Refitting the Model for End-Stage Liver Disease for the Eurotransplant Region

Ben F. J. Goudsmit [iD] ,[1-3] Hein Putter [iD] ,[4] Maarten E. Tushuizen [iD] ,[2] Serge Vogelaar [iD] ,[3] Jacques Pirenne [iD] ,[5**]
Ian P. J. Alwayn [iD] ,[1] Bart van Hoek [iD] ,[2*] and Andries E. Braat [iD] [1*]

BACKGROUND AND 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.

APPROACH AND RESULTS: All adult patients listed for a first liver transplantation between January 1, 2007, and December 31, 2018, were included. Data were randomly split in a training set (70%) and a validation set (30%). In the training data, generalized additive models 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. A total of 6,684 patients were included. Based on training data, refit parameters were capped at creatinine 0.7-2.5, bilirubin 0.3-27, international normalized ratio 0.1-2.6, and sodium 120-139. 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 compared to UNOS-MELD.

CONCLUSIONS: Refitting MELD resulted in new lower and upper bounds for each parameter. The predictive power of reMELD-Na was significantly higher than UNOS-MELD. ReMELD prioritized patients with higher 90-day mortality rates. Thus, reMELD(-Na) should replace UNOS-MELD for liver graft allocation in the Eurotransplant region. (HEPATOLOGY 2021;74:351-363).

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 for 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] Since 2016, the United Network for Organ Sharing (UNOS) regions also added serum sodium through the MELD-Na score,[3] but the Eurotransplant region remains MELD-based. The MELD was weighed, i.e., the relative importance of each parameter, based on a cohort from 1991 to 1995.[4] For clinical use, the lower

boundaries for the parameters were set to 1, to prevent negative MELD scores after natural logarithm transformation. Creatinine levels were capped at 4 mg/dL for patients not receiving dialysis. According to some of the proposers of MELD, these boundaries were "based entirely on the clinical intuition of the policymaking body when the MELD score was implemented."[5] 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."[6]

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 the UNOS-MELD has declined significantly in recent years.[7] However, an update of the MELD coefficients in UNOS data showed that performance could still be further improved.[5] As the Eurotransplant population differs from the original MELD cohort,[4,8] improvement of the Eurotransplant liver allocation is very 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 Eurotransplant patients as it was fit on the UNOS population. This could diminish MELD's 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 (reMELD) score for the Eurotransplant region by reweighing the MELD coefficients and reevaluating 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 Na levels at listing in a Eurotransplant refit MELD-Na (reMELD-Na) score will be evaluated. The performance of the constructed refit Eurotransplant models will be compared to the UNOS-MELD because this model is still being used at the basis of liver allocation in our region.

Patients and Methods

PATIENT DATA

The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement was used to report the development of the multivariate prediction models in this study.[9] Informed consent was waived upon IRB approval since this was a minimal risk study. Data were requested from the Eurotransplant Database. All adult patients actively listed for a first LT between January 1, 2007, and December 31, 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 ([N]SE) points, a high urgency (HU) status (i.e., UNOS status 1), living donor grafts, or multiorgan transplantations (other than kidney).[10] Patient data were 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 for both actively listed

ARTICLE INFORMATION:

From the [1]Division of Transplantation, Department of Surgery, Leiden University Medical Centre, ZA Leiden, The Netherlands; [2]Department of Gastroenterology and Hepatology, Leiden University Medical Centre, ZA Leiden, The Netherlands; [3]Eurotransplant International Foundation, KG Leiden, The Netherlands; [4]Department of Biomedical Data Sciences, Leiden University Medical Centre, ZA Leiden, The Netherlands; [5]Department of Abdominal Transplant Surgery, University Hospitals Leuven, Leuven, Belgium.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Ben F. J. Goudsmit, M.D.
Division of Transplantation, Department of Surgery
Leiden University Medical Centre
Albinusdreef 2

2333 ZA Leiden, The Netherlands
E-mail: b.f.j.goudsmit@lumc.nl
Tel.: +0031 (0) 71 526 9111

and removed patients. Thus, removed patients who died within 90 days after listing were also considered deceased. 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 December 31, 2018, or after receiving NSE points or an HU status during active waiting. The sample size for this study was set by the retrospective design. Missing data (in <0.01%) were not imputed.

STATISTICAL METHODS

The data were randomly split into a training set (70%) and a validation set (30%). For each recipient, the UNOS-MELD and MELD-Na scores at first active listing were calculated.[11,12] Then, the Eurotransplant reMELD score was constructed in the training data. For each MELD parameter, a multivariate Cox generalized additive 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 the 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 bounds 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 cutoffs. 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 above-mentioned 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 reMELD. To ensure equal distributions of the traditional UNOS-MELD and Eurotransplant reMELD scores in our data, the 25th and 75th quantiles were matched. Also, reMELD scores < 6 and >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 concentrations. Interactions between Na and each reMELD parameter were assessed and deemed relevant if $P < 0.01$. Thus, the final reMELD-Na model comprised of reMELD parameters, newly bound Na, and relevant interactions between the terms. Again, the 25th and 75th quantiles were matched, and the final scores of the reMELD-Na were set between 6 and 40. Finally, the refit Eurotransplant models were compared with likelihood ratio tests to the 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.[13] 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 the constructed reMELD-Na was investigated by plotting both scores. To assess whether the reMELD-Na would give more effective sickest-first allocation, survival estimates were calculated for patients prioritized by the UNOS-MELD and reMELD-Na. All statistical analyses were performed using R, v3.6.1 (RStudio, Inc., Boston, MA).

Results

In this study, 6,944 patients were included (Table 1). More male (68%) than female patients were included, and alcohol-associated cirrhosis was the most frequent cause of liver disease. The median UNOS-MELD and serum sodium at listing were 14 (interquartile range [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 (2.5% while actively listed

TABLE 1. Characteristics of Training and Validation Data

	Training Set (n = 4,860)	Validation Set (n = 2,084)	P
Age (median [IQR])	56 (49-62)	55 (49-62)	0.022
Gender female (%)	1563 (32.2)	659 (31.6)	0.680
Disease (%)			0.089
Cirrhosis, alcohol-associated	1,361 (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			0.508
Censored because of HU or (N)SE	1171 (24.2)	476 (22.9)	
Deceased			
While actively listed	115 (2.4)	59 (2.8)	
After removal but within 90 days	337 (7.0)	167 (8.0)	
Removed from the WL	624 (12.8)	257 (12.3)	
Still waiting on WL	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 (mg/dL)	1.40 (3.73)	1.46 (4.16)	0.563
Bilirubin (mg/dL)	5.74 (8.79)	5.84 (9.34)	0.669
INR	1.51 (0.72)	1.52 (0.72)	0.510
Sodium (mmol/L)	137.02 (4.99)	136.94 (4.88)	0.526
UNOS-MELD at listing (median [IQR])	14 (10-20)	14 (10-20)	

and 7.3% after removal but within 90 days of first listing). There were no relevant differences between the training and validation data.

MODEL DEVELOPMENT

The GAM plots for each parameter are shown in Fig. 1. For creatinine, the S-shaped curve displayed clear lower and upper boundaries; 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 <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 lowered. Therefore, the influence of renal failure on the chances for LT was reduced.

For bilirubin, the lower bound was found at 0.3 and the upper at 27 mg/dL. Varying 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 more relevant value, roughly between 20 and 40, without affecting the optimal fit to the data too much (Supporting Fig. S1).

The INR had no lower bound and was capped at a maximum of 2.6. However, assessment of the log-likelihood values showed that a range between 0.1 and 1.0 would be acceptable as the lower bound (Supporting Fig. S3) 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 the INR, of which the weight was increased in the refit models.

Overall, the reMELD and reMELD-Na models capped fewer patients at assumed values than the UNOS-MELD.

In Fig. 2, lines were plotted 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 included in the Supporting Information. 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 $7.728*\ln(\text{creatinine}) + 3.446*\ln(\text{bilirubin}) + 10.597*\ln(\text{INR}) + 8.422$. In this equation the above-mentioned 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 $9.025*\ln(\text{creatinine}) + 2.969*\ln(\text{bilirubin}) + 9.518*\ln(\text{INR}) - 0.392*(139-\text{Na}) - 0.351*\ln(139-\text{Na})*\ln(\text{creatinine})$. For the parameters in the reMELD-Na score, the above-mentioned boundaries were used. Compared to the UNOS-MELD, the reMELD and reMELD-Na used, respectively, 149% (n = 4,815) and 42% (n = 2,748) more patient measurements, i.e., fewer
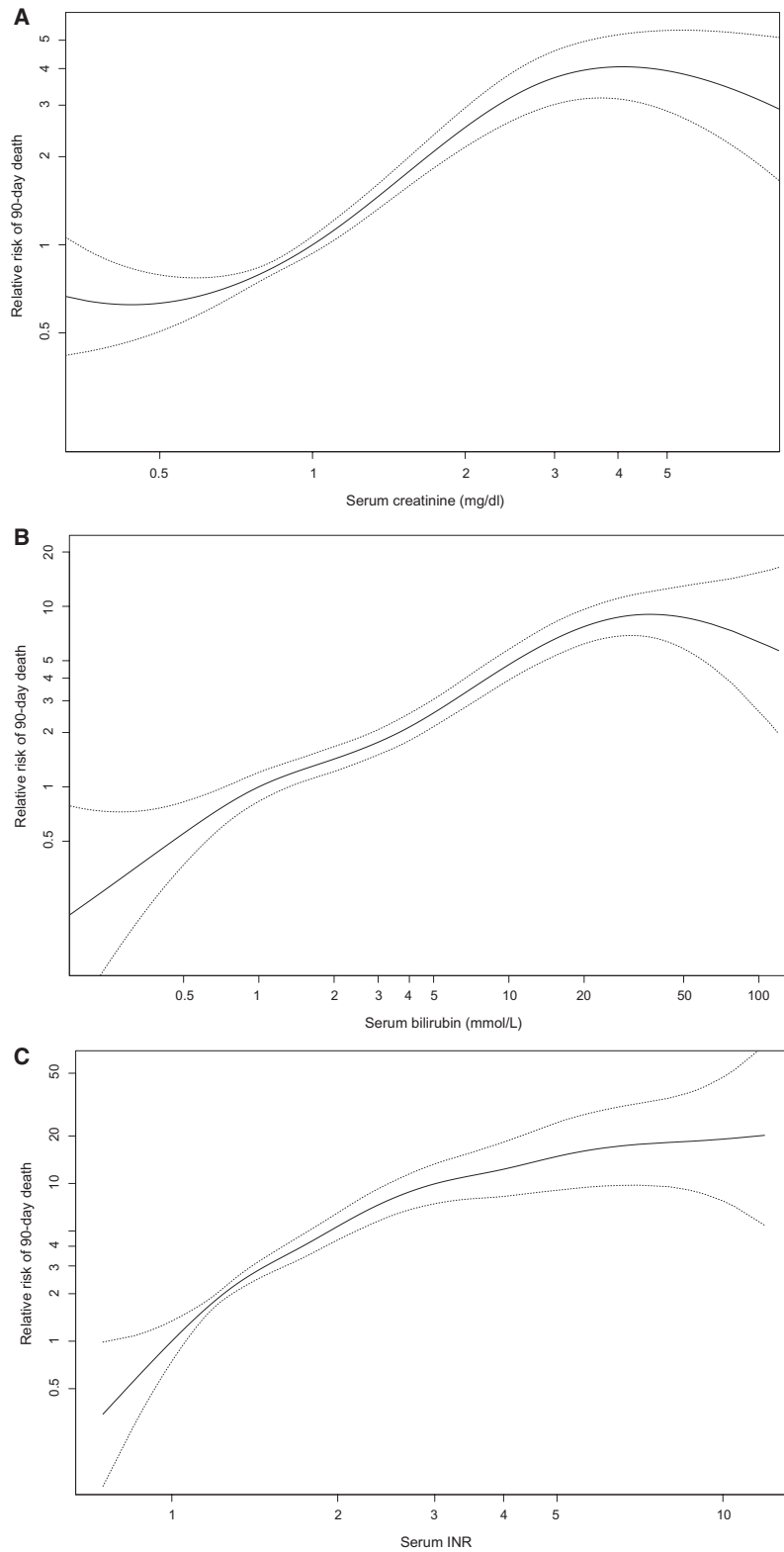
FIG. 1. GAM with splines for creatinine, bilirubin, and INR based on the training data. Each line represents the relation of the studied parameter with 90-day mortality, corrected for the other MELD parameters (creatinine is corrected for bilirubin and INR, and so on). (A) Relation between serum creatinine concentration and 90-day mortality. (B) Relation between serum bilirubin concentration and 90-day mortality. (C) Relation between serum INR and 90-day mortality.
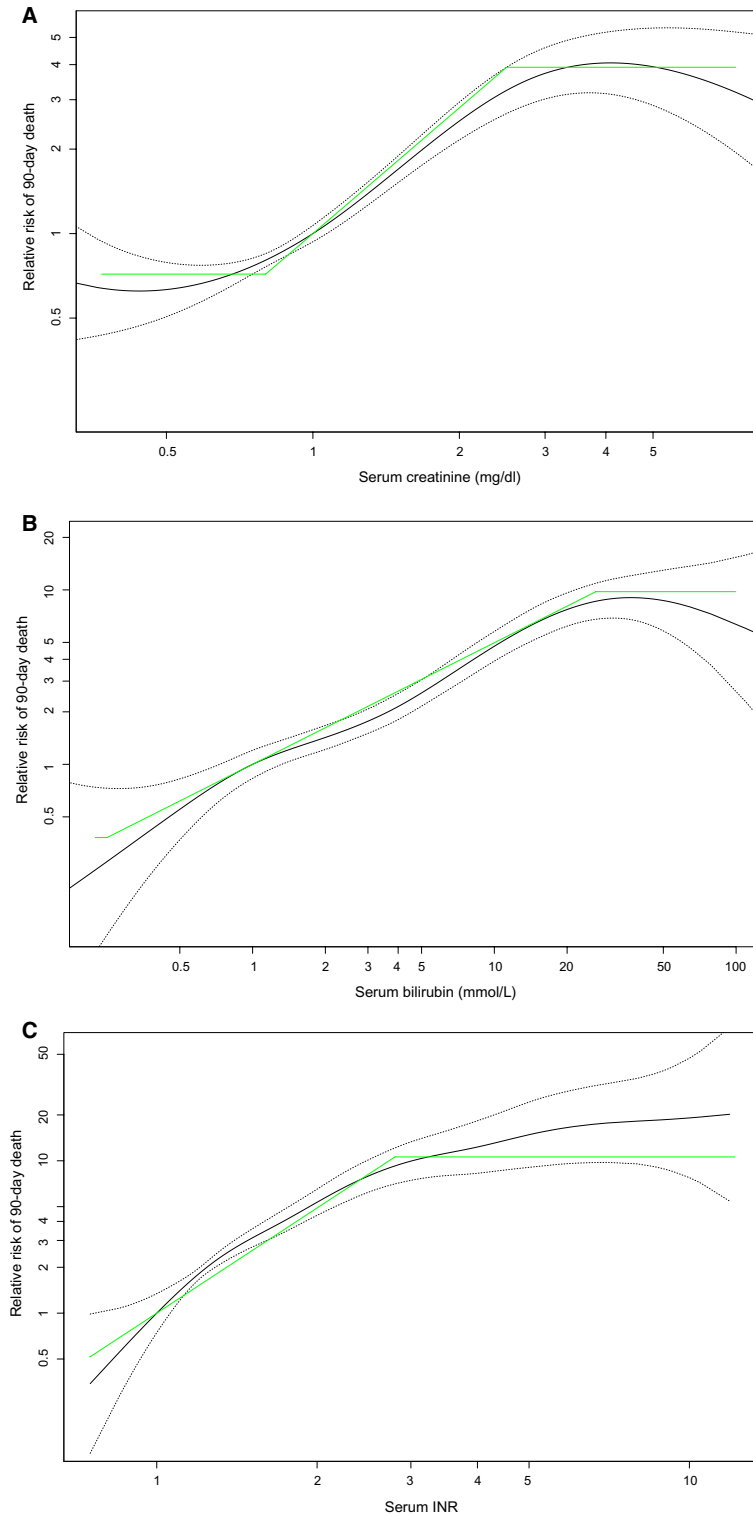
FIG. 2. Over the GAM with splines, lines (green) were plotted that represent the coefficient (slope of the diagonal) and lower and upper boundaries (horizontal segments) of the refit parameters. (A) For creatinine, straight lines represent the coefficient (slope of the diagonal) and lower and upper boundaries (horizontal segments) in refit MELD. (B) For bilirubin, straight lines represent the coefficient (slope of the diagonal) and lower and upper boundaries (horizontal segments) in refit MELD. (C) For the INR, straight lines represent the coefficient (slope of the diagonal) and lower and upper boundaries (horizontal segments) in refit MELD.

TABLE 2. Number of Patient Measurements Included in UNOS and Refit Models

		UNOS-MELD	Patients Capped (%)	Included Patients (%)	ReMELD(-Na)	Patients Capped (%)	Included Patients (%)
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		139	43	

For each parameter the lower and upper bounds are shown. "Patients capped" shows the percentage of the cohort that either lies under or above the chosen bounds. "Patients included" shows the percentage of patients whose measurements are included in the model. The total number of patients included per model is 1,933, 4,815, and 2,748 for the UNOS-MELD, reMELD, and reMELD-Na, respectively.

true patient measurements were capped, at listing with the boundaries, as shown in Table 2.

MODEL PERFORMANCE

Figure 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 the reMELD showed a better fit to the data for all three parameters. The calculated chi-squared values confirmed significant ($P < 0.001$) improvements in the refit models compared to the UNOS-MELD (Table 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 the 0.849 of the UNOS-MELD (Table 3). Furthermore, the refit models showed an 8% reduction in prediction error compared to the UNOS-MELD, with Brier scores of 0.053 (reMELD[-Na]) and 0.057 (UNOS-MELD). Compared to the UNOS-MELD-Na Brier score of 0.056, the refit models further reduced prediction errors by 5%.

IMPACT ON THE WL

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 4 shows the characteristics of these differently

prioritized patients. 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 (Fig. 4) illustrates which patients would be prioritized according to either UNOS-MELD or reMELD-Na allocation. The patients in the top left quadrant would have been prioritized by reMELD-Na allocation but not by the UNOS-MELD. They had estimated 90-day survival probabilities of 52.4% (95% CI 41.3–66.5) compared to 70.0% (95% CI 58.9-83.1) for patients prioritized by the UNOS-MELD but not by the reMELD-Na (bottom right quadrant). Thus, the reMELD-Na could have prioritized patients with a 90-day WL mortality hazard ratio of 1.6 compared to currently prioritized patients. Figure 4 also illustrates that after refitting, no scores >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, the 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, the MELD score was refitted to the Eurotransplant data. By establishing new and evidence-based lower and upper bounds for each
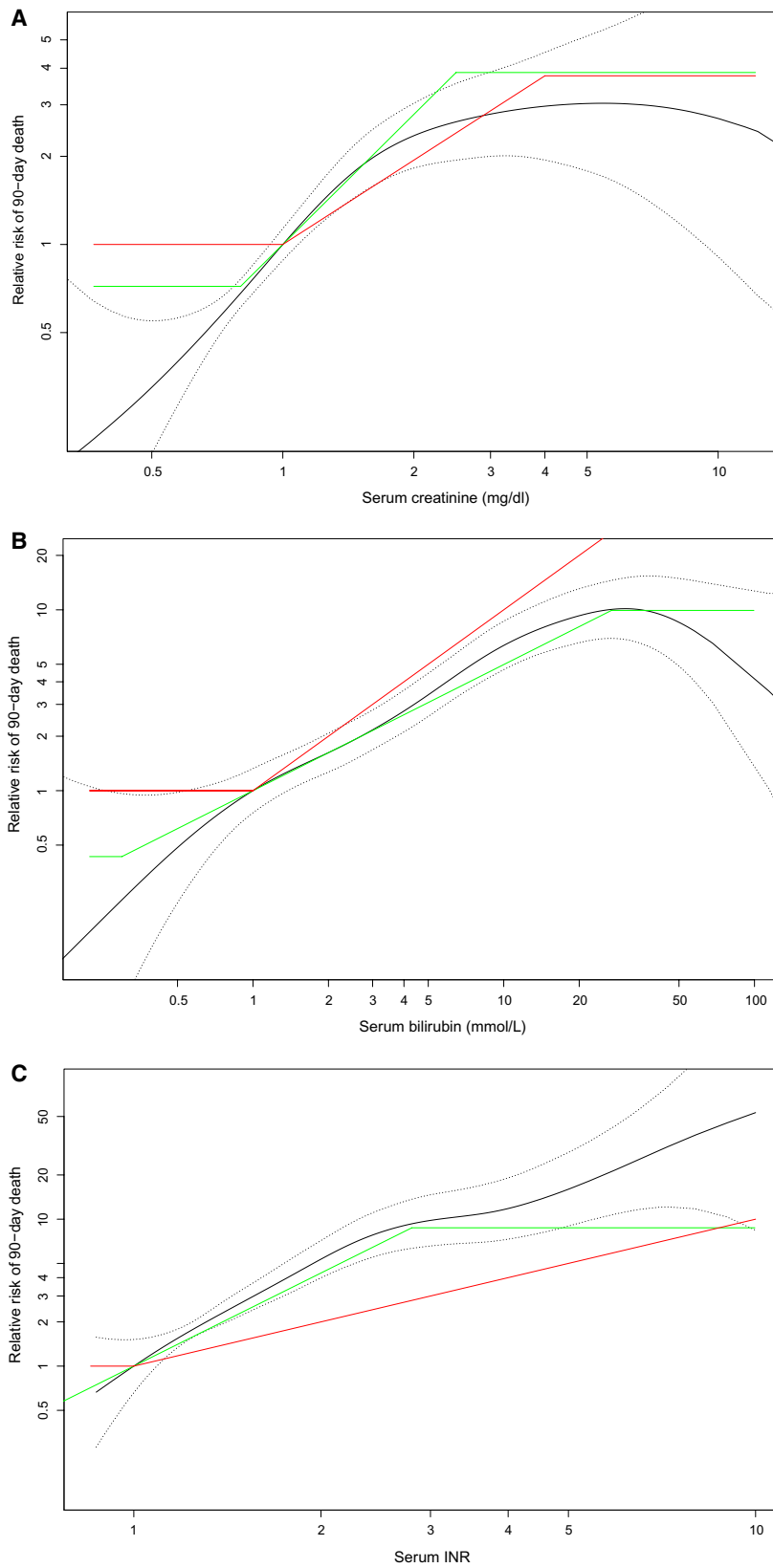
FIG. 3. In the validation data, the GAM with splines for each parameter is shown. The coefficients and boundaries of reMELD (green) and UNOS-MELD (red) were plotted. (A) In the validation data, the creatinine relation with 90-day mortality is shown. The coefficients and boundaries of creatinine in reMELD (green) and UNOS-MELD (red) were plotted to illustrate model fit. (B) In the validation data, the bilirubin relation with 90-day mortality is shown. The coefficients and boundaries of bilirubin in reMELD (green) and UNOS-MELD (red) were plotted to illustrate model fit. (C) In the validation data, the INR relation with 90-day mortality is shown. The coefficients and boundaries of the INR in reMELD (green) and UNOS-MELD (red) were plotted to illustrate model fit.

TABLE 3. Comparison of Models in Validation Data

	C-Index	Max Log-Likelihood	Chi-Squared	P
UNOS-MELD	0.849 (SE = 0.012)	−1376.6		
UNOS-MELD-Na	0.860 (SE = 0.010)	−1362.8	+27.660	<0.001
reMELD	0.866 (SE = 0.011)	−1347.1	+58.966	<0.001
reMELD-Na	0.869 (SE = 0.010)	−1347.1	+59.066	<0.001

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.

TABLE 4. Characteristics of Prioritized Patients

	Transplanted Both	UNOS-MELD Transplanted	reMELD-Na Transplanted	Not Transplanted	P
n	1,105	143	143	5,553	
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)	1,750 (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 (%)					<0.001
Cirrhosis, alcohol associated	390 (35.3)	48 (33.6)	65 (45.5)	1,458 (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)	1,325 (23.9)	
Other	207 (18.7)	44 (30.7)	19 (13.2)	739 (13.3)	
Status after 90 days					<0.001
Censored because of HU or NSE	52 (4.7)	9 (6.3)	8 (5.6)	1,578 (28.5)	
Deceased	338 (30.7)	28 (19.6)	36 (25.2)	276 (5.0)	
After removal but within 90 days	121 (11.0)	30 (21.0)	27 (18.9)	703 (12.7)	
While waiting on WL	56 (5.1)	19 (13.3)	28 (19.6)	2,370 (42.8)	
Transplanted	536 (48.6)	57 (39.9)	44 (30.8)	611 (11.0)	
Days on WL (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 (mg/dL)	2.95 (8.51)	2.67 (9.43)	1.26 (0.48)	1.09 (1.18)	<0.001
Bilirubin (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 (mmol/L)	134.26 (6.08)	138.21 (4.67)	127.34 (5.34)	137.76 (4.20)	<0.001
(re)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

Correlation plot UNOS MELD and reMELD-Na
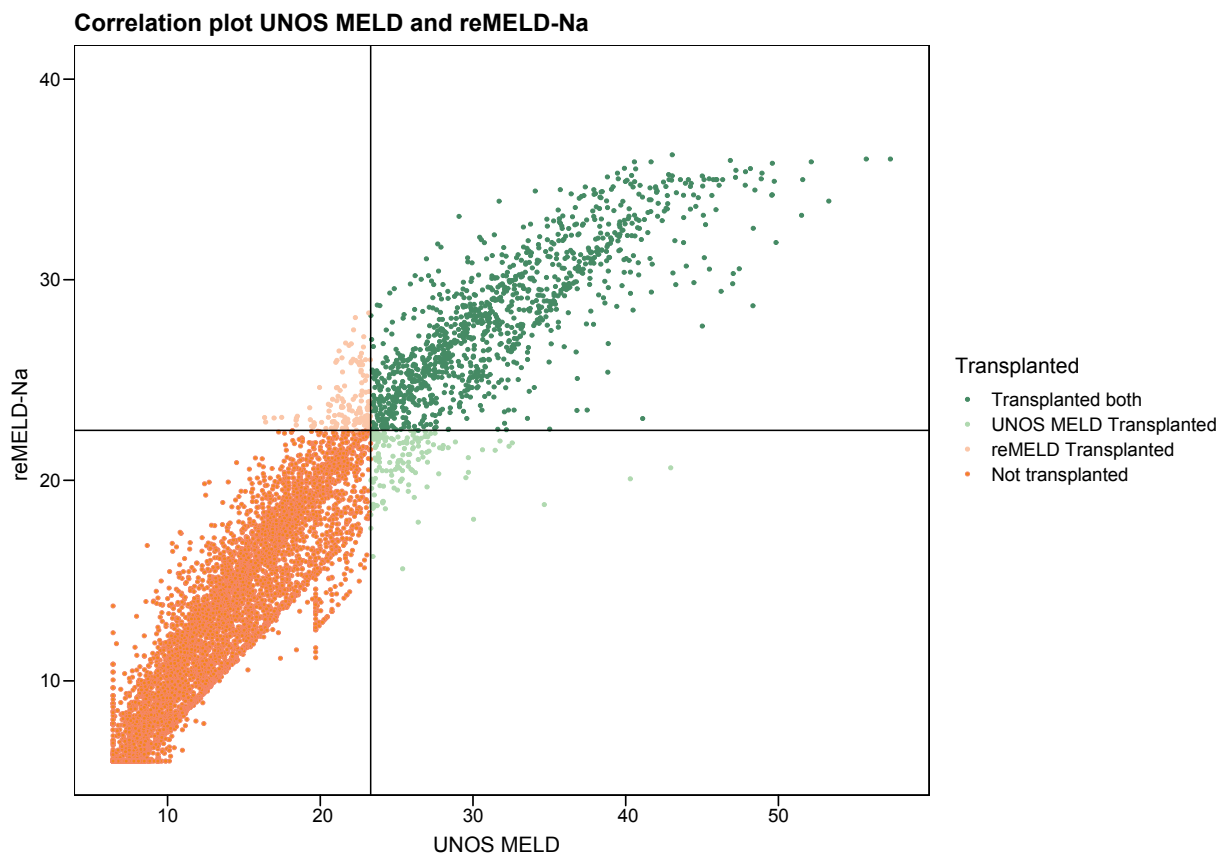
FIG. 4. Correlation plot of UNOS-MELD and reMELD-Na. Based on the number of transplanted patients after the first 90 days (n = 1,248), the highest-ranked patients according to both scores separately were assigned a liver graft, as represented by the horizontal (graft granted by reMELD-Na) and vertical (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).

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 data set. The reMELD and reMELD-Na gave convincingly higher C-indices than the 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, the UNOS-MELD is not fit for the changing LT candidate population, which showed through a decline in predictive power.[7] Refitting, i.e.,

reestablishing 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 the 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 the UNOS-MELD, the number of transplanted patients with renal failure increased significantly, possibly due to overweighed creatinine.[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 the UNOS-MELD when allocating scarce liver grafts.

High bilirubin levels led to unreliable measurements of the UNOS-MELD due to interaction with creatinine, which influenced scores because of the weight of creatinine in the 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 nontreatment) 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 the 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 Eurotransplant liver allocation. The addition of Na to the reMELD gave a small but significant improvement in discriminative ability (C-index, 0.866-0.869). Although the largest improvement in the C-index was achieved by the reMELD alone (0.849-0.866), the additional smaller gain still represented important changes for patients with hyponatremia. 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 changes. 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 patients (with hyponatremia) gives little improvement.[12,14] Based on the current findings, the reMELD-Na performed slightly but significantly better than the reMELD. Also, it seems just to consider the proven effect of Na levels on mortality. Therefore, use of the reMELD-Na is preferred.

IMPACT ON THE WL

Despite the seemingly small performance differences between the 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 (Fig. 3 and Table 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 it is both statistically and clinically very significant. A recent study showed that switching from the UNOS-MELD to the MELD-Na would significantly reduce WL 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 the MELD to the MELD-Na showed a similar increase in C-index (i.e., 0.868-0.883),[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, the UNOS MELD and the 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 (Table 4). As the 90-day mortality of the reMELD-Na-prioritized patients (Fig. 4) was 1.6 times higher than that of the

currently prioritized patients, the reMELD-Na could possibly better effectuate the sickest-first principle. Figure 4 also shows patients with MELD ≥ 40, which were rescaled to <40 after refitting. A UNOS-MELD score of 40 originally corresponded to a 100% 90-day WL mortality.[11] However, over the past decades, the WL 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 (Supporting Table S1) and allocates 25%-30% of 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 6-40 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 into 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 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 the 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.

In 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, the 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 are significant and relevant improvements compared to the currently used UNOS-MELD.

REFERENCES

1) Eurotransplant International Foundation. Statistics Library. https://www.eurotransplant.org/statistics/statistics-library/.

2) Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, et al. A model to predict survival in patients with end-stage liver disease. Hepatology 2001;33:464-470.

3) Organ Procurement and Transplantation Network. 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;31:864-871.

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;140:1952-1960.

6) Merion RM, Sharma P, Mathur AK, Schaubel DE. Evidence-based development of liver allocation: A review. Transpl Int 2011;24:965-972.

7) Godfrey EL, Malik TH, Lai JC, Mindikoglu AL, Galvan NTN, Cotton RT, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. Am J Transplant 2019;19:3299-3307.

8) Nagai S, Chau LC, Schilke RE, Safwan M, Rizzari M, Collins K, et al. Effects of allocating livers for transplantation based on Model for End-Stage Liver Disease-sodium scores on patient outcomes. Gastroenterology 2018;155:1451-1482.

9) Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 2015;162:W1-W73.

10) Jochmans I, Van Rosmalen M, Pirenne J, Samuel U. Adult liver allocation in Eurotransplant. Transplantation 2017;101:1542-1550.

11) Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, et al. Model for End-Stage Liver Disease (MELD) and allocation of donor livers. Gastroenterology 2003;124:91-96.

12) Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med 2008;359:1018-1026.

13) van Houwelingen HC, Putter H. Dynamic Prediction in Clinical Survival Analysis. 1st ed. Boca Raton, FL: CRC Press; 2011.

14) Goudsmit BFJ, Putter H, Tushuizen ME, Boer J, Vogelaar S, Alwayn I, et al. Validation of the Model for End-Stage Liver Disease sodium (MELD-Na) score in the Eurotransplant region. Am J Transplant 2020. https://doi.org/10.1111/ajt.16142.

15) Saxena V, Lai JC. Kidney failure and liver allocation: current practices and potential improvements. Adv Chronic Kidney Dis 2015;22:391-398.

16) Sharma P, Schaubel DE, Sima CS, Merion RM, Lok ASF. Re-weighting the model for end-stage liver disease score components. Gastroenterology 2008;135:1575-1581.

17) Verna EC, Connelly C, Dove LM, Adem P, Babic N, Corsetti J, et al. Center-related bias in MELD scores within a liver transplant UNOS region. Transplantation 2020;104:1396-1402.

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;10:1349-1353.

19) Nadim MK, DiNorcia J, Ji L, Groshen S, Levitsky J, Sung RS, et al. Inequity in organ allocation for patients awaiting liver transplantation: rationale for uncapping the model for end-stage liver disease. J Hepatol 2017;67:517-525.

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, Watt K, Wang B, Goldenberg A, et al. Applying machine learning in liver disease & transplantation: a comprehensive review. Hepatology 2020;0-3. https://doi.org/10.1002/hep.31103.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31677/suppinfo.