly split into a training (70%) and validation (30%) set. For each recipient, the UNOS-MELD and MELD-Na score at first active listing were calculated.[10,11] Then, the ET refit MELD (reMELD)

score was constructed in the training data. For each MELD parameter, a multivariate generalized additive Cox model (GAM) with smoothing splines was plotted. The GAM showed the (non-)linear effect of the specific parameter on 90-day mortality, corrected for the other uncapped MELD parameters. By visual inspection it was assessed whether upper and lower boundaries for the parameter were necessary, i.e. if there was any violation of the linearity relation between studied parameter and the 90-day mortality and at which time point. Then, the best boundaries for the parameter were sought within the visually apparent range by calculating the maximum log-likelihood and the concordance statistic (c-index) for each possible combination of upper and lower bounds. The combination with the maximum log-likelihood was chosen as the lower and upper bound for that MELD parameter. The impact of deviations from the maximum log-likelihood and c-index were visualized through heatmaps to facilitate discussion of weighing the maximum calculated values against clinically relevant cut-offs. After establishing the best boundaries for the parameter, a multivariate Cox model with the capped parameter was compared to a Cox model with the unbounded values through likelihood ratio tests. To visualize the fit of the studied reMELD parameter, the obtained bounds and coefficient were plotted in the training data. The abovementioned steps were repeated for all three MELD parameters. The three obtained capped parameters were then combined into a multivariate Cox model, thus forming the Eurotransplant refit MELD. To ensure equal distributions of the traditional UNOS-MELD and ET refit MELD scores in our data, the 25th and 75th quantiles were matched. Also, reMELD scores below 6 and above 40 were set to that value. Then, the addition of serum sodium to the reMELD was investigated in the training set as described above for the MELD parameters. In short: based on the GAM inspection, the optimal Na bounds were sought, i.e. calculating log-likelihood values and c-indices, and compared with likelihood ratio tests to uncapped Na levels. Interactions between Na and each refit MELD parameter were assessed and deemed relevant if $p<0.01$. Thus, the final reMELD-Na model comprised of reMELD parameters, newly bound sodium and relevant interactions between the terms. Again, the 25th and 75th quantiles were matched and the final scores of the refit MELD-Na were set between 6 to 40. Finally, the refit ET models were compared with likelihood ratio tests to UNOS-MELD. For each model, the c-index was calculated to calculate discriminative ability in the validation data. Brier scores were

calculated as a measure of error reduction in prediction estimates.[12] The fit of the models to the validation data was visualized by plotting the coefficients for each MELD parameter. The correlation between the currently used UNOS-MELD and constructed reMELD-Na was investigated by plotting both scores. To assess whether reMELD-Na would give more effective sickest-first allocation, survival estimates were calculated for patients prioritized by UNOS-MELD and reMELD-Na. All statistical analyses were performed using R v3.6.1(R Foundation for Statistical Computing, Vienna, Austria).

Results

In this study, 6,944 patients were included, see Table 3.1. More male (68%) than female patients were included, and alcohol induced cirrhosis was the most frequent cause of liver disease. The median UNOS-MELD and serum sodium at listing were 14 (IQR 10-20) and 138 (IQR 134-140) respectively. After 90 days of follow-up, 35.7% of the patients were still waiting for LT, 23.8% were censored due to HU status or (N)SE points, 18.0% were transplanted, 12.6% were removed from the WL and 9.8% died on the WL. There were no relevant differences between the training and validation data.

Model development

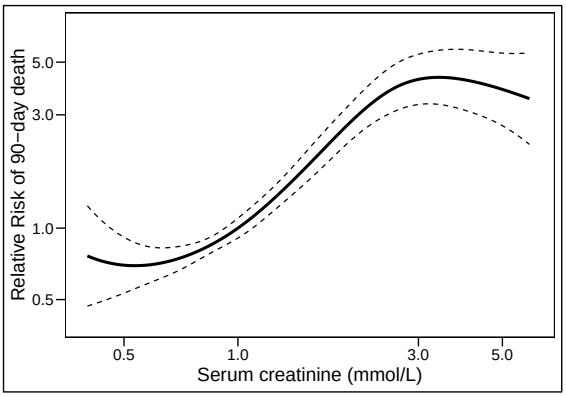
The GAM plots for each parameter are shown below. For creatinine, the S-shaped curve displayed clear lower and upper boundaries in Figure 3.1A, the maximum log-likelihood was calculated for the bounds of 0.7 and 2.5 mg/dL. Clinically, it seemed logical to include values of creatinine below 1.0 mg/dL, mainly because many patients (55%) had creatinine levels <=1 mg/dL. Through refitting, the serum creatinine was decreased in weight and its upper bound was lowered. Therefore, the influence of renal failure on the chances for LT was reduced.

For bilirubin, in Figure 3.1B, the lower bound was found at 0.3 and the upper at 27 mg/dL. Varying of the lower bound between 0.1 and 0.5 did not alter the log-likelihood significantly, i.e. would still be an acceptable fit to the data. Also, 23.7% of our population would no longer be capped at listing. The upper bound of 27 mg/dL could be altered to a clinically

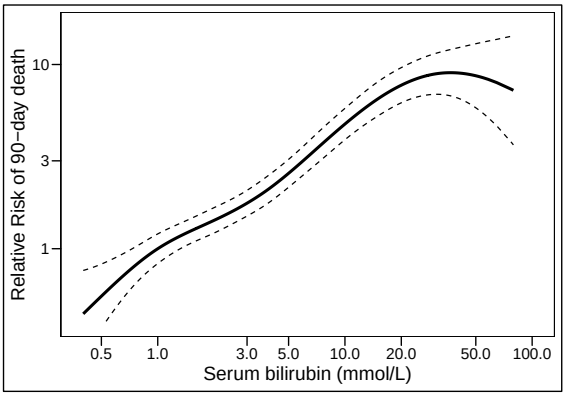
Table 3.1: Characteristics of training and validation data

characteristics	Training set	Validation set	p
n	4860	2084	
Age (median (IQR))	56 (49-62)	55 (49-62)	0.022
Gender female (%)	1563 (32.2)	659 (31.6)	0.680
Disease (%)			
Cirrhosis, Alcoholic	1361 (28.0)	600 (28.8)	
Cirrhosis, HCV	352 (7.2)	123 (5.9)	
Cirrhosis, other causes	825 (17.0)	353 (16.9)	
Cholestatic disease	652 (13.4)	295 (14.1)	
HCC and cirrhosis	953 (19.6)	421 (20.2)	
Other	717 (14.8)	292 (14.0)	
Status after 90 days			
Censored because of HU or NSE	1171 (24.2)	476 (22.9)	
Deceased	452 (9.30)	226 (10.8)	
Removed from the waiting list	624 (12.8)	257 (12.3)	
Still waiting on waiting list	1734 (35.8)	739 (35.5)	
Transplanted	867 (17.9)	381 (18.3)	
Days follow-up (mean (SD))	44.22 (39.48)	44.06 (39.27)	0.875
Serum measurement at listing (mean (SD))			
Creatinine in mg/dL	1.40 (3.73)	1.46 (4.16)	0.563
Bilirubin in mg/dL	5.74 (8.79)	5.84 (9.34)	0.669
INR	1.51 (0.72)	1.52 (0.72)	0.510
Sodium in mmol/L	137.02 (4.99)	136.94 (4.88)	0.526
UNOS MELD at listing (median (IQR))	14 (10-20)	14 (10-20)	

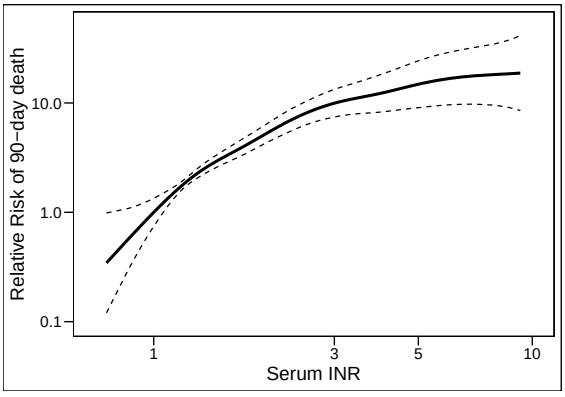
Note:
IQR: inter quartile range, HCV: hepatitis C induced cirrhosis, HCC: hepatocellular carcinoma, HU: high urgency, NSE: (non)standard exception, SD: standard deviation, INR: international normalized ratio, UNOS: united network for organ sharing



(a) Creatinine



(b) Bilirubin



(c) INR

Figure 3.1: For each parameter, the relation to 90-day mortality is shown based on the training data

more relevant value, roughly between 20 and 40, without affecting the optimal fit to the data too much (supplement heatmap bilirubin).

The INR had no lower bound and was capped at a maximum of 2.6, see Figure 3.1C. However, assessment of the log-likelihoods values showed that a range between 0.1 and 1.0 would be acceptable as lower bound (supplement heatmap INR) and would affect few patients (2.7%). For the INR an upper bound of 2.6 was chosen, which still acknowledged, i.e. did not cap, 93% of the patients. Although it may seem controversial to cap the INR, this meant that if patients reached 2.6, they would receive the maximum refit points for INR, of which the weight was increased in the refit models.

Overall, the reMELD and reMELD-Na models capped less patients at assumed values than UNOS-MELD. In Figure 3.2, lines were plotted for respectively creatinine, bilirubin, and the INR to represent the refit coefficient (slope of the diagonal) and the boundaries (horizontal lines).

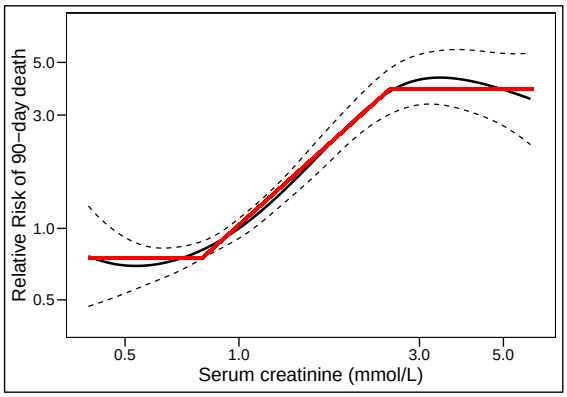
The heatmaps of the calculated log-likelihoods and c-indices per combination of boundaries are attached in the (online) supplement. After checking for interactions and matching the 25th and 75th quantiles of the reMELD to the UNOS-MELD in the training data, the reMELD equation was:

$$7.728 * ln(creatinine) + 3.446 * ln(bilirubin) + 10.597 * ln(INR) + 8.422$$

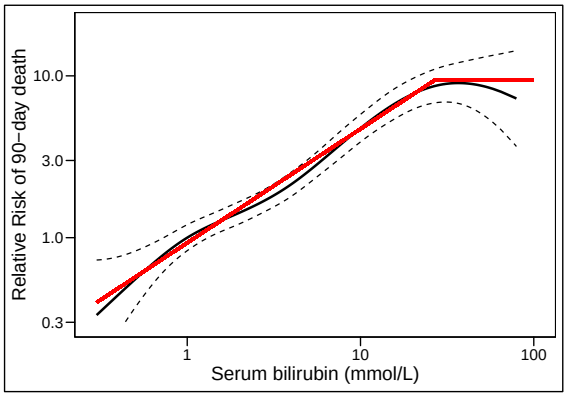
In this equation the abovementioned boundaries were used for the parameters. The maximum log-likelihood for Na levels was found between 120 and 139 mmol/L. Combining the reMELD and Na showed a significant interaction between Na and creatinine. Thus, after quantile matching in the training data, the reMELD-Na formula was:

$$\begin{aligned} 9.025 \times ln(creatinine) + 2.969 \times ln(bilirubin) + 9.518 \times ln(INR) - \\ 0.392 \times (139 - Na) - 0.351 \times ln(139 - Na) \times ln(creatinine) \end{aligned}$$

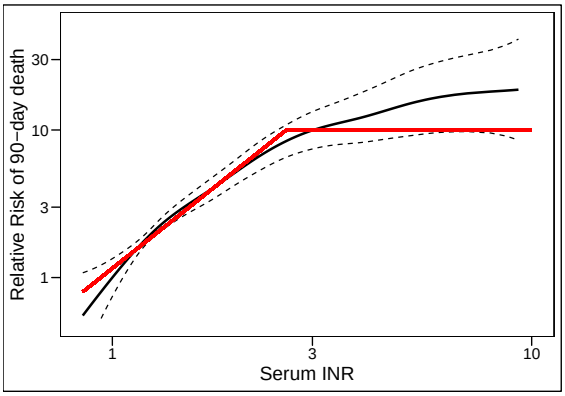
For the parameters in the reMELD-Na score, the abovementioned boundaries were used. Compared to the UNOS-MELD, re-MELD and reMELD-Na used respectively 149% (n=4815) and 42% (n=2748) more patient measurements, i.e. less true patient measurements were capped, at listing with the boundaries as shown in Table 3.2.



(a) Creatinine



(b) Bilirubin



(c) INR

Figure 3.2: For each parameter, the diagonal line represent the coefficient (slope of the diagonal) and lower and upper boundaries (horizontal segments) in refit MELD

Table 3.2: Parameter bounds and number of patient measurements included in UNOS and refit models

		UNOS MELD(-Na)			refit MELD(-Na)		
		bounds	capped (%)	included (%)	bounds	capped (%)	included (%)
Creatinine	lower	1	55.0	41.9	0.7	20.1	73
	upper	4	3.1		2.5	6.9	
Bilirubin	lower	1	23.7	76.3	0.3	2.0	93.5
	upper	NA			26.9	4.5	
INR	lower	1	9.8	91.2	0.1	NA	94.8
	upper	NA			2.6	5.2	
Sodium	lower	125	2.7	72.9	120	0.7	56.3
	upper	140	24.4		138.6	43	

Note:
For each parameter the lower and upper bounds are shown. 'capped' shows the percentage of the cohort that either lies under or above the chosen bounds. 'included' shows the percentage of patients whose measurements are included in the model.

Table 3.3: Comparison of models in validation data

Model	C-index	Max log-likelihood	Chisq	p
UNOS MELD	0.849 (se = 0.012)	-1376.6		
UNOS MELD-Na	0.860 (se = 0.010)	-1362.8	27.660	< 2.2e-16
reMELD	0.866 (se = 0.011)	-1347.1	58.966	< 2.2e-16
reMELD-Na	0.869 (se = 0.010)	-1347.1	59.066	< 2.2e-16

Note:
For each model the C- index and maximum log- likelihood are calculated in the validation data. The likelihood ratio comparisons of the models to UNOS- MELD are shown by chi- squared and P values.

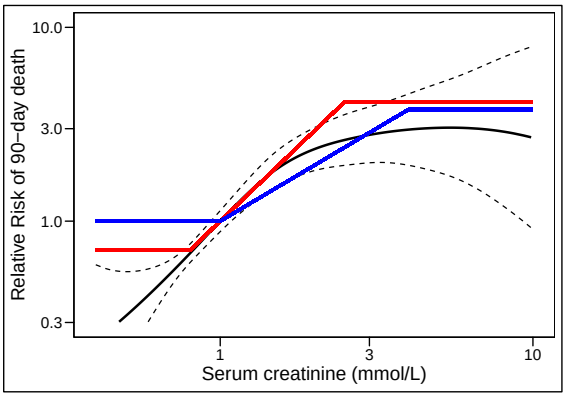
Model performance

Figure 3.3 shows the effect of each MELD parameter, corrected for the others, on 90-day mortality in the validation data. The red and blue lines represent the coefficients of the reMELD and UNOS-MELD respectively. It was visually apparent that refit MELD showed a better fit to the data for all three parameters.

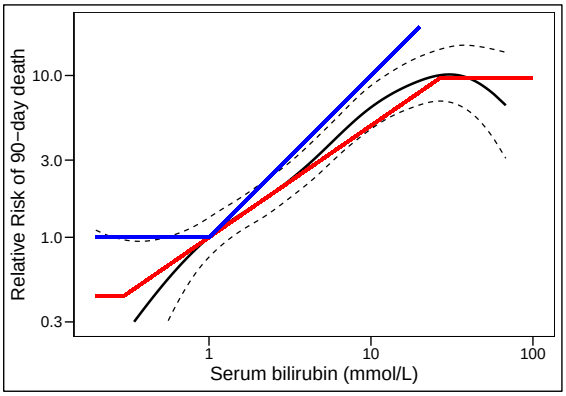
The calculated chi-square values confirmed significant ($p<0.001$) improvements in the refit models compared to the UNOS-MELD, shown in Table 3.3. The reMELD and reMELD-Na models showed c-indices of 0.866 and 0.869 respectively, which were significantly ($p<0.001$) higher than 0.849 of the UNOS-MELD, see Table 3.3. Furthermore, the reMELD-Na showed a 8% reduction in prediction error as compared to UNOS-MELD with Brier scores of 0.053 (reMELD-Na) and 0.057 (UNOS-MELD) respectively.

Impact on the waiting list

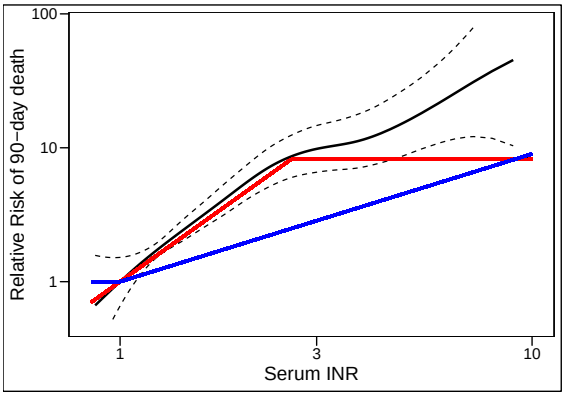
After 90 days of follow-up, 1,248 patients of our cohort were transplanted. By using the reMELD-Na compared to the UNOS-MELD to allocate the 1,248 available liver grafts, 134/1,248 (11.5%) of the transplanted patients would have been within the top 1,248 candidates under one of these models but not under the other; i.e., prioritization would differ. Table 3.4 shows the characteristics of these differently prioritized patients.



(a) Creatinine



(b) Bilirubin



(c) INR

Figure 3.3: In the validation data, the relation with 90-day mortality is shown. The coefficients and boundaries of creatinine in reMELD (red) and UNOS-MELD (blue) illustrate model fit.

Table 3.4: Characteristics of Prioritized Patients

Characteristics	Transplanted both	UNOS MELD Transplanted	reMELD-Na Transplanted	Not transplanted	p
n	1105	143	143	5553	
Age at listing (mean (SD))	53.42 (10.48)	48.73 (13.62)	55.29 (9.53)	54.09 (10.77)	<0.001
Gender female (%)	362 (32.8)	66 (46.2)	44 (30.8)	1750 (31.5)	0.003
Length (mean (SD))	172.87 (10.88)	171.73 (8.85)	173.59 (10.16)	173.03 (9.56)	0.368
Weight (mean (SD))	81.42 (18.43)	77.33 (18.19)	79.30 (18.30)	79.03 (17.41)	<0.001
Disease (%)					
Cirrhosis, Alcoholic	390 (35.3)	48 (33.6)	65 (45.5)	1458 (26.3)	
Cirrhosis, HCV	74 (6.7)	6 (4.2)	10 (7.0)	385 (6.9)	
Cirrhosis, other causes	285 (25.8)	27 (18.9)	33 (23.1)	833 (15.0)	
Cholestatic disease	113 (10.2)	15 (10.5)	7 (4.90)	811 (14.6)	
HCC and cirrhosis	37 (3.3)	3 (2.1)	9 (6.3)	1325 (23.9)	
Other	207 (18.7)	44 (30.7)	19 (13.2)	739 (13.3)	
Status after 90 days					
Censored because of HU or NSE	52 (4.7)	9 (6.3)	8 (5.6)	1578 (28.5)	
Deceased	338 (30.7)	28 (19.6)	36 (25.2)	276 (5.0)	
Removed from the list	121 (11.0)	30 (21.0)	27 (18.9)	703 (12.7)	
Still waiting on waiting list	56 (5.1)	19 (13.3)	28 (19.6)	2370 (42.8)	
Transplanted	536 (48.6)	57 (39.9)	44 (30.8)	611 (11.0)	
Days on the waiting list (mean (SD))	24.94 (78.46)	51.32 (114.64)	72.64 (132.97)	175.21 (304.96)	<0.001
Serum measurement at listing (mean (SD))					
Creatinine in mg/dL	2.95 (8.51)	2.67 (9.43)	1.26 (0.48)	1.09 (1.18)	<0.001
Bilirubin in mg/dL	19.29 (14.10)	10.69 (9.08)	8.01 (5.96)	2.89 (3.51)	<0.001
INR	2.43 (1.20)	2.37 (1.40)	1.74 (0.32)	1.30 (0.28)	<0.001
Sodium in mmol/L	134.26 (6.08)	138.21 (4.67)	127.34 (5.34)	137.76 (4.20)	<0.001
(refit)MELD score	30.95 (5.48)	25.57 (2.95)	21.10 (2.26)	12.91 (4.60)	<0.001
Dialysis dependent (%)	165 (15.3)	21 (15.1)	0 (0.0)	87 (1.6)	<0.001

Most notably, reMELD-Na-prioritized patients were slightly older, were more often male, and had a higher prevalence of cirrhosis. Unsurprisingly, these patients had significantly lower serum sodium levels (138 vs. 127 mmol/L). As hyponatremia is most often seen in alcohol-associated cirrhosis,[14] the sex and age differences are largely explained. The correlation plot Figure 3.4 shows which patients would be prioritized according to either UNOS-MELD or re-MELD-Na allocation.

The patients in the top left quadrant would have been prioritized by reMELD-Na allocation but not by UNOS-MELD. They had estimated 90-day survival probabilities of 52.4% (95CI 41.3 – 66.5), as compared to 70.0% (95CI 58.9 – 83.1) for patients prioritized by UNOS-MELD, but not by reMELD-Na (bottom right quadrant), Thus, re-MELD-Na would have prioritized patients with a 90-day WL mortality HR of 1.6 as compared to currently prioritized patients. Figure 3.4 also illustrated that after refitting, no scores above 40 were calculated and thus that all high MELD scores were acknowledged correctly. By using more recent data and the true 90-day mortality rates of our population, reMELD-Na showed that very few patients actually approached 100% 90-day WL mortality, i.e. MELD 40. Thus, the refit models restored the clinical meaning of the 6-40-point range.

Discussion

In this study, for the first time the MELD score was refitted to the Eurotransplant data. By establishing new and evidence-based lower and upper bounds for each MELD parameter, the role of each MELD component was reweighed. The reweighed coefficients performed significantly better than the currently used UNOS-MELD in the independent validation dataset. The reMELD and reMELD-Na gave convincingly higher c-indices than UNOS-MELD and were based on the best fit to the current Eurotransplant data. The reMELD-Na prioritized patients with 1.6 times higher 90-day mortality rates than the currently prioritized patients. Thus, refitting MELD results in more accurate, effective and just mortality prediction and subsequent sickest-first allocation.

The UNOS-MELD has remained unchanged ever since it was constructed 20 years ago in a cohort of 231 patients.[4] Its parameter bounds were chosen arbitrarily.[5,6,11] Thus, UNOS-MELD is not fit for the

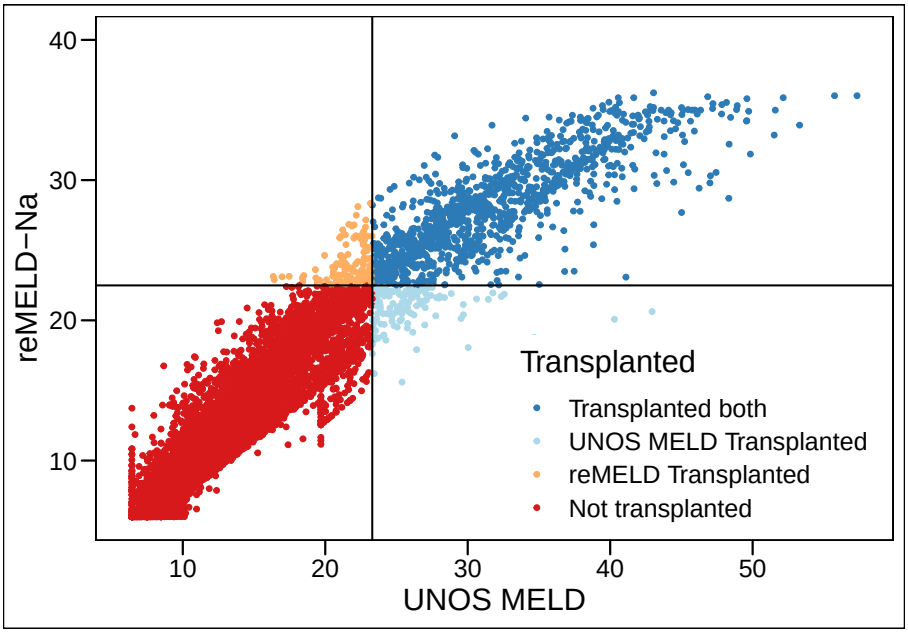


Figure 3.4: Correlation plot of UNOS- MELD and reMELD- Na. Based on the number of transplanted patients after the first 90 days (n = 1,248), the highest-ranked patients according to both scores separately were assigned a liver graft, as represented by the horizontal (graft granted by reMELD- Na) and vertical (by UNOS- MELD) lines. Patients in the top left quadrant (reMELD- Na-prioritized) had a 1.58 times higher risk of 90- day death compared to patients in the lower right quadrant (UNOS- MELD- prioritized).

changing LT candidate population, which showed through a decline in predictive power.[7] Refitting, i.e. re-establishing parameter bounds and weights, enables prediction models to change along with the population they serve. Indeed, the principle of refitting could be applied to any model used for survival prediction.

Lower bounds

By refitting, the lower border of creatinine was set to 0.7. A creatinine of 1.0 mg/dL might already indicate disease in LT candidates, as measured creatinine overestimates kidney function in e.g. sarcopenia, females and patients with high bilirubin.[15] Evaluation of the lower bounds of bilirubin and the INR showed that multiple combinations of bounds provided a good fit to the data, while preserving the predictive power of the model. Thus, the exact lower bounds should be determined through expert-based discussion. By acknowledging more low values (which most patients had at listing), the higher values were placed in a more appropriate context than with the UNOS lower bounds of 1.0.

Upper bounds

The upper bounds found in this study were perhaps more controversial, as UNOS-MELD uses none for bilirubin and INR. However, the new bounds resulted in better-performing models. Through refitting, serum creatinine became less important. Under UNOS-MELD, the number of transplanted patients with renal failure increased significantly, possibly due to overweighed creatinine in UNOS-MELD.[6,16] As these patients have increased morbidity and mortality both before and after LT, the principle of the sickest-first system was to prioritize them. However, one could question the prioritization of renal failure above liver failure, through the high weight of creatinine in UNOS-MELD, when allocating scarce liver grafts. High bilirubin levels led to unreliable measurements of UNOS-MELD due to interaction with creatinine, which influenced scores because of the weight of creatinine in UNOS-MELD.[17] Therefore, decreasing the weight of creatinine and establishing an upper bound for bilirubin should give more reliable reMELD scores. Of the three MELD parameters, INR is the most unreliable. This is in

part because the INR varies significantly depending on the method of laboratory measurement.[18] Also, medical treatment (or non-treatment) can decrease or increase the INR. Therefore, an upper bound for the INR would also be an improvement, as it would reduce the influence of outliers in INR measurements.[5]

Sodium addition

The UNOS regions have used MELD-Na for liver allocation since 2016.[3] Despite the proven impact of serum sodium levels on LT candidate survival,[12,14] Na is not used (yet) for the Eurotransplant liver allocation. The addition of Na to the reMELD gave a small but significant improvement in discriminative ability (c-index 0.866 to 0.869). Although the largest improvement in c-index was achieved by reMELD alone (0.849 to 0.866), the additional smaller gain still represented important changes for hyponatremic patients. The c-index measures the proportion of patient pairs whose ranking is correctly ordered. Hence, a difference in c-index can be thought of as the proportion of patients whose ranking change. It however does not measure the degree of change within ranks, i.e. for each patient. Thus, a small difference for many patients will give a high c-index increase, whereas a large change for a smaller number of (hyponatremic) patients gives little improvement.[12,14] Based on the current findings, reMELD-Na performed slightly but significantly better than reMELD. Also, it seems just to consider the proven effect of Na levels on mortality. Therefore, use of reMELD-Na is preferred.

Impact on the WL

Despite the seemingly small performance differences between UNOS and refit models, the refit models were very different at their bases, which was the goal of this study. Refitting established new parameter bounds, notably different coefficients and a superior fit to the data, see Figure 3.3 and Table 3.3. This improved both model discrimination (c-index) and calibration (prediction errors). The increase in c-index from 0.849 to 0.869 may seem small, but is both statistically and clinically very significant. Recent study showed that switching from UNOS MELD to MELD-Na would significantly reduce waiting list mortality in the Eurotransplant region, although the difference in c-index was 0.015 (0.832

vs 0.847).[14] The study that formed the basis of the US switch from MELD to MELD-Na, showed a similar increase in c-index (i.e. 0.868 to 0.883),[12] which was considered an important increase and convincing evidence for possible MELD-Na implementation. Another large UNOS cohort study on improving MELD showed a c-index increase from 0.75 to 0.77.[16] This illustrates that improving an already-high c-index is very difficult, as it increases in an asymptotic fashion when approaching its maximum. The highest obtainable baseline c-index is probably around 0.9 or lower because of possible imperfections and biological variation in the data.[5,12,14] Moreover, compared to respectively UNOS MELD and MELD-Na, refitting reduced prediction errors by 8% and 5%, which is a major improvement considering the already-high accuracy of the scores. To estimate the possible clinical impact of refitting, differences in prioritization were assessed, see Table 3.4. As the 90-day mortality of the reMELD-Na-prioritized patients (Figure 3.4) was 1.6 times higher than the currently prioritized patients, reMELD-Na could possibly better effectuate the sickest-first principle. Figure 3.4 also shows patients with MELD>=40, which were rescaled below 40 after refitting. An UNOS-MELD score of 40 originally corresponded to a 100% 90-day WL mortality.[11] However, over the past decades, the waitlist population and the risks of death per MELD score have changed,[7] which also shows through the increasing number and survival of MELD>=40 patients.[19] This has important implications for the Eurotransplant exception point system, which is based on MELD mortality rates dating from 2006 (supplement 3) and allocates 25-30% of the LT candidates.[10,20] Regardless of possible refit score implementation, the Eurotransplant exception point system would benefit from an accurate rescaling. Still, by quantile matching and refitting specifically in the 6 to 40 range, the refit scores restored their old mortality equivalents, i.e. MELD 40 represented a 100% 90-day mortality risk.

Limitations

Estimating the impact of a new allocation system based on another system's data inadequately reflects the possible effects of new allocation. Before implementation, one aims to answer important questions concerning counterfactual outcomes in causal inference, e.g. what would have happened to patients had they not been transplanted. The best way to evaluate a new allocation system is to bring it in practice and

measure the difference. Evaluating a new system through simulation is probably the next best option. One should be aware, however, that assessment through simulation is based on intrinsically unverifiable assumptions, namely that with changing the allocation priorities nothing else in the system will change. The Eurotransplant region does not yet have a simulation model of its liver allocation, like the Liver Simulation Allocation Model (LSAM) in the UNOS. Therefore, new allocation systems, e.g. refit models, cannot be formally evaluated before possible implementation. Instead, only a rough estimate of possible impact could be given by assessing differences in prioritized patients. Still, this was likely a less-biased method compared to proposed UNOS MELD-Na estimations of impact.[12] Finally, the role of clinical intuition and logic of reasoning should not be underestimated. Optimizing MELD for our region makes clinical sense and the log-likelihood-based approach is statistically solid and logical. Regions without simulation programs cannot know for certain what the effect of new allocation systems will be. Still, evidence can form a strong suggestion of improvement, which can be confirmed after possible implementation.

Conclusion

This study showed that updating the boundaries and coefficients on more recent region-specific data increased the predictive power of MELD again. The discussion on the establishment of refit models should consider at least three aspects: the parameter boundaries, fit of the model to the data and the prediction performance of the model. With the increasing interest in more advanced computational possibilities, the transplant community should investigate alternative models to the current allocation system.21 However, as the MELD still is the basis of liver allocation in many regions, efforts should be made to keep the model as relevant as possible, and we believe the current study serves this purpose. In conclusion, refitting MELD acknowledged more patient measurements at listing and prioritized patients with higher 90-day mortality. The discriminative ability and accuracy of refit models was a significant and relevant improvement compared to the currently used UNOS-MELD.

References

1. Eurotransplant International Foundation. Statistics Library.
2. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. Hepatology. 2001;33(2):464-470. doi:10.1053/jhep.2001.22172
3. OPTN. Organ procurement and transplantation network policies.
4. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology. 2000;31(4):864-871. doi:10.1053/he.2000.5852
5. Leise MD, Kim WR, Kremers WK, Larson JJ, Benson JT, Therneau TM. A revised model for end-stage liver disease optimizes prediction of mortality among patients awaiting liver transplantation. Gastroenterology. 2011;140(7):1952-1960. doi:10.1053/j.gastro.2011.02.017
6. Merion RM, Sharma P, Mathur AK, Schaubel DE. Evidence-based development of liver allocation: A review. Transpl Int. 2011;24(10):965-972. doi:10.1111/j.1432-2277.2011.01274.x
7. Godfrey EL, Malik TH, Lai JC, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. Am J Transplant. 2019;19(12):3299-3307. doi:10.1111/ajt.15559
8. Nagai S, Chau LC, Schilke RE, et al. Effects of Allocating Livers for Transplantation Based on Model for End-Stage Liver Disease-Sodium Scores on Patient Outcomes. Gastroenterology. 2018;155(October):1451-1482. doi:10.1053/j.gastro.2018.07.025
9. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73. doi:10.7326/M14-0698
10. Jochmans I, Van Rosmalen M, Pirenne J, Samuel U. Adult Liver Allocation in Eurotransplant. Transplantation. 2017;101(7):1542-1550. doi:10.1097/TP.0000000000001631
11. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology. 2003;124(1):91-96. doi:10.1053/gast.2003.50016
12. 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(10):1018-1026.

doi:10.1007/s11250-017-1262-3
13. van Houwelingen HC, Putter H. Dynamic Prediction in Clinical Survival Analysis. 1st ed. CRC Press; 2011.
14. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Validation of the Model for End-stage Liver Disease sodium (MELD-Na) score in the Eurotransplant region. Am J Transplant. 2020. doi:10.1111/ajt.16142
15. Saxena V, Lai JC. Kidney Failure and Liver Allocation: Current Practices and Potential Improvements. Adv Chronic Kidney Dis. 2015;22(5):391-398. doi:10.1053/j.ackd.2015.05.002
16. Sharma P, Schaubel DE, Sima CS, Merion RM, Lok ASF. Reweighting the Model for End-Stage Liver Disease Score Components. Gastroenterology. 2008;135(5):1575-1581. doi:10.1053/j.gastro.2008.08.004
17. Verna EC, Connelly C, Dove LM, et al. Center-Related Bias in MELD Scores Within a Liver Transplant UNOS Region. Transplantation. 2019;Published. doi:10.1097/tp.0000000000003031
18. Porte RJ, Lisman T, Tripodi A, Caldwell SH, Trotter JF. The international normalized ratio (INR) in the MELD score: Problems and solutions. Am J Transplant. 2010;10(6):1349-1353. doi:10.1111/j.1600-6143.2010.03064.x
19. Nadim MK, DiNorcia J, Ji L, et al. Inequity in organ allocation for patients awaiting liver transplantation: Rationale for uncapping the model for end-stage liver disease. J Hepatol. 2017;67(3):517-525. doi:10.1016/j.jhep.2017.04.022
20. Eurotransplant. Chapter 5 - ET Liver Allocation System (ELAS).; 2019. https://www.eurotransplant.org/cms/index.php?page=et_manual.
21. Spann A, Yasodhara A, Kang J, et al. Applying Machine Learning in Liver Disease & Transplantation: A Comprehensive Review. Hepatology. 2020:0-3. doi:10.1002/hep.31103

Part II: Disease over time

It is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.

— Abraham Maslow

Chapter 4

Joint modelling of liver transplant candidates outperforms the model for end-stage liver disease: the effect of disease development over time on patient outcome

Goudsmit BFJ, Braat AE, Tushuizen ME, et al. Joint modeling of liver transplant candidates outperforms the model for end-stage liver disease: The effect of disease development over time on patient outcome. *American Journal of Transplantation.* 2021; doi:10.1111/ajt.16730

Abstract

Background & Aims: Liver function is measured regularly in liver transplantation (LT) candidates. Currently, these previous disease development data are not used for survival prediction. By constructing and validating joint models (JMs), we aimed to predict outcome based on all available data, using both disease severity and its rate of change over time.

Methods: Adult LT candidates listed in Eurotransplant between 2007-2018 (n=16,283) and UNOS between 2016-2019 (n=30,533) were included. Patients with acute liver failure, exception points or priority status were excluded. Longitudinal MELD(-Na) data was modeled using spline-based mixed effects. Waiting list survival was modeled with Cox proportional hazards models. The JMs combined the longitudinal and survival analysis. JM 90-day mortality prediction performance was compared to MELD(-Na) in the validation cohorts.

Results: MELD(-Na) score and its rate of change over time significantly influenced patient survival. The JMs significantly outperformed the MELD(-Na) score at baseline and during follow-up. Baseline MELD-JM AUC was 0.94 (0.92-0.95) versus MELD AUC 0.87 (0.85-0.89). MELDNa-JM AUC was 0.91 (0.89-0.93) and MELD-Na AUC was 0.84 (0.81-0.87). The JMs were significantly ($p<0.001$) more accurate than MELD(-Na). After 90 days, we ranked patients for LT based on their MELD-Na and MELDNa-JM survival rates, showing that MELDNa-JM-prioritized patients had 3x higher waiting list mortality.

Conclusion: The MELD(Na)-JM significantly outperformed current models that drive liver allocation. Thus, patient survival can be dynamically predicted based on past and current disease. These predictions could more accurately direct treatment to those most in need.

Introduction

The shortage of available donor livers creates waiting lists of liver transplant (LT) candidates with end-stage liver disease.[1] In many countries, candidates with the lowest expected survival are ranked highest and thus usually treated first.[2] In the Eurotransplant and United Network for Organ Sharing (UNOS) regions, the survival prediction and subsequent ranking of LT candidates is based on the Model for End-stage Liver Disease (MELD) or MELD sodium (MELD-Na) score.[2] The MELD(-Na) score estimates 90-day mortality based on the last known measurement of serum creatinine, bilirubin and the INR (and sodium).[3–5] For patients awaiting LT, MELD(-Na) scores are repeatedly and regularly measured. These data are valuable for outcome prediction as they show the patient-specific disease development over time.[6,7] Clinically, it also makes sense to account for past disease and its severity when estimating prognosis. However, currently only the last available MELD(-Na) measurement is used for survival prediction and subsequent LT allocation. Previous data is ignored.

Joint models (JMs) are a recent statistical development that join longitudinal and survival analysis.[8] JMs can handle complex follow-up data, i.e. irregularity in number, interval and missing of measurements.[9] Also, JMs can use both the disease severity and its rate of change for survival prediction. This approximates disease as a dynamic process, whereas MELD(-Na) is static and underestimates fast-changing disease severity.[10,11] Previous work has shown that JMs can outperform Cox models.[12–14] JMs have however never been used to model patients with end-stage liver disease or any other large cohort data. The LT setting is interesting for evaluating JMs because statistical models, i.e. currently the MELD(-Na) score, determine who is offered transplantation first.

The goal of this study is to use joint models to improve prediction of waiting list mortality, by considering disease severity and its rate of change over time. Therefore, this study develops and validates JMs for LT waiting list survival prediction based on repeated MELD(-Na) measurements. We constructed and validated JMs both in the Eurotransplant and the United Network for Organ Sharing (UNOS) regions. Online survival prediction tools of the resulting MELD-JM and MELDNa-JM were created to allow predictions based on single-patient data.

Methods

The analyses were done separately for the Eurotransplant and UNOS regions, MELD- and MELD-Na based JMs were constructed and validated respectively.

Study population

For this study, waiting list data was used from Eurotransplant and the UNOS regions. For the Eurotransplant region, patients were followed between January 1ˆst, 2007 until December 31ˆst, 2018. For the UNOS, the study interval was from January 16ˆth 2016 (MELD-Na implementation) to December 31ˆst, 2019. Patients with acute liver failure, exception points or priority status at registration and listing for multiple organs were excluded. All other adult patients listed for a first LT were included. Longitudinal exception points were not modeled, as they do not reflect disease severity within the patient. Separate training (67% of the patients) and testing (33%) sets were constructed through random sampling. The longitudinal data of the waiting list contained repeated measurements of the MELD(-Na) score.ˆ4 Data from first active listing until delisting were used. Reasons for delisting were death, transplantation, removal or the end of study. Patients who were removed due to deteriorating clinical condition or who died within 30 days of removal were also counted as deceased. "Removal" comprised of removal from the waiting list due to improved clinical condition and censoring for exception points or priority status acquired during follow-up. The primary outcome of survival analysis was the overall waiting list mortality. Predictors were (repeated) MELD(-Na) scores. In table S7, results are shown of an additional model that also considers e.g. age, region and sex. For the longitudinal analysis, patients were censored at the end of the study follow-up. Also, patients receiving priority status or exception points during waiting were censored from that date, as transplant and thus death chances would change from that time point on. The sample size was set by the retrospective study design. Complete-case analysis was done.

Statistical analysis

Study variables following normal distributions are presented as mean±SD (standard deviation) and non-normal variables as median±IQR (interquartile range). Categorical variables are reported as counts and percentages.

Longitudinal analysis

The longitudinal MELD(-Na) data were modeled with mixed effect models. These calculate both the average (population) and individual (deviation of each patient from the average) MELD(-Na) development over time. Importantly, they model developments as continuous trajectories, which can also be non-linear, e.g. hyperbolical. This gives a natural approximation of disease over time, which contrasts the last measurement carried-on-forward approach of Cox models (figure S4). The fixed effects included: intercept (representation of disease severity at baseline) and time on the waiting list which were modeled with natural cubic splines (3 degrees of freedom). The random effect components, which varied to randomly deviate from the average for each individual, were intercept (baseline disease severity) and follow-up time on the waiting list.

Combining longitudinal and survival analysis

Next, the abovementioned mixed effects model was combined with a Cox model. The latter was fit to the outcome of waiting list mortality, censoring for all other outcomes, with MELD(-Na) as predictor. Thus, the MELD(-Na) joint models (MELD-JM and MELDNa-JM) were constructed using the R package "JMbayes."[15] The JMs predicted survival using both the value of the MELD(-Na) score and its rate of change at each moment in time (i.e. time-dependent slope). By considering time-dependent slopes, a more nuanced definition of disease severity is used for survival prediction, see Figure 4.1. Also, predictions are updated for each newly-available measurement, i.e. the model is dynamic.

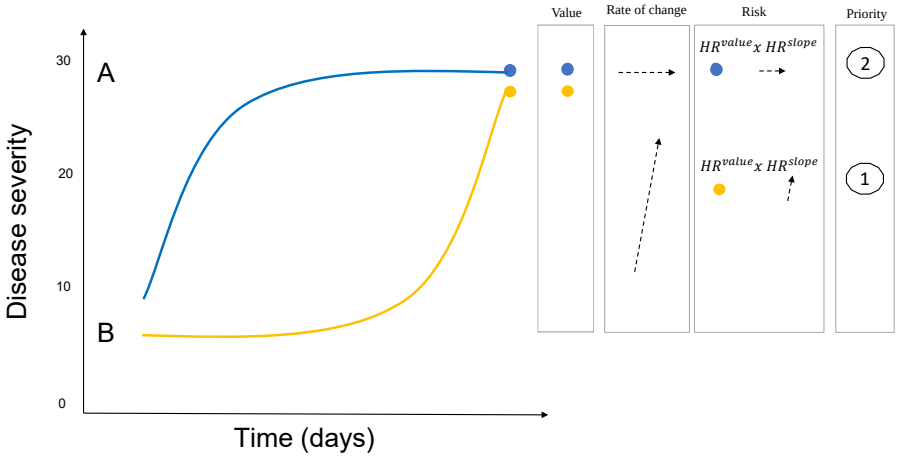


Figure 4.1: Two hypothetical patient trajectories on the LT waiting list are shown. Patient A initially increases and then stabilizes in disease severity. B is initially stable and later deteriorates. Under the current MELD(-Na) allocation, patient A would be prioritized over patient B in liver allocation, because the most recent MELD(-Na) is used. However, the JM uses both the past and current disease severity (value) and the rate of change at each moment in time (slope). At any given time, the JM combines the hazard ratio's for value and slope to calculate the risk of death. Thus, the JM would calculate a higher mortality risk and thus LT priority for patient B, because the disease is increasing fast.

Prediction performance

The JMs ability to predict 90-day mortality was assessed by calculating the area under the receiver operator curve (AUC) and prediction errors (Brier scores). Model performance was assessed at baseline (start of waiting list follow-up) and 3-monthly during follow-up of two years through bootstrap cross validation with 100 repetitions. To clarify, patients were censored if they did not die, but their data up until censoring would still be used when calculating performance. For comparison to currently-used models, MELD(-Na) prediction performance was also calculated at these time points.

Impact on the waiting list

Next, we estimated the possible impact of using the JMs instead of MELD(-Na) for waiting list prioritization. To do this, data from baseline to 90 days was used. At day 90, patients still on the waiting list were ranked highest-to-lowest based on their predicted 90-day mortality probability. This created a different ranking for the MELD(Na)-JM and MELD(-Na) models. The number of available donor livers in the first 90 days was then assigned to the highest ranking patients. This created a rough estimate who would have been offered LT first.[16,17] To further explain the possible differences in prioritization, baseline characteristics and the MELD(-Na) developments over time were compared between patients prioritized either by the MELD(Na)-JM or MELD(-Na).

Online LT-JM prediction tool

Lastly, online prediction tools of the MELD-JM (https://predictionmodels.shinyapps.io/meld-jm/) and MELDNa-JM (https://predictionmodels.shinyapps.io/MELDNa-JM/) were created. This allows interested readers to predict survival probabilities based on individual patient data. For the instruction manual, see supplement page 3. All the analyses were done with R v4.0.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Population characteristics

Table 4.1 shows the baseline characteristics for the Eurotransplant and UNOS populations. The 16,283 Eurotransplant LT candidates had a median age of 55 (48-61) at listing. Most (66.3%) patients were male and the most common liver diseases were (post)alcoholic (39.5%), cholestatic (11.7%) and hepatitis-C (10.7%) induced cirrhosis. At the end of follow-up, 50.2% were transplanted, 20.9% deceased, 20.2% were removed either due to improved clinical condition, priority status or exception points and 8.7% were censored at the end of study. The 30,533 UNOS patients had a median age of 58 (50-64) years and were mostly (63.3%) male. Alcohol- (30.5%) and NASH (20.7%) related liver cirrhosis were most common. The median MELD at listing was 18 (13-26), which was higher than the MELD 15 (11-21) for the Eurotransplant region. Median MELD-Na at listing was 19 (12-27) points in the UNOS cohort. At the end of follow-up, 52.2% was transplanted, 13% had died while waiting or was removed because of worsening clinical condition 31% was removed due to improved condition, exception point or status 1 approval during follow-up and 3.8% was censored at the end of study.

JM properties

The JMs calculates hazard ratios at a specific time (HRt) though the following equations, for MELD-JM:

$$HR_t = \left(1.29^{MELD_{value}}\right) * (8.12^{MELD_{slope}})$$

and MELDNa-JM:

$$HR_t = \left(1.24^{MELDNa_{value}}\right) * \left(8.02^{MELDNa_{slope}}\right)$$

The MELD-JM coefficient for MELD values is 1.29 with 95% CI (1.28-1.31). The MELD-JM slope coefficient is 8.12 (95% CI 1.27-50.38). For the MELDNa-JM these are 1.23 (95% CI 1.24-1.26) and 8.02 (95% CI 3.65-17.1) respectively. This means that at a given moment in time, a 1-point increase in MELD value will increase mortality risk by a factor 1.29, and a 1-point faster or slower change gives a factor 8.12

Table 4.1: Baseline characteristics for the Eurotransplant and UNOS regions

Region	Eurotransplant	UNOS
study interval	2007-2018	2016-2019
n	16283	30533
Age (median (IQR))	55.0 [48.0, 61.0]	58.0 [50.0, 64.0]
Gender male (%)	10,796 (66.3)	19,334 (63.3)
BMI (median (IQR))	25.6 [22.9, 29.2]	29.0 [25.0, 33.0]
Disease (%)		
Cirrhosis, Alcoholic	6432 (39.5)	9309 (30.5)
Cirrhosis, HCV	1742 (10.7)	4001 (13.1)
Cirrhosis, NASH	NA	6328 (20.7)
Cirrhosis, other causes	3794 (23.3)	4754 (15.6)
Cholestatic disease	1905 (11.7)	2422 (7.9)
Other	2410 (14.8)	3725 (12.2)
Serum measurement at listing (mean (SD))		
Creatinine in mg/dL	1.3 (3.0)	1.5 (1.4)
Bilirubin in mg/dL	6.0 (10.6)	7.0 (9.4)
INR	1.5 (0.6)	1.8 (0.9)
Sodium in mmol/L	NA	136 (5.0)
Dialysis dependency (%)	937 (5.8)	3223 (10.6)
MELD at listing (median(IQR))	15.0 [11.0, 21.0]	18.0 [13.0, 26.0]
MELD-Na at listing (median(IQR))	NA	19.0 [12.0, 27.0]
Status at delisting (%)		
Transplanted	8174 (50.2)	15928 (52.2)
Deceased	3404 (20.9)	3974 (13.0)
Removed from the waiting list	3289 (20.2)	9460 (31.0)
Censored at study end	1417 (8.7)	1171 (3.8)

Note:
NA: Eurotransplant has no complete data regarding this item, HCV: hepatitis-C induced, HCC: hepatocellular carcinoma,HU: high urgent status, NSE: (non)standard exception points, MELD: Model of End-stage Liver Disease

Table 4.2: 90-day mortality AUCs of the MELDNa-JM versus the MELD-Na, at baseline and during waiting list follow-up in the validation cohort.

	MELDNa-JM			MELD-Na			
Time (months)	AUC	low95	upp95	AUC	low95	upp95	p
0	0.91	0.89	0.93	0.84	0.81	0.87	<0.001
3	0.79	0.75	0.82	0.67	0.62	0.73	<0.001
6	0.80	0.76	0.84	0.69	0.61	0.75	<0.001
9	0.81	0.75	0.86	0.75	0.69	0.81	<0.001
12	0.74	0.66	0.81	0.69	0.58	0.79	NS
15	0.76	0.67	0.84	0.70	0.54	0.83	<0.001
18	0.78	0.69	0.86	0.76	0.62	0.87	NS
21	0.88	0.78	0.97	0.83	0.62	0.96	NS
24	0.72	0.60	0.85	0.68	0.42	0.86	NS

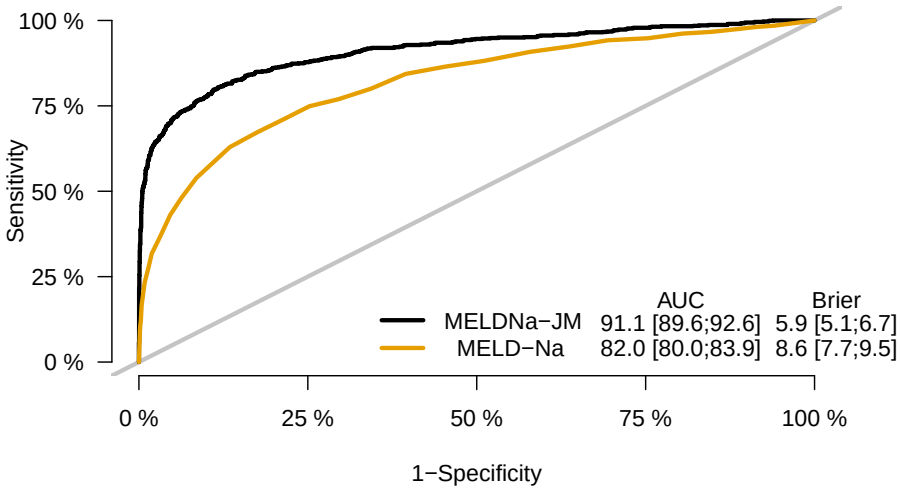
Note:
AUC: area under receiver operator curve, JM: joint model, MELD-Na: model for end-stage liver disease sodium score

difference. These equations, combined with the baseline risks, can be used to calculate specific risks. However, the JM is needed to calculate the MELD(-Na) value and slope at a given time point. To enable easy access to JM predictions, we developed online applications of the MELD-JM (https://predictionmodels.shinyapps.io/meld-jm/) and MELDNa-JM (https://predictionmodels.shinyapps.io/MELDNa-JM/). Interested readers can upload repeated MELD(-Na) measurements of individual patients into these applications, to generate personalized predictions. See supplement page 3 for an instruction manual. The performance of these JMs is tested below.

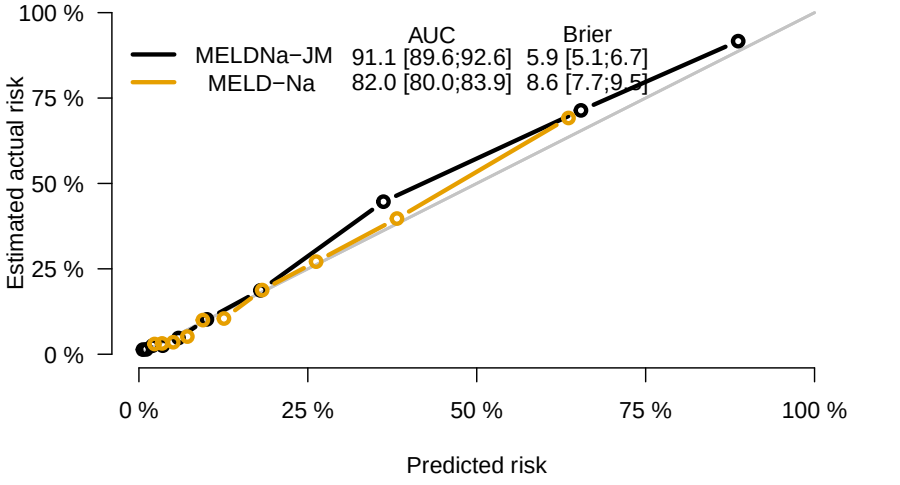
JM performance

The JM performance was assessed in the independent validation data at baseline (Figure 4.2 and figure S1) and during follow-up (Table 4.2: UNOS, table S1: Eurotransplant).

At baseline, MELDNa-JM AUC was 0.91 (0.89-0.93) and MELD-Na AUC was 0.84 (0.81-0.87). In Eurotransplant, MELD-JM AUC was 0.94



(a) 90-day mortality ROC plot of the MELDNa-JM and MELD-Na.



(b) Calibration plot of the MELDNa-JM and MELD-Na score. Each dot represents 10 percent of the population. The lines show how well the predicted risks match the observed risks

Figure 4.2: Performance measures for the MELDNa-JM and MELD-Na.

(95% CI 0.92-0.95) compared to 0.87 (0.85-0.89) for MELD (figures S1 and S2). For both the MELD(Na)-JM and MELD(-Na), prediction performance was best in the first months of follow-up. The MELD(Na)-JMs AUCs were significantly ($p<0.001$) better than the MELD(-Na) for the first 12 months of follow-up. During this period, the majority of transplantations was done, i.e. 94% (UNOS) and 84% (Eurotransplant). After 12 months, JMs AUCs were still notably but not significantly better than MELD(-Na). Over time, MELD(-Na) might be less representative of disease severity in LT candidates, which could explain the decrease in AUC over time for both models. MELD(Na)-JM prediction errors were always significantly lower than the MELD(-Na) (figure 2B, figure S2, tables S1 and S2). In other words, the JMs predictions were more accurate and thus better resembled the observed risks in the population. Subset analysis of prior (2007-2012) versus recent (2013-2018) years showed slightly better performance in the 2007-2012 cohort (table S4). Excluding HCV patients as sensitivity analysis increased AUCs (table S5). MELDNa-JM performed better in males (Figure S5), possibly because MELD-Na tends to underestimate female disease severity through lower creatinine levels.18 Performance was comparable for most diseases and worst in HCV disease (Figure S6). The implications for LT candidates might be limited, as the number of listed HCV patients is decreasing.19 Performance for non-black candidates was slightly better than for black candidates (Figure S7).

JM impact on the waiting list

The possible differences in MELDNa-JM and MELD-Na prioritization were assessed. Table 4.3 shows the baseline characteristics of patients that would have been prioritized both by MELDNa-JM and MELD-Na, by one of the models or by neither (table S6: MELD and MELD-JM comparison).

Table 4.3: 90-day mortality AUCs of the MELDNa-JM versus the MELD-Na, at baseline and during waiting list follow-up in the validation cohort.

Characteristics	Both	MELDNa-JM prioritized	MELD-Na prioritized	Not prioritized	p
n	3196	611	611	5658	
Age (median [IQR])	55.0 [47.0, 62.0]	56.0 [48.0, 63.0]	58.0 [51.0, 63.0]	59.0 [53.0, 64.0]	<0.001
Female sex (%)	1209 (37.8)	284 (46.5)	216 (35.4)	1978 (35.0)	<0.001
BMI (mean (SD))	29.9 (6.6)	28.5 (6.6)	28.6 (5.8)	29.1 (5.9)	<0.001
Death within 90 days (%)	498 (15.6)	94 (15.4)	26 (4.3)	135 (2.4)	<0.001
Disease (%)					
Cirrhosis HCV	235 (7.4)	39 (6.4)	87 (14.2)	973 (17.2)	
NASH	597 (18.7)	140 (22.9)	138 (22.6)	1204 (21.3)	
Cirrhosis Alcoholic	1413 (44.2)	209 (34.2)	230 (37.6)	1245 (22.0)	
Cirrhosis Other	575 (18.0)	108 (17.7)	83 (13.6)	761 (13.4)	
Cholestatic disease	185 (5.8)	68 (11.1)	33 (5.4)	533 (9.4)	
Metabolic disease	73 (2.3)	16 (2.6)	13 (2.1)	107 (1.9)	
Malignant/benign tumor	52 (1.6)	12 (2.0)	22 (3.6)	705 (12.5)	
Other	66 (2.1)	19 (3.1)	5 (0.8)	130 (2.3)	
MELD (median [IQR])	30.0 [26.0, 37.0]	21.0 [18.0, 24.0]	22.0 [19.0, 24.0]	14.0 [10.0, 17.0]	<0.001
MELD-Na (median [IQR])	31.0 [27.0, 35.0]	21.0 [19.0, 22.0]	25.0 [24.0, 27.0]	13.0 [9.0, 17.0]	<0.001

Compared to MELD-Na, the MELDNa-JM prioritized slightly younger (56 vs 58 years) and female (46.5% vs 35.4%) patients, who less often had hepatitis-C-induced liver cirrhosis. Most importantly, MELDNa-JM-prioritized patients had a 3.6 times higher 90-day mortality rate, i.e. 15.4% versus 4.3%. For the Eurotransplant region, MELD-JM prioritized patients with 5.0 times higher 90-day mortality compared to MELD, i.e. 23.2% versus 4.6% (table S6). A possible cause of this difference in mortality is illustrated in Figure 4.3.

The JM prioritized patients with lower median MELD-Na scores, see Table 4.3, but these patients had increasing disease severity at the time of liver graft allocation. This illustrates how not only the MELD-Na value, but also the rate of change is considered when estimating survival (figure S3 for Eurotransplant plots). The MELDNa-JM could therefore have prioritized patients with a higher waiting list mortality, possibly not captured by MELD-Na.

Online prediction tools

To access MELD-JM or MELDNa-JM predictions for the individual patient, please visit respectively https://predictionmodels.shinyapps.io/meld-jm/ or https://predictionmodels.shinyapps.io/MELDNa-JM/. See page 3 of the supplement for instructions. For clinical JM implementation in individual patients, repeated measurements of MELD(-Na) can be loaded into the online app. This essentially is the same data as uploaded to organ procurement organizations. The JM app then calculates prognosis based on these measurements and lets the user choose the moment in time and prediction horizon, e.g. assess 90-day survival probabilities after five months of waiting. These individual predictions can improve clinical decision making.

Discussion

This retrospective cohort analysis aimed to improve LT candidate survival prediction by using longitudinal data. Therefore, we developed and validated the MELD-JM and MELDNa-JM for waiting list mortality prediction in the Eurotransplant and UNOS regions. We report several important findings. First, the JM-calculated MELD(-Na) values and their time-dependent rate of change are significantly associated

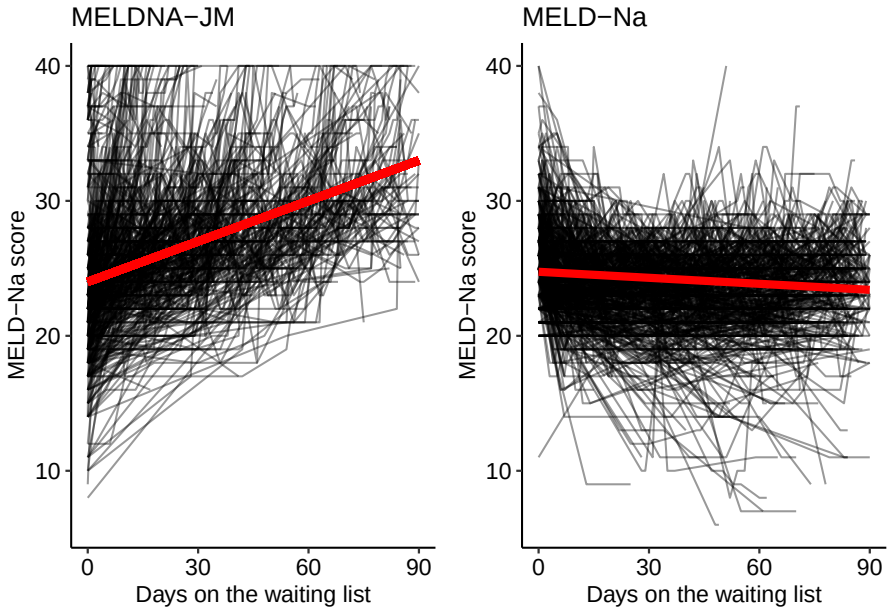


Figure 4.3: The MELDNa- JM and MELD- Na would prioritize different patients for liver transplantation. For these patients, we plotted the individual (black lines) and average (red line) MELD-Na score development during 90 days. Although the MELD-Na-prioritized patients had a higher initial MELD-Na score (value), their average scores remained stable (slope). In contrast, the JM-prioritized patients had lower MELD-Na (value) scores but with faster increasing disease severity (slope). Interestingly, the JM- prioritized patients had a five times higher 90- day mortality rate. Indicating that JM prioritization could possibly be more just.

with LT candidate waiting list mortality. Second, using time-dependent value and slope, the JMs significantly outperformed both MELD and MELD-Na when predicting mortality. Third, the JMs would have prioritized patients with three to five times higher mortality on the waiting list, who would not have been prioritized under MELD(-Na).

Longitudinal analysis

The progression of liver disease changes within and between patients over time. The current models that determine transplantation priority for patients with end-stage liver disease, i.e. the MELD(-Na), ignore previous disease development. However, for the clinician it is evident that the history of disease is important when estimating prognosis. Therefore, JMs were used to combine longitudinal and survival analysis.[8] The resulting MELD(Na)-JM estimate both the value and slope - i.e. current disease severity and the current rate of change- at each new measurement in time to predict survival, while also considering all previous measurements, see Figure 4.1. The resulting disease developments are a continuous and flexible trajectory over time, whereas e.g. time-dependent Cox (TDC) models carry the last measured value on forward.[20] This can fail to adequately model changing disease severity (figure S4) and can lead to underestimation of mortality in severely-ill LT candidates.[11] The idea of using MELD(-Na) rate of change for survival prediction is not new. Previously, the MELD spike and delta-MELD have been proposed.[6,21] The MELD spike indicates a 30% or higher difference between current MELD and the MELD score measured 7 days ago. It is a binary parameter based on cut-offs (30% and 7 days). However, through joint-modelling, we achieved a continuous representation of disease based on all data (not only assessing 30% differences or the past 7 days). MELD spike was intended as tiebreaker between patients with the same MELD scores. The JMs could however prioritize patients even if their MELD-Na values are lower, as long as the product of the value and slope is higher, see Figure 4.1 and 4.3. The delta-MELD is the difference between lowest MELD in previous 30 days and current MELD. It averages the slope over a varying number of previous days or measurements (depending on the date of lowest MELD). In our view, this makes it an imprecise approximation of current rate of change. Still, it is often considered as predictor in LT analysis.[22–26] However, Bambha et al. already showed that the effect of delta-MELD

depends on the frequency of measurements.[27] In contrast, the estimated slope of the MELD(Na)-JM is updated with each new measurement and is not altered by the frequency of measurements.

Prediction performance

The MELD(Na)-JM prediction performance was significantly better than MELD(-Na). The predictions also more accurately resembled the actual survival rates on the waiting list. Models on which treatment decisions are based should ascertain excellent accuracy.[28] Using additional predictors in JMs, such as age and sex, slightly improved AUCs after 12 months (table S7). However, this was a small improvement, while using these predictors adds to complexity and might be considered unethical. Therefore, MELD(-Na)-only JMs were primarily constructed. Others have also studied possible improvements to MELD(-Na). Recently, a machine-learning MELD-Na alternative was constructed by Bertsimas et al., i.e. the optimized prediction of mortality (OPOM) model.[29] Although OPOM outperformed MELD-Na, it also considered more (n=25 or 28) variables. Moreover, OPOM is based on classification analysis, i.e. is the patient alive after 90 days yes/no, instead of survival analysis, i.e. how much time passed until death or censoring. Other machine-learning techniques, like random survival forests and neural networks, do not seem to outperform Cox models, even in high-dimensional data.[30] Previous work did show that JMs outperform time-dependent Cox (TDC) models,[12–14] which is interesting considering the frequent use of TDC analysis for LT candidates.[6,7,24,27,31–33] We believe that the TDC last measurement carried-on-forward can give a suboptimal representation of disease (supplement figure 4). With changing disease severity, the TDC model either underestimates or overestimates disease severity. This is especially the case if few measurements are available or data is missing, which often occurs in LT candidate data.

Impact on the waiting list

We investigated the prioritization differences between the MELDNa-JM and MELD-Na, to give clinical meaning to the found statistical improvements. Considering the rate of change in disease severity helped to identify patients with worse prognosis, which illustrates the concepts shown in Figure 4.1. To optimize the sickest-first allocation and

transplantation benefit, it could therefore be interesting to use the JM-approximated course of disease for LT evaluation. Physicians can use the MELD(Na)-JM as online tool (see above) to predict outcome based on individual patient data. Also, on a center or waiting list population level, JMs can be applied to predict survival of each eligible patient every time a donor liver graft is offered. These predictions can be used alongside or eventually perhaps instead of MELD(-Na), because JM performance is good compared to MELD-Na and the same data is used. This is practical, because no changes would have to be made in the centers' routine of collecting and uploading data.

Limitations

A limitation is that data could be missing dependent on unobserved values. Statistical methods, like the JM and Cox model, assume missing at random (MAR) data. For the waiting list, this means that MELD(-Na) missingness should not depend on unobserved values, but it may depend on observed values. Because unobserved values cannot be observed, MAR cannot be proven in this study or any other Eurotransplant/UNOS registry analysis. We did however assess the relation between MELD(-Na) value and reporting frequency (supplement "missingness analysis"). Involuntary updates of low MELD(-Na) scores were done in only a small part of the data. Also, despite the fact that the most recent score was lower than the previous one, centers still reported these values and often well in time. The average time between measurements that were previously higher or lower did not differ substantially. Dependent missingness in low MELD(-Na) scores could lead to overestimation of waiting list mortality. A solution to alleviate possible bias could be to increase the mandatory update frequency of MELD(-Na) scores. Another limitation is that patients with exception points were excluded, because longitudinal modelling of arbitrarily assigned MELD points does not reflect disease severity. However, JMs could be used to model repeated AFP measurements, tumor characteristics and response to therapy. Also, the difference in waiting list prioritization between the MELD(-Na) and MELD(Na)-JM is a rough estimate, which depends on the chosen interval, i.e. for a shorter follow-up, presumably prioritization of the two indices would be more similar and vice versa. Furthermore, we did not study postoperative survival if the MELD(Na)-JM would have been used for allocation. This is because the JMs were not used to

drive allocation. We therefore only could have assessed postoperative survival after MELD(-Na) allocation and would not know how the JMs would have changed that. These questions concern counterfactual outcomes in causal inference, e.g. what would have happened to patients had they not been transplanted.[34] The best way to evaluate a new allocation system is to bring it in practice and measure the difference. Evaluating a new allocation system through simulation is probably the next best option. These extensive simulations were beyond the scope of this study. One should be aware, however, that assessment through simulation is based on intrinsically unverifiable assumptions, namely that with changing the allocation priorities nothing else in the system will change. Lastly, JMs are statistically complex and can give biased results if mis-specified. Therefore, construction should be done with care. To aid clinicians, we made online versions of our models freely available.

Conclusion

This study developed and validated the MELD-JM and MELDNa-JM prediction models for respectively the Eurotransplant and UNOS regions. The MELD(Na)-JM significantly outperformed current models that drive liver allocation. Thus, patient survival can be dynamically predicted based on past and current disease. These predictions could more accurately direct treatment to those most in need.

References

1. Eurotransplant. Annual Report 2017.; 2018. www.eurotransplant.org.
2. Tschuor C, Ferrarese A, Kuemmerli C, et al. Allocation of liver grafts worldwide – Is there a best system? J Hepatol. 2019;71(4):707-718. doi:10.1016/j.jhep.2019.05.025
3. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology. 2000;31(4):864-871. doi:10.1053/he.2000.5852
4. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology. 2003;124(1):91-96. doi:10.1053/gast.2003.50016
5. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and Mortality among Patients on the Liver-Transplant Waiting List. N Engl J Med. 2008;359(10):1018-1026. doi:10.1007/s11250-017-1262-3
6. Merion RM, Wolfe RA, Dykstra DM, Leichtman AB, Gillespie B, Held PJ. Longitudinal assessment of mortality risk among candidates for liver transplantation. Liver Transplant. 2003;9(1):12-18. doi:10.1053/jlts.2003.50009
7. Sharma P, Schaubel DE, Sima CS, Merion RM, Lok ASF. Reweighting the Model for End-Stage Liver Disease Score Components. Gastroenterology. 2008;135(5):1575-1581. doi:10.1053/j.gastro.2008.08.004
8. Rizopoulos R. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. 1st ed. Chapman and Hall/CRC; 2012.
9. Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes. Annu Rev Stat Its Appl. 2019;6(1):223-240. doi:10.1146/annurev-statistics-030718-105048
10. Goudsmit BFJ, Tushuizen ME, Putter H, Braat AE, van Hoek B. The role of the model for end-stage liver disease-sodium score and joint models for 90-day mortality prediction in patients with acute-on-chronic liver failure. J Hepatol. 2021;74(2):475-476. doi:10.1016/j.jhep.2020.08.032
11. Hernaez R, Liu Y, Kramer JR, Rana A, El-serag HB, Kanwal F. Model for end-stage liver disease-sodium underestimates 90-day

mortality risk in patients with acute-on-chronic liver failuare. J Hepatol. 2020. doi:10.1016/j.jhep.2020.06.005

12. Arisido MW, Antolini L, Bernasconi DP, Valsecchi MG, Rebora P. Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. BMC Med Res Methodol. 2019;19(1):1-13. doi:10.1186/s12874-019-0873-y
13. Rizopoulos D, Takkenberg JJM. Tools & techniques - Statistics: Dealing with time-varying covariates in survival analysis - joint models versus Cox models. EuroIntervention. 2014;10(2):285-288. doi:10.4244/EIJV10I2A47
14. Campbell KR, Juarez-Colunga E, Grunwald GK, Cooper J, Davis S, Gralla J. Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: Application to kidney transplantation. BMC Med Res Methodol. 2019;19(1):1-12. doi:10.1186/s12874-019-0773-1
15. Rizopoulos D. The R package jmbayes for fitting joint models for longitudinal and time-to-event data using MCMC. J Stat Softw. 2016;72(7). doi:10.18637/jss.v072.i07
16. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Refitting the model for end-stage liver disease for the Eurotransplant region. Hepatology. 2020;in press.
17. Leise MD, Kim WR, Kremers WK, Larson JJ, Benson JT, Therneau TM. A revised model for end-stage liver disease optimizes prediction of mortality among patients awaiting liver transplantation. Gastroenterology. 2011;140(7):1952-1960. doi:10.1053/j.gastro.2011.02.017
18. Myers RP, Shaheen AAM, Aspinall AI, Quinn RR, Burak KW. Gender, renal function, and outcomes on the liver transplant waiting list: Assessment of revised MELD including estimated glomerular filtration rate. J Hepatol. 2011;54(3):462-470. doi:10.1016/j.jhep.2010.07.015
19. Kwong AJ, Kim WR, Lake JR, et al. OPTN/SRTR 2019 Annual Data Report: Liver. Am J Transplant. 2021;21(S2):208-315. doi:10.1111/ajt.16494
20. Dekker FW, De Mutsert R, Van Dijk PC, Zoccali C, Jager KJ. Survival analysis: Time-dependent effects and time-varying risk factors. Kidney Int. 2008;74(8):994-997. doi:10.1038/ki.2008.328

21. Massie AB, Luo X, Alejo JL, Poon AK, Cameron AM, Segev DL. Higher Mortality in Registrants With Sudden Model for End-Stage Liver Disease Increase: Disadvantaged by the Current Allocation Policy. Liver Transplant. 2015;21(5):683-689. doi:10.1002/lt
22. Belli LS, Berenguer M, Cortesi PA, et al. Delisting of liver transplant candidates with chronic hepatitis C after viral eradication: A European study. J Hepatol. 2016;65(3):524-531. doi:10.1016/j.jhep.2016.05.010
23. Cholankeril G, Li AA, Dennis BB, et al. Pre-Operative Delta-MELD is an Independent Predictor of Higher Mortality following Liver Transplantation. Sci Rep. 2019;9(1):8312. doi:10.1038/s41598-019-44814-y
24. Györi GP, Silberhumer GR, Rahmel A, et al. Impact of dynamic changes in MELD score on survival after liver transplantation – a Eurotransplant registry analysis. Liver Int. 2016;36(7):1011-1017. doi:10.1111/liv.13075
25. Schlegel A, Linecker M, Kron P, et al. Risk Assessment in High- and Low-MELD Liver Transplantation. Am J Transplant. 2017;17(4):1050-1063. doi:10.1111/ajt.14065
26. Northup PG, Berg CL. Preoperative delta-MELD score does not independently predict mortality after liver transplantation. Am J Transplant. 2004;4(10):1643-1649. doi:10.1111/j.1600-6143.2004.00593.x
27. Bambha K, Kim WR, Kremers WK, et al. Predicting survival among patients listed for liver transplantation: An assessment of serial MELD measurements. Am J Transplant. 2004;4(11):1798-1804. doi:10.1111/j.1600-6143.2004.00550.x
28. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7
29. Bertsimas D, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. Am J Transplant. 2019;19(4):1109-1118. doi:10.1111/ajt.15172
30. Kantidakis G, Putter H, Lancia C, Boer JD, Braat A, Fiocco M. Survival prediction models since liver transplantation - comparisons between Cox models and machine learning techniques. BMC Med Res Methodol. 2020;20(277):1-14. doi:10.21203/rs.3.

rs-22670/v1.
31. Luo X, Leanza J, Massie AB, et al. MELD as a metric for survival benefit of liver transplantation. Am J Transplant. 2018;18(5):1231-1237. doi:10.1111/ajt.14660
32. Schaubel DE, Guidinger MK, Biggins SW, et al. Survival benefit-based deceased-donor liver allocation. Am J Transplant. 2009;9(4 PART 2):970-981. doi:10.1111/j.1600-6143.2009.02571.x
33. Sharma P, Schaubel DE, Gong Q, Guidinger M, Merion RM. End-stage liver disease candidates at the highest model for end-stage liver disease scores have higher wait-list mortality than status-1A candidates. Hepatology. 2012;55(1):192-198. doi:10.1002/hep.24632
34. van Geloven N, Swanson SA, Ramspek CL, et al. Prediction meets causal inference: the role of treatment in clinical prediction models. Eur J Epidemiol. 2020;35(7):619-630. doi:10.1007/s10654-020-00636-1

Chapter 5

Development and validation of a dynamic survival prediction model for patients with acute-on-chronic liver failure

Goudsmit BFJ, Braat AE, Tushuizen ME, et al. Development and validation of a dynamic survival prediction model for patients with acute-on-chronic liver failure. *JHEP Reports*. 2021; doi: 10.1016/j.jhepr.2021.100369.

Abstract

Background & Aims: Acute-on-Chronic-Liver Failure (ACLF) involves an acute deterioration of liver function in patients with chronic liver disease. ACLF is usually associated with a precipitating event and results in the failure of other organ systems and high short-term mortality. Currently-used prediction models fail to adequately estimate prognosis and need for liver transplantation (LT) in ACLF. This study develops and validates a dynamic prediction model for ACLF patients, that uses both longitudinal and survival data.

Methods: Adult patients on the UNOS waitlist for LT between 11.01.2016-31.12.2019 were included. Repeated model for end-stage liver disease sodium (MELD-Na) measurements were jointly-modeled with Cox survival analysis to develop the ACLF joint model (ACLF-JM). Model validation was done in separate testing data with area under curve (AUC) and prediction errors. An online ACLF-JM tool was created for clinical application.

Results: In total, 30,533 patients were included. ACLF grade 1 to 3 was present in respectively 16.4, 10.4 and 6.2% of the patients. The ACLF-JM predicted survival significantly ($p<0.001$) better than the MELD-Na, both at baseline and during follow-up. For 28- and 90-day predictions, ACLF-JM AUCs ranged between 0.840-0.871 and 0.833-875, respectively. Compared to MELD-Na, AUCs and prediction errors were improved by 23.1%-62.0% and 5%-37.6% respectively. Also, the ACLF-JM could have prioritized patients who had four times higher waiting list mortality, possibly not identified by MELD-Na.

Conclusion: The ACLF-JM dynamically predicts outcome based on current and past disease severity. Prediction performance is excellent over time, even in ACLF-3 patients. Therefore, the ACLF-JM could be used as clinical tool in the evaluation of prognosis and treatment in patients with ACLF.

Introduction

Liver transplantation (LT) is a lifesaving treatment for patients with acute-on-chronic-liver failure (ACLF). ACLF is characterized by an acute deterioration of liver function in patients with chronic liver disease, often started by a precipitating event. ACLF results in the failure of one or more organs and is associated with high short-term mortality.[1–3] The current model that prioritizes patients for LT, the Model for End-stage Liver Disease sodium (MELD-Na) score,[4,5] underestimates disease severity in ACLF.[6,7] This is because MELD-Na does not consider temporal development of single or multiorgan failure (involving the 6 major organs/systems—i.e. liver, kidney, brain, coagulation, circulation, and respiration). This underestimation of predicted waitlist mortality results in lower access to transplantation for ACLF patients.[7] Sundaram et al. showed that ACLF death and waiting list removal rate were highest in ACLF-3 patients with MELD-Na <25.[8] Given that 20.9% of UNOS LT candidates between 2005-2016 had a form of ACLF,8 the inequal transplantation access might be substantial.

The MELD-Na uses one moment in time, i.e. the most recent measurement, to predict outcome.[4,5] It therefore ignores previous data valuable for survival estimation. However, ACLF is a dynamic disease with a clinical course that can change within days, resulting in very different outcomes.[9,10] Thus, there is a need for prediction models that estimate ACLF survival based on disease development over time.[7] The Chronic Liver Failure Consortium Organ Failure score (CLIF-C OFs) and CLIF-C ACLF score were developed for this purpose and showed better performance than the MELD-Na.[3,6] However, they also assessed only one moment in time. A joint model (JM) is a novel prediction model that simultaneously uses longitudinal and survival data.[11] It approximates changing disease severity over time and uses this for survival prediction.[12] JMs have shown superior predictive performance over Cox models.[12–14] However, they have not been applied to ACLF.

We hypothesized that using disease development over time to dynamically predict prognosis could improve survival prediction in ACLF patients. Much like a clinician, we aimed to use disease severity and its rate of change to predict outcome. We believe this is warranted in ACLF, because of the dynamic nature of ACLF disease and the current under-

estimation of mortality by MELD-Na.[9,10,15] Therefore, we constructed and validated a multivariate prediction model for survival prediction in ACLF patients: the ACLF Joint Model (ACLF-JM). We investigated the ACLF-JM 28- and 90-day survival prediction performance in the United Network for Organ Sharing (UNOS) registry and compared its performance to the MELD-Na score. We also investigated whether the ACLF-JM would identify patients in whom MELD-Na underestimates mortality. For easy clinical application, an online ACLF-JM tool was developed for dynamic survival prediction in ACLF patients.

Methods

The TRIPOD statement was used for the development and validation of this multivariate prediction model.[16]

Study population

Data of LT candidates was requested from the UNOS. We included adult (>=18 years) patients listed for a first LT between January 11th, 2016 (after MELD-Na implementation) and December 31st, 2019. We excluded candidates with acute liver failure (ALF) and hepatocellular carcinoma (HCC) at baseline. Data were used from first active listing until the earliest of patient death, transplantation, removal or censor at December 31st, 2019. Death was defined both as death while listed and removal for being too sick to transplant.[8] If patients received exception points or a status 1 (i.e. high urgency status) after first listing, they were censored from that date. MELD-Na data was missing in 0.05%, therefore complete-case analysis was done. Missing values for the predictors life support dependency (variable CAN_LIFE_SUPPORT, 0.00009% missing) and spontaneous bacterial peritonitis (CAN_BACTERIA_PERIT, 0.005% missing) were set to 'no.'

Identification of ACLF

Baseline ACLF was defined according to the to the European Foundation for the Study of Chronic Liver Failure (EF Clif) criteria.[3] Specifically, liver failure was defined as serum bilirubin >=12 mg/dL, kidney

failure as serum creatinine >=2.0 mg/dL or renal replacement therapy, cerebral failure as presence of hepatic encephalopathy grade 3-4, coagulation failure as INR >=2.5. Like other authors that used United Network for Organ Sharing (UNOS) data, we used mechanical ventilation as replacement for respiratory failure, since data on PaO2/FiO2 were not available. Also, life-support dependency was used to designate circulatory failure.[6,8,10,17]

Development of the ACLF-JM

Data were randomly split in a training (67% of the patients) and a testing (33%) set, for model development and validation respectively. The ACLF-JM consists of two parts: a longitudinal (mixed-effect) and survival (Cox proportional hazards) model. Mixed-effect models were used because they estimate disease development over time as a continuous trajectory and can model both linear (chronic, stable disease) and non-linear (fast deterioration in ACLF) developments. See figure S4 for an illustration. Thus, repeated measurements of MELD-Na scores were modeled with mixed-effects. Additional predictors were used to correct the longitudinal data. To start, 50 candidate variables were assessed (table S2). We excluded some variables a priori, because they referred to pediatric recipients, exclusion criteria, or donor characteristics. Variable relation to mortality was studied in univariate analysis and then variables were backwards selected variables for multivariate Cox analysis. The final variables included in the model contributed most significantly besides the those used for ACLF scoring through EF CliF criteria (serum bilirubin, creatinine, renal replacement therapy, encephalopathy grade, INR, mechanical ventilation, and life-support dependency). Thus, we additionally corrected for candidate age (years), sex (male/female), life support dependency (yes/no), presence of bacterial peritonitis (yes/no), presence of cirrhosis (alcohol-induced, hepatitis-C virus, non-alcoholic steatohepatitis (NASH) or other cirrhosis) (yes/no) and CLIF-C OF score (No ACLF or ACLF grade 1 to 3) (table S1). Next, a Cox proportional hazards model was constructed for waiting list mortality, using the same predictors as the mixed-effect model. Then, the ACLF-JM was constructed by joint-modelling the longitudinal (mixed-effect) and survival (Cox) model.[18] A key feature is that the ACLF-JM uses both the estimated MELD-Na value and the rate of change in MELD-Na (the slope of the decrease/increase) over

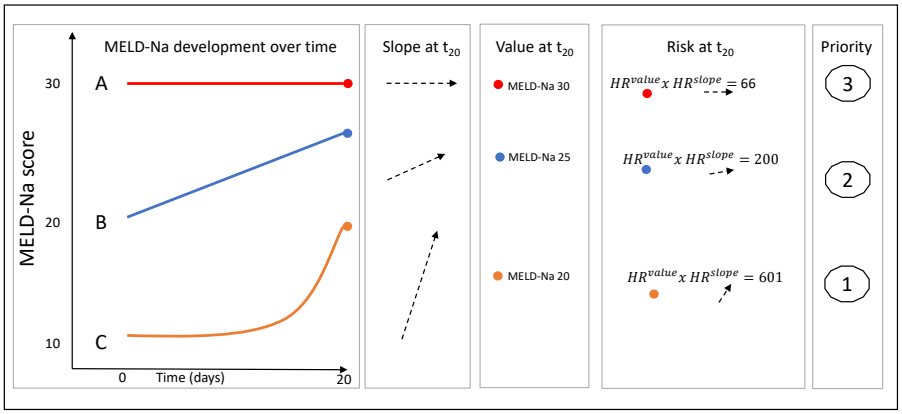


Figure 5.1: For three hypothetical patients A, B and C, the 20-day MELD-Na development is shown. After 20 days, patient A has a MELD-Na score of 30 and is thus prioritized by the current allocation system. However, the ACLF-JM uses both the estimated value (measured MELD-Na score) and slope (rate of change) at time=20 for survival prediction. Calculation of the HRs shows that the ACLF-JM gives patient C the greatest risk of death, because of the fast increase in MELD-Na scores (positive slope). See supplement 4 for the precise explanation and calculation.

time for survival prediction. For clarity, these concepts of value and slope were illustrated in Figure 5.1.

Validation of the ACLF-JM

Next, the prediction performance of the ACLF-JM was compared to the MELD-Na at various points in time in the separate testing data. Specifically, predictions were assessed at baseline and after a follow-up of 48 hours, 7 days and 14 days (similar to the validation study of the CLIF-C OF).6 Outcomes were 28-day and 90-day survival. For both the ACLF-JM and MELD-Na Cox model, the area under the receiver-operator-characteristic curve (AUC) and prediction errors were calculated and compared (see supplement 3 for detailed information). These measures and their 95% confidence intervals (95%CI) and p-values were calculated using the R package JM and bootstrapping.[18]

ACLF-JM impact on the transplantation waiting list

Next, we assessed the possible effect of using the ACLF-JM instead of MELD-Na to estimate mortality and subsequently prioritize patients for LT. This was of interest, because ACLF patients are likely underserved in the current LT allocation.[15] To assess possible differences in MELD-Na and ACLF-JM waitlist prioritization of patients, we followed patients from baseline until day 28.[6] Within this period, each time a liver graft was offered, patients were ranked two times from most to least ill based on their estimated survival without transplant. One ranking was made with the ACLF-JM predictions and one based on MELD-Na. Thus, for each model, patients were ranked 2636 times, i.e., the total number of available liver grafts within the first 28 days. After a liver graft offer, the transplanted patient was removed from the waiting list. We assumed that the highest ranked patients were transplanted, which is not necessarily true, and thus that the number of available transplants in the first 28 days represented the threshold of receiving transplantation. We then assessed which patients were prioritized according to what model. After 28 days and 2636 rankings, patients were stratified in four groups: those who are prioritized and possibly transplanted within 28 days according to both scores, those who are prioritized by either the ACLF-JM or MELD-Na score (but not by both) and those who are not prioritized by both. We also assessed the characteristics of the differently-prioritized patients, to see why patients were prioritized differently.

Clinical application of the ACLF-JM

Lastly, an online version of the ACLF-JM was created (https://predictionmodels.shinyapps.io/aclf-jm/), which allows clinicians to assess ACLF-JM survival predictions for their individual patient(s). Plots can be created of these dynamic predictions, to show the updating survival estimate for every new available measurement during follow-up. For a instruction manual, see supplement 1 and 2. All statistical analyses were performed using R v4.0.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study population

In total, we included 30,533 patients with 249,030 measurements. Table 5.1 shows the baseline characteristics of the study population. ACLF at baseline was seen in 33.3% of the patients; 15.9% had ACLF grade 1, 10.3% had grade 2 and 7.1% had grade 3. In these patients, liver (47.2%) and kidney (63.6%) failure were the most common. With increasing ACLF grade, median [IQR] age decreased, ranging from 59 [52-64] (no ACLF) to 53 [43-60] years (ACLF-3). Most patients were male (no ACLF: 65.0%, ACLF: 60%) and had alcoholic liver disease (no ACLF 25.8%, ACLF 40%). For ACLF grades 0 to 3, median [IQR] MELD-Na scores at listing were 15 [10-22], 27 [23-31], 33 [29-37] and 37 [31-42]. Average time on the waiting list was 150 days for patients without ACLF, 89 for ACLF grade 1, 24 for grade 2 and 10 days for grade 3. Cumulative incidence plots showed significantly higher death and transplantation rates in ACLF patients (figure S1). At the end of follow-up, 10.9% of the patients without ACLF died. For patients with ACLF grade 1 to 3, death rates were 16.7%, 14.3% and 22.4%, respectively.

Table 5.1: Baseline characteristics of UNOS liver transplantation candidates between 2016 to 2019 (n=30,533)

Characteristics	No ACLF	ACLF (any grade)	p	ACLF-1	ACLF-2	ACLF-3	p
Number of patients (%)	20,384 (66.7)	10,149 (33.3)		4,843 (15.9)	3,147 (10.3)	2,159 (7.1)	
Age (median [IQR])	59 [52, 64]	55 [47, 62]	<0.001	58 [50, 64]	53 [44, 60]	53 [43, 60]	<0.001
Male gender	13240 (65)	6094 (60)	<0.001	2905 (60.0)	1919 (61.0)	1270 (58.8)	<0.001
BMI (median [IQR])	28 [25, 33]	29 [25, 33]	<0.001	28 [24, 33]	29 [25, 34]	30 [26, 35]	<0.001
Days waiting (median [IQR])	58 [14, 193]	12 [4, 40]	<0.001	27 [9, 93]	8 [4, 20]	5 [3, 10]	<0.001
Status after waiting							
Censored (December 31, 2019)	986 (4.8)	185 (1.8)		129 (2.7)	43 (1.4)	13 (0.6)	
Deceased	2229 (10.9)	1745 (17.2)		810 (16.7)	451 (14.3)	484 (22.4)	
Transplanted	8681 (42.6)	7247 (71.4)		3187 (65.8)	2472 (78.6)	1588 (73.6)	
Removed	8488 (41.6)	972 (9.6)		717 (14.8)	181 (5.8)	74 (3.4)	
Grouped cause of disease (%)							
Cirrhosis HCV	3084 (15.1)	917 (9.0)		556 (11.5)	205 (6.5)	156 (7.2)	
NASH	4359 (21.4)	1969 (19.4)		1184 (24.4)	500 (15.9)	285 (13.2)	
Cirrhosis Alcoholic	5252 (25.8)	4057 (40.0)		1680 (34.7)	1431 (45.5)	946 (43.8)	
Cirrhosis Other	2976 (14.6)	1778 (17.5)		682 (14.1)	616 (19.6)	480 (22.2)	
Cholestatic disease	1810 (8.9)	612 (6.0)		343 (7.1)	182 (5.8)	87 (4.0)	
Metabolic disease	408 (2.0)	245 (2.4)		112 (2.3)	81 (2.6)	52 (2.4)	
Malignant/benign tumor	2119 (10.4)	266 (2.6)		194 (4.0)	42 (1.3)	30 (1.4)	
Other	376 (1.8)	305 (3.0)		92 (1.9)	90 (2.9)	123 (5.7)	
MELD-Na score (median [IQR])	15 [10, 20]	30 [25, 35]	<0.001	27 [22, 31]	33 [29, 37]	37 [31, 42]	<0.001
Bacterial peritonitis (%)	1560 (7.7)	1533 (15.1)	<0.001	643 (13.3)	508 (16.1)	329 (17.4)	<0.001
Failure organ/system (%)							
Liver	540 (2.6)	4789 (47.2)	<0.001	1018 (21.0)	2007 (63.8)	1764 (81.7)	<0.001
Kidney	0 (0.0)	6457 (63.6)	<0.001	2958 (61.1)	1717 (54.6)	1782 (82.5)	<0.001
Coagulation	254 (1.2)	3699 (36.4)	<0.001	667 (13.8)	1613 (51.3)	1419 (65.7)	<0.001
Cerebral	806 (4.0)	2095 (20.6)	<0.001	164 (3.4)	697 (22.1)	1234 (57.2)	<0.001
Circulatory	22 (0.1)	1193 (11.8)	<0.001	36 (0.7)	221 (7.0)	936 (43.4)	<0.001
Respiratory	0 (0.0)	662 (6.5)	<0.001	0 (0.0)	39 (1.2)	623 (28.9)	<0.001

Model properties

The ACLF-JM is summarized by the equation:

$$Hazard\ Ratio\ death_t = 1.15^{MELDNa_{value\ t}} * 1.02^{MELDNa_{slope\ t}} * 1.38^{age} * 0.75^{female\ gender} * 0.95^{cirrhosis} * \left(if : 1.06^{ACLF1}\right) * \left(if : 1.98^{ACLF2}\right) * \left(if : 5.90^{ACLF3}\right) * 1.18^{SBP} * 1.35^{lifesupport}$$

The ACLF-JM estimates the MELD-Na value and slope at a given timepoint and calculates the HR of death. For each MELD-Na point increase, the risk of 1-year death increases with 15% (95% CI 14-16). For every 1-point increase in slope, i.e. acceleration of disease increase, the mortality risk increases with 2% (95% CI 1-2). Of course, in clinical practice, disease severity often changes more rapidly, especially for ACLF patients. A more intuitive illustration of the effect of MELD-Na value and slope is provided in Figure 5.1, where three hypothetical awaiting LT are shown. The example calculation (details in supplement 4) shows that considering the rate of change (slope) in disease severity adds important information. Considering both MELD-Na value and slope would give priority to patient C (MELD-Na score 20, accelerating disease severity), whereas using the current MELD-Na-based allocation would prioritize patient A (MELD-Na 30, stable disease).

Model validation

The ACLF-JM prediction performance was validated in separate testing data. Table 5.2 shows the 28- and 90-day prediction performance of the ACLF-JM and MELD-Na, stratified for patients with and without ACLF, at baseline and during follow-up. For all time points and studied outcomes, the JM performance was significantly better than MELD-Na. At baseline in ACLF patients, the ACLF-JM AUC was 0.875 (95% CI 0.840-0.909) and MELD-Na AUC was 0.780 (95% CI 0.737-0.823). During follow-up, AUCs of both models declined to 0.833 (0.799-0.868) and 0.719 (0.677-0.761) respectively, which is still excellent for the ACLF-JM and respectable for the MELD-Na (also see figure S2A and S3).

Table 5.2: Mortality prediction AUC of the ACLF-JM versus the MELD-Na in patients with and without ACLF, at baseline and during follow-up

	ACLF				No ACLF			
Time	ACLF-JM	95% CI	MELD-Na	95% CI	ACLF-JM	95% CI	MELD-Na	95% CI
28-day mortality								
Baseline	0.871	0.844-0.898	0.788	0.754-0.822	0.774	0.717-0.831	0.706	0.643-0.769
48 hours	0.871	0.844-0.898	0.786	0.752-0.820	0.794	0.741-0.847	0.728	0.668-0.788
7 days	0.862	0.833-0.890	0.753	0.716-0.789	0.810	0.761-0.859	0.740	0.684-0.796
14 days	0.840	0.803-0.878	0.731	0.685-0.777	0.833	0.788-0.879	0.748	0.694-0.802
90-day mortality								
Baseline	0.875	0.840-0.909	0.780	0.737-0.823	0.836	0.807-0.865	0.734	0.700-0.768
48 hours	0.870	0.837-0.903	0.777	0.735-0.818	0.838	0.810-0.867	0.736	0.703-0.770
7 days	0.861	0.832-0.891	0.755	0.717-0.792	0.835	0.806-0.864	0.722	0.687-0.757
14 days	0.833	0.799-0.868	0.719	0.677-0.761	0.837	0.809-0.865	0.717	0.682-0.752

Note:
ACLF: acute-on-chronic liver failure, AUC: area under receiver operator curve, JM: joint model, MELD-Na: model for end-stage liver disease sodium score
* All AUCs differed significantly ($p<0.001$)

Figure 5.2 show that with increasing ACLF grade, JM performance remains significantly better than the declining MELD-Na (also see table S3 and figure S3). Especially for 90-day prediction in ACLF grade 3 patients, JM performance is excellent with AUCs ranging from 0.841 to 0.853, contrasting the MELD-Na AUCs between 0.613 and 0.693. MELD-Na AUCs (almost) equal chance when predicting 28-day mortality in ACLF-3 patients, ranging from 0.497 to 0.605. Importantly, the ACLF-JM also better estimated risks, i.e. is better calibrated, than the MELD-Na (figure S2B). With increasing ACLF grade, prediction errors were improved up to 37.6% (figure S3B). An accurate model is important for clinical decision-making, because decisions are often based on risks.[19]

ACLF-JM impact on the transplantation waiting list

To study the difference in survival prediction and subsequent allocation priority between the ACLF-JM and the MELD-Na, patients were followed the first 28 days. In total, 2636 transplants were done within this period. Figure 5.3 shows the correlation plot between MELD-Na scores and ACLF-JM mortality estimates after 28 days of waiting list follow-up. For 2186 patients (in green), transplantation priority was given according to both the ACLF-JM and MELD-Na, as estimated mortality without LT was highest. More interestingly, 450 patients (in blue) could possibly have been prioritized by the ACLF-JM, but not by MELD-Na.

Importantly, although these patients had lower median MELD-Na scores, they also had four times higher 28-day mortality rates, i.e., 13.1% versus 3.1%, see Table 5.3. Compared to the 450 MELD-Na-prioritized patients (orange), ACLF-JM-prioritized patients were older, more often female, had lower ACLF-1 rates, more NASH, less alcohol-induced liver disease and were more often dependent on life-support. After 28 days, 190 patients were delisted due to increased disease severity. In these patients, survival prediction AUC (95%CI) of the ACLF-JM and MELD-Na was 88.0 (85.1-90.9) and 82.5 (79.0-85.9), respectively (figure S6).

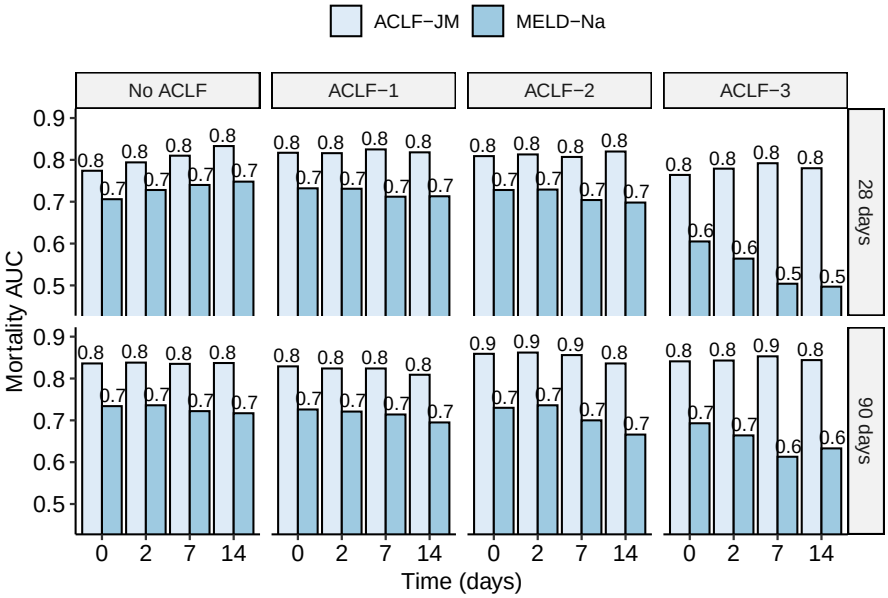


Figure 5.2: The AUCs for 28- and 90-day mortality prediction of the ACLF-JM and the MELD-Na, stratified for ACLF severity.

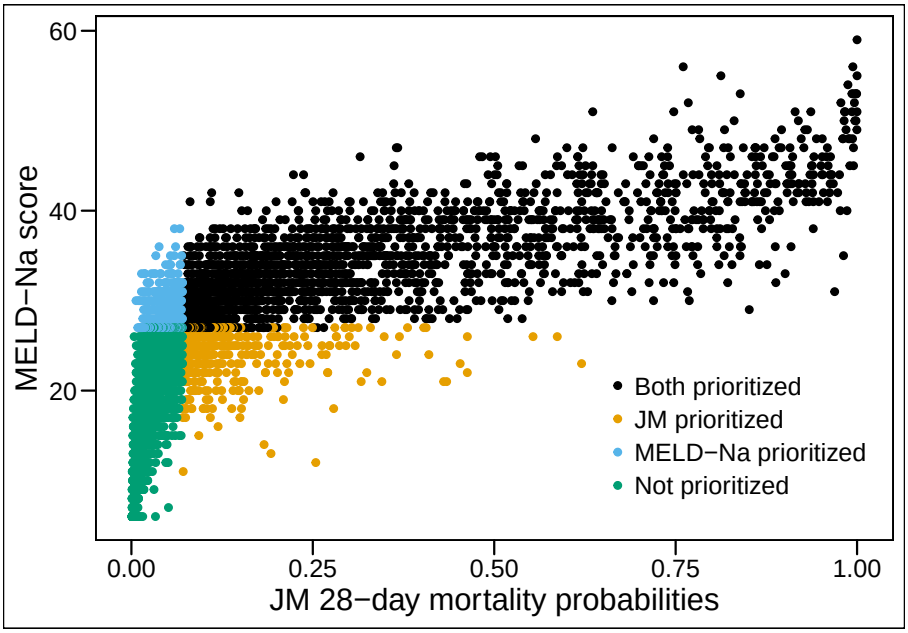


Figure 5.3: The correlation plot of MELD-Na score and 28-day ACLF-JM survival predictions. Patients are stratified in 4 groups: orange and blue patients would have been prioritized differently under either the ACLF-JM or MELD-Na. Blue patients had a 4x higher 28-day waiting list mortality than orange patients.

Table 5.3: Characteristics of patients prioritized differently for liver transplantation within 28 days

Characteristics	Both prioritized	ACLF-JM prioritized	MELD-Na prioritized	Not prioritized	p*
n	2186	450	450	6990	
Age (median [IQR])	56.0 [47.0, 62.0]	62.0 [55.0, 67.0]	50.0 [42.0, 56.8]	59.0 [52.0, 64.0]	<0.001
Male sex (%)	1336 (61.1)	175 (38.9)	326 (72.4)	4552 (65.1)	<0.001
Death within 28 days (%)	289 (13.2)	59 (13.1)	14 (3.1)	90 (1.3)	<0.001
ACLF (%)					
No ACLF	172 (7.9)	191 (42.4)	162 (36.0)	6155 (88.1)	
ACLF-1	585 (26.8)	95 (21.1)	248 (55.1)	720 (10.3)	
ACLF-2	792 (36.2)	91 (20.2)	39 (8.7)	105 (1.5)	
ACLF-3	637 (29.1)	73 (16.2)	1 (0.2)	10 (0.1)	
Disease (%)					
Cirrhosis HCV	165 (7.5)	31 (6.9)	39 (8.7)	1099 (15.7)	
NASH	392 (17.9)	147 (32.7)	61 (13.6)	1479 (21.2)	
Cirrhosis Alcoholic	964 (44.1)	130 (28.9)	235 (52.2)	1768 (25.3)	
Cirrhosis Other	416 (19.0)	68 (15.1)	55 (12.2)	988 (14.1)	
Cholestatic disease	104 (4.8)	41 (9.1)	36 (8.0)	638 (9.1)	
Metabolic disease	56 (2.6)	5 (1.1)	13 (2.9)	135 (1.9)	
Malignant/benign tumor	39 (1.8)	11 (2.4)	7 (1.6)	734 (10.5)	
Other	50 (2.3)	17 (3.8)	4 (0.9)	149 (2.1)	
MELD (median [IQR])	34.0 [29.0, 39.0]	24.0 [21.0, 28.0]	26.0 [23.0, 29.0]	15.0 [11.0, 19.0]	<0.001
MELD-Na (median [IQR])	33.0 [30.0, 38.0]	25.0 [23.0, 26.0]	28.0 [27.0, 30.0]	15.0 [10.0, 20.0]	<0.001
Life support dependent	291 (13.3)	84 (18.7)	3 (0.7)	50 (0.7)	<0.001

Note:

Clarification: JM-prioritized patients are not prioritized by MELD-Na, and vice versa.

* Difference tested between ACLF-JM-prioritized and MELD-Na-prioritized patients

Clinical application of the ACLF-JM

After constructing and validating the ACLF-JM in this large cohort, an online application was developed, which allows clinicians to easily calculate individual patient survival probabilities based on the ACLF-JM. Available at: https://predictionmodels.shinyapps.io/aclf-jm/. Excel files with repeated MELD-Na measurements can be uploaded into this tool, to generate dynamic survival predictions during follow-up. The ACLF-JM simulates individual patient data to calculate personalized predictions. See supplement 1 for precise instructions for the data upload and supplement 2 for a step-by-step manual.

Discussion

In this study, we developed and validated the ACLF-JM prediction model, to estimate survival of ACLF patients. We report several important findings. First, both current and past disease severity and its rate of change are strongly associated with survival in ACLF. Second, by using these data, the ACLF-JM gives excellent prediction performance, even in ACLF-3, and significantly outperforms MELD-Na. Third, the ACLF-JM could have prioritized patients with low median MELD-Na scores, i.e. not identified by MELD-Na, but four times higher mortality rates than MELD-Na prioritized patients. Fourth, the ACLF-JM can be clinically applied online to estimate and visualize patient-specific survival, which can be updated with every new measurement.

Disease development over time

ACLF disease severity is dynamic and can change rapidly. During the first week, disease severity changes for most patients, resulting in different survival outcomes.[9,10] The current liver allocation system does not consider change, as it uses only the most recent measurement for survival prediction and ignores previous data. Moreover, survival is estimated based on the MELD-Na score, which ignores relevant factors for ACLF and therefore underestimates mortality.[7,8] Hernaez et al. showed that mortality was higher than expected in low MELD-Na score patients. They also showed that, despite their high(er) ACLF grade, these low MELD-Na patients were often not considered for LT.[7]

Interestingly, Hernaez et al. mentioned that "Future research should also focus on developing and validating prognostic scores that incorporate dynamic changes in patients clinical course," i.e. the goal of this study. Sundaram et al. showed that ACLF death and removal rate did not correlate well with the MELD-Na score, as mortality rates were highest in ACLF-3 patients with MELD-Na <25.[8] In this study, ACLF was present in 33.3% of the patients. Therefore, the MELD-Na underestimation of ACLF disease severity could be substantial, which possibly leads to unequal treatment access and surplus mortality.[7] Therefore, the ACLF-JM was developed to predict ACLF patient survival based on disease development over time. The model provides several important improvements over the MELD-Na (table S4).[21] Most importantly, predictions are based on all available previous data and update for every new measurement.[22] Predictions should be update based on accumulating evidence, because ACLF is a dynamic disease. Also, the ACLF-JM considers both the value of disease severity and the rate at which disease severity is changing, see Figure 5.1. It uses more nuanced aspects of ACLF disease development to predict survival. Thus, like a clinician, past and current disease developments are used to estimate patient prognosis. Updating prognosis is important in ACLF, as disease can increase fast and non-linearly (e.g. exponential).[1,3] ACLF-JM survival predictions could therefore be used to aid clinical decision making for ACLF patients on the waiting list for LT, as current models result in unequal transplantation access and post-LT survival rates.[8,10,17] Furthermore, In this cohort, we showed that ACLF-JM prioritization identified patients with low MELD-Na scores, but high mortality, see Table 5.3. Mortality is underestimated in these patients and subsequently they receive a lower priority for LT. Since ACLF patients benefit from fast LT,[17] use of the ACLF-JM for the evaluation of prognosis could perhaps help to resolve the underestimation of waiting list mortality for ACLF patients.[7]

ACLF-JM validation

The ACLF-JM showed excellent short-term survival prediction performance at baseline and with increasing follow-up. Increasing ACLF grade did not lead to a decrease in predictive accuracy. This is important, because risk of death and need for LT should be reliably estimated in the sickest patients. Our data showed that both the ACLF-JM

and MELD-Na AUCs declined with increasing follow-up. This is likely due to population changes, i.e. the sickest patients die or are transplanted first and less patients remain with increasing follow-up.[23] Also, with increasing disease severity, generally a shorter follow-up period is available. The ACLF-JM approximation of disease does not depend on the number of measurements per patient, because it estimates disease over time as a continuous trajectory (figure S4). This is important, because frequency of measurement confounded previous (Cox-based) survival predictions for patients in need of LT.[24] The ACLF-JM performed comparable and sometimes even better compared to the reported performance of the CLIF-C OF score.[6] This could possibly indicate that ACLF-JM performance was adequate enough for clinical application. Because the UNOS registry does not contain data on white blood cell counts, CLIF-C ACLF scoring was not possible in this study. ACLF-JM performance could however be externally validated in the cohorts used to construct the CLIF-C scores.[6]

Clinical application of the ACLF-JM

After training and ascertaining excellent performance, an online tool of the ACLF-JM was created for clinical use. Especially in ACLF, both the patient and treating clinician benefit from patient-specific modelling, which shifts the focus of prediction from the population to the individual patient level. Jalan et al. already stated that there is a need for models that "update on a daily basis providing additional prognostic information," and that "currently, no validated evidence-based tools guide the decision-making."[6] The ACLF-JM meets these demands and more, with excellent performance leading to personalized prediction, readily available online for any clinician.

Limitations

A limitation is that longitudinal MELD-Na measurements are not best to model ACLF disease development, as they can underestimate ACLF disease severity.[7] Ideally, longitudinal CLIF-C ACLFs data would be available in the UNOS data. However, MELD-Na was one of the few consistently available longitudinal measurements, which allowed analysis on a large scale and comparison to previous studies. The retrospec-

tive analysis of large databases also has several disadvantages. Misclassification of disease severity could give bias, e.g. subjective scoring of ascites and encephalopathy. Also, surrogate markers, suggested by authors of other large UNOS ACLF analyses, were used for ventilatory and circulatory failure.[6,8,10,17] For example, mechanical ventilation was used as replacement for respiratory failure, it is however very well possible that a patient with respiratory failure did not receive mechanical ventilation, or vice versa. Despite these shortcomings, the ACLF-JM showed excellent performance with increasing disease severity (ACLF grade).

Conclusion

ACLF survival is dynamically predicted by the ACLF-JM prediction model, using both longitudinal and survival data. Updating prognosis on new measurements is important, as ACLF is a dynamic disease. The ACLF-JM prediction performance was excellent in this cohort, even in ACLF-3 patients. The ACLF-JM could therefore be used as a tool for the personalized evaluation of prognosis and clinical decision making in patients with ACLF.

References

1. Jalan R, Gines P, Olson JC, et al. Acute-on chronic liver failure. J Hepatol. 2012;57(6):1336-1348. doi:10.1016/j.jhep.2012.06.026
2. Hernaez R, Kramer JR, Liu Y, et al. Prevalence and short-term mortality of acute-on-chronic liver failure: A national cohort study from the USA. J Hepatol. 2019;70(4):639-647. doi:10.1016/j.jhep.2018.12.018
3. Moreau R, Jalan R, Gines P, et al. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. Gastroenterology. 2013;144(7):1426-1437.e9. doi:10.1053/j.gastro.2013.02.042
4. 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(10):1018-1026. doi:10.1007/s11250-017-1262-3
5. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Validation of the Model for End-stage Liver Disease sodium (MELD-Na) score in the Eurotransplant region. Am J Transplant. 2020. doi:10.1111/ajt.16142
6. Jalan R, Saliba F, Pavesi M, et al. Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. J Hepatol. 2014;61(5):1038-1047. doi:10.1016/j.jhep.2014.06.012
7. Hernaez R, Liu Y, Kramer JR, Rana A, El-Serag HB, Kanwal F. Model for end-stage liver disease-sodium underestimates 90-day mortality risk in patients with acute-on-chronic liver failure. J Hepatol. 2020:1-9. doi:10.1016/j.jhep.2020.06.005
8. Sundaram V, Jalan R, Wu T, et al. Factors Associated with Survival of Patients With Severe Acute-On-Chronic Liver Failure Before and After Liver Transplantation. Gastroenterology. 2019;156(5):1381-1391.e3. doi:10.1053/j.gastro.2018.12.007
9. Gustot T, Fernandez J, Garcia E, et al. Clinical Course of acute-on-chronic liver failure syndrome and effects on prognosis. Hepatology. 2015;62(1):243-252. doi:10.1002/hep.27849
10. Sundaram V, Kogachi S, Wong RJ, et al. Effect of the clinical course of acute-on-chronic liver failure prior to liver transplantation on post-transplant survival. J Hepatol. 2020;72(3):481-488. doi:10.1016/j.jhep.2019.10.013

11. Rizopoulos R. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. 1st ed. Chapman and Hall/CRC; 2012.
12. Papageorgiou G, Mokhles MM, Takkenberg JJM, Rizopoulos D. Individualized dynamic prediction of survival with the presence of intermediate events. Stat Med. 2019;38(30):5623-5640. doi: 10.1002/sim.8387
13. Arisido MW, Antolini L, Bernasconi DP, Valsecchi MG, Rebora P. Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. BMC Med Res Methodol. 2019;19(1):1-13. doi:10.1186/s12874-019-0873-y
14. Campbell KR, Juarez-Colunga E, Grunwald GK, Cooper J, Davis S, Gralla J. Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: Application to kidney transplantation. BMC Med Res Methodol. 2019;19(1):1-12. doi:10.1186/s12874-019-0773-1
15. Hernaez R, Liu Y, Kramer JR, Rana A, El-serag HB, Kanwal F. Model for end-stage liver disease-sodium underestimates 90-day mortality risk in patients with acute-on-chronic liver failuare. J Hepatol. 2020. doi:10.1016/j.jhep.2020.06.005
16. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73. doi:10.7326/M14-0698
17. Thuluvath PJ, Thuluvath AJ, Hanish S, Savva Y. Liver transplantation in patients with multiple organ failures: Feasibility and outcomes. J Hepatol. 2018;69(5):1047-1056. doi:10.1016/j.jhep.2018.07.007
18. Rizopoulos D. The R package jmbayes for fitting joint models for longitudinal and time-to-event data using MCMC. J Stat Softw. 2016;72(7). doi:10.18637/jss.v072.i07
19. Gerds TA, Blanche P, Mortensen R, Tollenaar N, Brasch Mogensen U, Ozenne B. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks. 2020. https://cran.r-project.org/web/packages/riskRegression/index.html.
20. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration:

The Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7

21. Goudsmit BFJ, Tushuizen ME, Putter H, Braat AE, van Hoek B. The role of the model for end-stage liver disease-sodium score and joint models for 90-day mortality prediction in patients with acute-on-chronic liver failure. J Hepatol. 2021;74(2):475-476. doi: 10.1016/j.jhep.2020.08.032
22. Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes. Annu Rev Stat Its Appl. 2019;6(1):223-240. doi:10.1146/annurev-statistics-030718-105048
23. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Refitting the model for end-stage liver disease for the Eurotransplant region. Hepatology. 2020;in press.
24. Bambha K, Kim WR, Kremers WK, et al. Predicting survival among patients listed for liver transplantation: An assessment of serial MELD measurements. Am J Transplant. 2004;4(11):1798-1804. doi:10.1111/j.1600-6143.2004.00550.x

Part III: Survival with and without transplantation

"What can be controlled is never completely real;

what is real can never be completely controlled."

— Vladimir Nabokov

Chapter 6

Survival benefit from liver transplantation for patients with and without hepatocellular carcinoma

Goudsmit BFJ, Prosepe I, Tushuizen ME, et al. Survival benefit from liver transplantation for patients with and without hepatocellular carcinoma. *Under review.*

Background & Aims: In the US, inequal liver transplantation (LT) access exists between patients with and without hepatocellular carcinoma (HCC). Survival benefit considers survival without and with LT and could equalize LT access. We calculated and compared LT survival benefit scores for patients with(out) HCC, based on longitudinal data in a recent US cohort.

Methods: Adult LT candidates with(out) HCC between 2010-2019 were included. Waitlist survival over time was contrasted to posttransplant survival, to estimate 5-year survival benefit from the moment of LT. Waitlist survival was modeled with bias-corrected time-dependent Cox regression and posttransplant survival was estimated through Cox proportional hazards regression.

Results: Mean HCC survival without LT was always lower than non-HCC waitlist survival. Below MELD(-Na) 30, HCC patients gained more life-years from LT than non-HCC patients at the same MELD(-Na) score. Only non-HCC patients below MELD(-Na) 9 had negative benefit. Most HCC patients were transplanted below MELD(-Na) 14 and most non-HCC patients above MELD(-Na) 26. Liver function (MELD(-Na), albumin) was the main predictor of 5-year benefit. Therefore, during five years, most HCC patients gained 0.12 to 1.96 years from LT, whereas most non-HCC patients gained 2.48 to 3.45 years.

Conclusion: On an individual level, transplanting patients with HCC resulted in survival benefit. However, on a population level, benefit was indirectly wasted, as non-HCC patients were likely to gain more survival due to decreased liver function. Based on these data, we now provide an online calculator to estimate 5-year survival benefit given specific patient characteristics. Survival benefit scores could serve to equalize LT access.

Introduction

Adult liver transplantation (LT) relies on scarce donor grafts. Therefore, allocation prioritizes patients that likely will die soon without transplantation.[1] For most patients on the LT waiting list in the United States (US), the Model for End-stage Liver Disease sodium (MELD-Na) score adequately predicts expected survival without transplantation.[2,3]

However, MELD-Na is less predictive of survival for transplant candidates with hepatocellular carcinoma (HCC). This is because HCC mortality is typically caused by tumor progression and not by liver failure.[4] The number of HCC patients listed for transplantation has tripled the past 10 years.[5] HCC is the single most important cause of death in cirrhotic patients, and treatment through LT still has the best long-term results.[6–8] The exception point system was developed to compensate liver graft allocation based on inadequate MELD(-Na) survival prediction for most notably HCC patients. In this system, HCC patients receive artificial MELD points that increase automatically every 90 days, to mimic HCC progression.[9–11] Unintendedly, the exception points created inequity between non-HCC and HCC patients, because HCC LT access increased too much,[12–14] and inequity among HCC patients, because all patients within one region receive the same priority with only waiting time as tiebreaker.[15]

Instead of arbitrary points, patient characteristics should be used to model the risk of waiting list dropout.[10,14,16–18] Moreover, to balance the principles urgency and utility,[19] the risk of waiting list dropout should be compared to expected post-transplant survival. The difference is survival benefit, or the life-years gained from transplantation.[20] Considering LT survival benefit is valuable because donor grafts are scarce and some patients gain more life-years than others.[20–23]

Benefit of non-HCC and HCC patients has been previously evaluated,[13,14,16,17] but re-evaluation is warranted. Firstly, because contradicting findings have been reported. Most notably, Vitale et al. showed that HCC patients in Italy benefited twice as much from LT compared to non-HCC patients,[17] whereas Berry et al. stated that US HCC patients derived negative or little benefit.[13] Secondly, previous work defined survival benefit as the difference between post-transplant survival and waiting list survival counted from first registration. We hypothesized that counting waiting list survival from