Methods

Patient population

This retrospective cohort analysis included adult (>=18 years) patients listed for a first LT between January 1st, 2010 and April 30th, 2019 on the UNOS waiting list (Figure S1). This interval ended before the May 14th, 2019 implementation of median MELD at transplant.[15] It also compromised the most recent data with adequate 5-year follow-up completeness. We aimed to calculate benefit for two patient groups: patients without HCC and without exception points (non-HCC group), and patients with HCC and with exception points (HCC group). Although other diseases also qualify for exception points, like primary sclerosing cholangitis and biliary cirrhosis, we only assessed HCC patients, as this is by far the largest group and incidence is increasing.[5] Current OPTN policy allows standard exception points for 1) HCC patients within Milan criteria (henceforth T2 HCC),[27] and 2) HCC patients initially outside Milan criteria but successfully downstaged within criteria through loco-regional treatment before LT (henceforth HCC outside criteria). Although previous study found that outcomes of these groups were similar,[28] we separately analyzed these groups, as the initial HCC disease severity and non-LT treatment are different. We excluded patients with previous LT, acute liver failure, listing for living donation, listing for multiple organs, and non-HCC malignancy (Figure S1). We randomly split our population in training data (67% of patients) and validation data (the remaining 33% of patients).

Benefit definition

Survival benefit was defined as the life-years gained from transplantation during the next five years, see Figure 6.2.[21,29] To estimate benefit for a given transplanted patient, post-transplantation survival (henceforth 'with LT') was contrasted to the hypothetical waiting list survival if LT would not have happened (henceforth 'without LT'), again see Figure 6.1.[21] Crucially, we estimated future waiting list survival from the moment of LT and not from baseline, as patients are not transplanted at baseline. To model without LT survival, we chose time-dependent Cox regression corrected with inverse probability censoring weighting (IPCW), in accordance to previous studies.[21–23,26] IPCW is used when

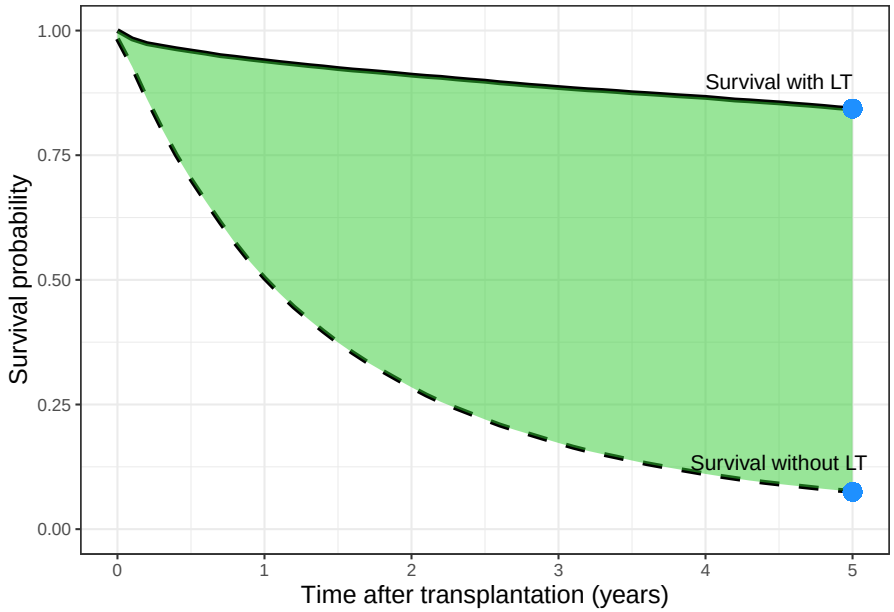


Figure 6.2: The survival with (solid line), without (dashed line) and benefit from transplantation (green area) are shown. In this example, survival is averaged for non-HCC patients with MELD-Na 25. Please note the difference in survival during five years (lines) and at five years (dots).

treatment is initiated after baseline and the chance of treatment depends on patient characteristics, that is changing MELD-Na scores over time.[21,26,30] This dependence confounds analysis of waiting list survival upon which allocation is based. These risks therefore must be corrected with statistical methods, preferably IPCW.[21,26,30] Unlike previous work,[13,16,17,28,31,32] we specifically did not use intention-to-treat (ITT) or competing risk analysis, please see supplement 1 for a detailed explanation. In short, because 1) they predict a different risk than without LT survival, 2) could result in undertreatment of patients,[30] and 3) we wanted to model changes in waiting list disease over time beyond baseline. The IPCW analyses are more complex and therefore less often applied, but this does not mean we should not use them.[33]

Statistical analysis

Waiting list survival

The waiting list population was divided in biweekly cross-sections, because in allocation liver grafts are offered to active patients on the waiting list at a certain date, not whole study cohorts of patients.[21] In time-dependent Cox analysis, repeated MELD-Na scores were modeled over time. Date and type of pre-LT HCC treatments were specifically included to account for their effects on waiting list survival. Additional predictors were used to correct the longitudinal data (Table S1), which we selected from available UNOS candidate variables deemed clinically relevant in published studies.[10,16–18,21] We excluded some variables a priori, because they referred to pediatric recipients, exclusion criteria, or donor characteristics. The outcome of analysis was waiting list mortality, which comprised death while awaiting LT and removal because of worsened condition. We censored for all other outcomes (e.g., transplantation, removal due to recovery, end of study) and corrected for dependent censoring with IPCW. Through IPCW we also estimated without LT survival of transplanted patients based on logic continuation of disease trajectories of similar patients at the same moment in time that were not transplanted (yet), please see supplement 1 for further explanation.

Post-transplantation survival

We then used Cox proportional hazards regression to model post-transplant survival. Predictors were selected by assessing relations of available UNOS recipient and donor variables to 5-year survival in univariate models, with backwards selection of significant variables in multivariate analysis. The outcome was 5-year post-transplant survival, defined as the difference between the date of transplantation and the earliest date of death, loss to follow-up or end of study on April 30th, 2019.

Calculating benefit scores

After establishing the Cox models in the training data, 5-year survival benefit from LT was calculated for each transplanted patient in the in-

dependent validation data. Benefit scores were averaged per biochemical MELD or MELD-Na [MELD(-Na)] score at transplantation, respectively for transplantations before or after January 11th, 2016, and stratified for non-HCC and HCC patients. We visualized benefit with smoothed general additive model plots per MELD(-Na) score and (non-)HCC disease. We assessed model discrimination for 5-year survival by calculating the area under the receiver-operating-curve (AUC). Cox proportional hazards model calibration (i.e., model accuracy) at five years was assessed based on bootstrapping with 200 repetitions, to obtain overfitting-corrected estimates of predicted survival, which were compared to observed survival probabilities.[34]

Online benefit score calculator

It was of interest to calculate LT benefit scores based on individual patient and donor characteristics. These benefit predictions had to be readily available online for the clinician and patient, in an intuitive calculator. Therefore, we fit a regression model to the previously calculated 5-year survival benefit scores. To compromise clinical ease of use and predictive power, only the most predictive variables were used in the benefit regression model. Variable importance for benefit prediction was assessed based on ANOVA tests. We used the overfitting-corrected R^2 to assess how much variation in benefit was explained by the predictors.[34] A R^2 value of 1 indicates that all variability in predictions is accounted for and a value above 0.9 therefore indicates excellent model predictions. The online calculator also gives graphical summaries of benefit, averaged per MELD-Na score and (non-)HCC disease, to illustrate the gain of life years during the next five years.

Results

Patient characteristics at transplantation

Characteristics for non-HCC and HCC patients at transplantation between 2010-2019 are shown in Table 6.1 . Compared to non-HCC patients, HCC patients were slightly older, more often male, and less often of white race/ethnicity. HCC patients also more frequently had diabetes mellitus, were less dependent on renal replacement therapy, and

had lower median MELD(-Na) scores. HCC patients were mostly transplanted in medium (2, 4, 6, 7, and 8) and long (1, 5, and 9) UNOS waiting time regions, whereas non-HCC patients were mostly transplanted in short (3, 10, and 11) waiting time regions. Until the moment of transplantation, the vast majority (93%) of HCC patients were at home and therefore significantly less often in hospital or ICU than non-HCC patients. Accordingly, non-HCC patients were more often dependent on life-support. Median MELD-Na scores in non-HCC, T2 HCC, and HCC beyond criteria patients were 25, 12, and 11, respectively. The AFP at transplantation for within Milan/T2 criteria and initially outside Milan/T2 criteria HCC patients was on average (SD) 67 (294) and 61 (262) ng/mL, respectively. The average AFP levels were higher in T2 HCC patients than HCC patients beyond criteria, which was likely due to the higher frequency of downstaging non-LT treatment. At time of transplantation, HCC outside criteria patients more frequently had two or three tumors. Average total tumor diameter for T2 and non-T2 HCC was 2.79 (1.11) cm and 3.17 (1.89) cm, respectively. Donor risk index scores were comparable for (non-)HCC patients, therefore HCC patients on average received the same donor quality organs as non-HCC patients.

Table 6.1: Recipient and donor characteristics at transplantation between 2010-2019

Characteristics	No HCC	T2 HCC	HCC outside criteria	p
n	24503	6922	5448	
Age (median [IQR])	56.0 [48.0, 62.0]	60.0 [56.0, 65.0]	62.0 [58.0, 65.0]	<0.001
Female sex (%)	8926 (36.4)	1614 (23.3)	1133 (20.8)	<0.001
Race/ethnicity (%)				
White	18897 (77.1)	4907 (70.9)	3705 (68.0)	
Black	1956 (8.0)	683 (9.9)	542 (9.9)	
Hispanic	2790 (11.4)	873 (12.6)	782 (14.4)	
Other	860 (3.5)	459 (6.6)	419 (7.7)	
BMI (median [IQR])	28.0 [25.0, 33.0]	28.0 [25.0, 32.0]	28.0 [25.0, 32.0]	NS
Indication for transplantation (%)				
Alcoholic	6938 (28.3)	-	-	
Cholestatic	2805 (11.4)	-	-	
Hepatitis C virus	4666 (19.0)	-	-	
NASH	4688 (19.1)	-	-	
Other	5406 (22.1)	-	-	
T2 HCC	-	6922 (100)	-	
HCC outside criteria	-	-	5448 (100)	
Diabetes (%)	6113 (24.9)	2237 (32.3)	1863 (34.2)	<0.001
Dialysis independent (%)	20998 (85.7)	6803 (98.3)	5389 (98.9)	<0.001
MELD score (median [IQR])	25.0 [18.0, 33.0]	12.0 [9.0, 16.0]	11.0 [8.0, 14.0]	<0.001
MELD-Na score (median [IQR])	27.0 [20.0, 34.0]	13.0 [9.0, 17.0]	11.0 [8.0, 16.0]	<0.001

Table 6.1: Recipient and donor characteristics at transplantation between 2010-2019 *(continued)*

Characteristics	No HCC	T2 HCC	HCC outside criteria	p
Region waiting time* (%)				
long	4614 (18.8)	1643 (23.7)	1401 (25.7)	
medium	9135 (37.3)	3093 (44.7)	2255 (41.4)	
short	10754 (43.9)	2186 (31.6)	1792 (32.9)	
Location (%)				
home	14142 (57.7)	6385 (92.2)	5124 (94.1)	
hospital	6423 (26.2)	392 (5.7)	251 (4.6)	
ICU	3938 (16.1)	145 (2.1)	73 (1.3)	
Life-support dependent (%)	2251 (9.2)	79 (1.1)	39 (0.7)	<0.001
AFP in ng/mL (mean (SD))	-	67 (294)	61 (262)	<0.001
Number of HCC lesions (%)				
1	-	74.2	65.5	
2	-	19.3	24.6	
3	-	6.5	9.9	
Total tumor diameter (mean (SD))	-	2.79 (1.11)	3.17 (1.89)	<0.001
Donor risk index (median [IQR])	1.35 [1.11, 1.64]	1.36 [1.11, 1.65]	1.37 [1.11, 1.65]	NS

Note:

HCC: hepatocellular carcinoma, AFP: alpha-fetoprotein

* Long wait time is UNOS regions 1, 5, and 9; mid wait time is regions 2, 4, 6, 7, and 8; and short-wait time is regions 3, 10, and 11.

Waiting list survival model

The significant predictors of the waiting list Cox model are shown in Table S1. In summary, the most important predictors of survival without LT were age, MELD(-Na) score, serum sodium, serum AFP, serum albumin, presence of diabetes mellitus, presence of ascites, and liver disease etiology. By correcting coefficients through IPCW, the importance of MELD(-Na) increased (data not shown), which was expected as we aimed to correct for dependent censoring bias.

Post-transplantation survival model

The significant predictors for the post-transplantation survival model are shown in Table S2. Most important were age, liver disease etiology, being of black race/ethnicity, presence of diabetes mellitus, mechanical ventilation, total tumor diameter, serum AFP, and DRI score. HCC patients with MELD(-Na)>19, AFP>24 ng/mL, and total tumor diameter>3.2 cm had the worst posttransplant 5-year survival rates (58.1%; 95% CI 50.2-67.2). For all other HCC patients, 5-year survival was above 60% (Figure S2).[29] Post-transplant model AUC of 5-year survival was 61.9 (61.2-62.6), indicating respectable discrimination. More importantly,[35] model calibration was excellent (Figure S3), which meant that our predicted risks closely resembled observed risks. After establishing model accuracy, survival estimates and benefit were calculated in the validation data.

Survival without and with LT

The distribution of MELD(-Na) scores at transplantation is shown in Figure 6.3. Non-HCC patients were mostly transplanted at MELD(-Na) scores above 14 and HCC patients mostly below MELD(-Na) 14. This distribution is important for the interpretation of the survival and benefit estimates presented below.

Figure 6.4 shows the smoothed average survival probabilities during the next five years, both for post-transplantation (with LT: solid lines) and for remaining on the waiting list (without LT: dashed lines). Because life years are gained over time, Figure 6.4 shows the mean survival during five years, i.e., the mean of the lines shown in Figure 6.2. The

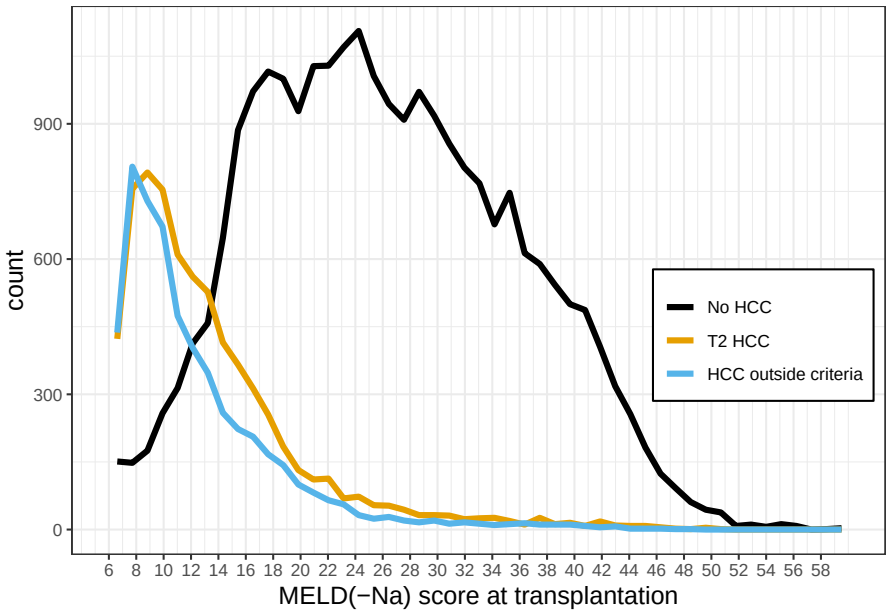


Figure 6.3: Distribution of MELD(-Na) scores at transplantation, per (non-)HCC disease. Non-HCC patients are mostly transplanted at MELD(-Na) scores >14. On the other hand, HCC patients are mostly transplanted below MELD(-Na) 14. Also, a significant part of non-HCC patients is transplanted above MELD(-Na) 30, whereas only 3% of HCC patients is transplanted at MELD(-Na) above 30.

survival probabilities at five years without and with LT are presented in Table S3, which are perhaps more intuitive survival measures for the clinician and patient. However, these hold no information regarding the survival trajectory during five years, which is what the average survival and benefit do encompass. For non-HCC patients below MELD(-Na) 10, i.e., a small number of patients, see Figure 6.3, mean survival probability without LT was better than with LT survival. In other words, on average these patients should not be transplanted. At equal MELD(-Na) scores, waiting list survival without LT for HCC patients was notably lower than for non-HCC patients. Survival without LT probabilities converged at the lowest levels, i.e., mortality could not increase much more at high MELD(-Na) scores. The average survival with LT in both groups declined above approximately MELD(-Na) 24. However, HCC survival decreased more at higher MELD(-Na) scores, most for HCC outside criteria. This decrease in posttransplant survival was possibly due to disease recurrence.

Survival benefit: life-years gained per 5 years

The 5-year transplantation survival benefit per MELD(-Na) score and per (non-)HCC disease is shown in Figure 6.5 and Table 6.2 (see Table S4 for the averages per MELD(-Na) score). Please note that the y-values correspond to the surface area shown in Figure 6.2, e.g., for a non-HCC MELD(-Na) 25 patient, LT would give 2.35 years survival benefit during the next five years.

For the 2.2% of non-HCC patients transplanted at MELD(-Na) below 9, benefit was negative, because mean postoperative life-expectancy was lower than survival without LT. With increasing MELD(-Na) scores, non-HCC benefit increased approximately linearly, up to 70% mean 5-year survival improvement for MELD(-Na) 40. The HCC benefit curves flattened with increasing MELD(-Na), whereas non-HCC benefit continued to increase. HCC MELD(-Na) >=30 benefit estimates should be interpreted carefully as they represent a small number of patients, i.e., 4.5% of the T2 HCC and 2.8% of the outside criteria HCC patients. The HCC benefit flattened at higher MELD(-Na) scores because of decreasing post-transplant survival, see Figure 6.4. Below MELD(-Na) 30, HCC patients would gain more benefit than non-HCC patients at the same MELD(-Na) score, which was mainly due to the lower expected HCC waiting list survival in absence of LT. However, the likelihood

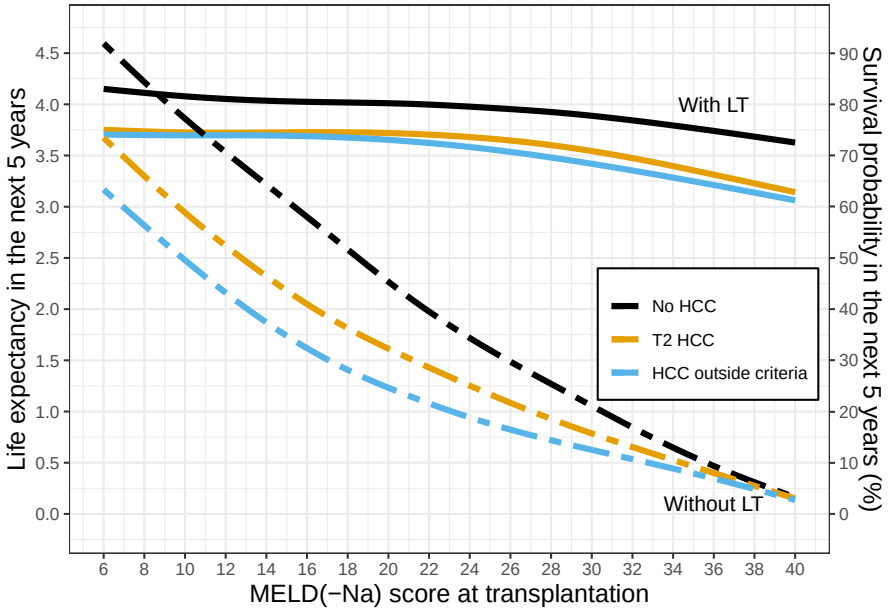


Figure 6.4: The mean survival during the next five years per MELD(-Na) score, for the waiting list (dashed lines) and after transplantation (solid lines). Note that the dashed lines represent future 'without LT' waiting list survival. The left y-axis shows life expectancy in years, the right y-axis shows survival probability.

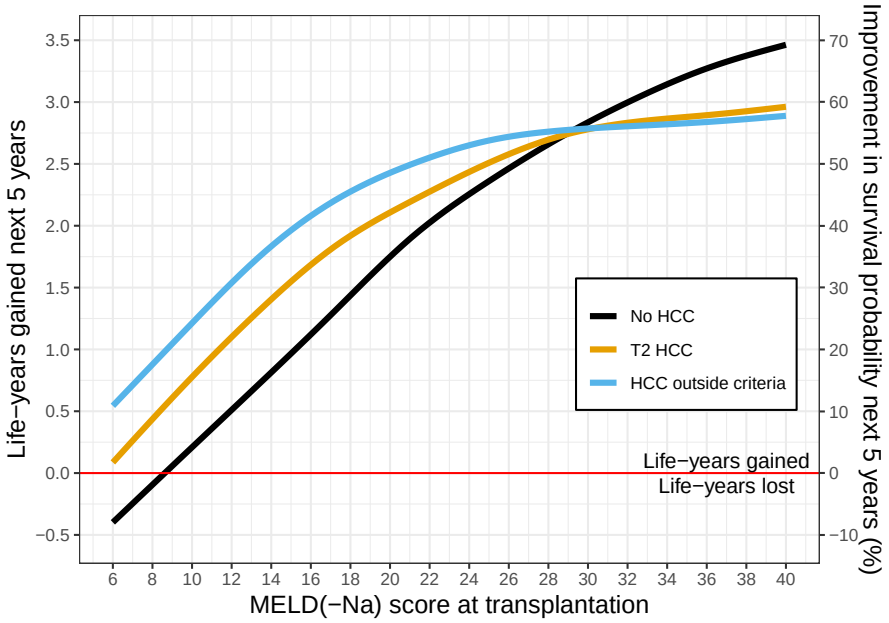


Figure 6.5: The survival benefit of liver transplantation per MELD(-Na) score. The left y-axis shows the average life-years gained in the next 5 years, i.e., the green surface area in Figure 8.2. The right y-axis shows the percentage improvement in mean survival probability during the next 5 years. Thus, for example, a T2/within Milan HCC patient with MELD-Na score 25 will gain 2.5 extra life-years in the next 5 years, i.e., the mean survival increases with 50% through transplantation.

Table 6.2: Liver transplantation 5-year survival benefit per MELD(-Na) score and etiology of disease

	No HCC			T2 HCC			HCC outside criteria		
MELD(-Na)	n	%	benefit	n	%	benefit	n	%	benefit
6-9	175	2.2	-0.14	729	32.0	0.39	717	39.2	0.82
10-13	425	5.3	0.46	675	29.6	0.98	525	28.7	1.40
14-17	943	11.7	1.08	416	18.2	1.61	304	16.6	2.03
18-21	1134	14.1	1.67	197	8.6	2.02	153	8.4	2.37
22-25	1260	15.6	2.20	106	4.6	2.37	59	3.2	2.65
26-29	1064	13.2	2.60	56	2.5	2.69	17	0.9	2.72
30-34	1159	14.4	2.99	41	1.8	2.78	21	1.1	2.72
35-40	1900	23.6	3.38	61	2.7	2.92	31	1.7	2.85
All patients	**8060**	**100**	**2.30**	**2281**	**100**	**1.19**	**1827**	**100**	**1.45**

Note:
n: number of patients per MELD(-Na) group, % : percentage of patients per MELD(-Na) group

of transplantation at lower MELD(-Na) was much lower for non-HCC patients. Figure 6.3 and Table 6.2 show that most non-HCC patients were transplanted at higher benefit scores than most HCC patients. Indeed, over 50% of HCC patients were transplanted below MELD(-Na) 14, whereas over 50% of non-HCC patients were transplanted above MELD(-Na) 26. In terms of benefit, most HCC patients gained 0.10 to 1.96 years from LT, whereas most non-HCC patients gained 2.48 to 3.46 years (Table S4). For all patients across all MELD(-Na) scores, non-HCC patients gained 3.2 years in the next 5 years through transplantation, T2 HCC gained 1.19 and HCC outside criteria gained 1.45 life-years, see Table 6.2.

Liver transplant benefit scores

Liver transplant benefit scores could be used as a continuous, equalizing metric for (non-)HCC LT access. There might be a need to calculate benefit given specific patient characteristics. This is now possible in the online benefit calculator: https://predictionmodels.shinyapps.io/benefit_calculator/. The calculator was based on a secondary regression analysis with only the most important benefit predictors, which showed an optimism corrected R^2 of 0.93. We therefore assumed that the calculator adequately predicted benefit and could serve as transla-

tion from our complex analyses to clinical practice. Variable importance in regression was summarized in Figure S4. When predicting benefit, the MELD(-Na) score was by far most important. Next were serum albumin, (non-)HCC disease, serum sodium levels, and recipient age. In line with Schaubel et al.,[21] liver function therefore remained the strongest predictor of survival benefit. Lastly, the online app also allows users to plot mean benefit per MELD-Na and (non-)HCC disease, like Figure 6.1. This can be used to inform clinicians and patients on the expected survival gain from transplantation. It also shows for selected HCC patients which non-HCC patients have equal benefit, i.e., which patients would compete for transplant based on benefit scores.

Discussion

Organ allocation aims to equally distribute donor organs to all patients in need. However, inequities on the LT waiting list exist. As a result, liver allocation has become increasingly relevant and complex. Survival benefit has gained increased attention,[13,14,16,17,29] as its optimization could improve life-years gained from transplantation for all listed patients.[21] Also, considering survival with and without LT based on patient characteristics closer resembles clinical reasoning. The objective of this study was to estimate and compare LT survival benefit for patients with and without HCC in a recent US waiting list cohort. The novelty was estimating benefit from the moment of transplantation based on longitudinal disease development up until that moment. Our results showed that mean LT survival benefit was positive across all MELD(-Na) scores, except for non-HCC patients with MELD(-Na) scores below 9. Non-HCC patients gained most life years from transplantation, as these patients were mostly transplanted above MELD(-Na) 26, where benefit was highest. HCC patients were mostly transplanted below MELD(-Na) 14, which yielded lower survival benefit. Liver function was the most important predictor of benefit. It is now possible online to calculate 5-year survival benefit based on specific patient characteristics through https://predictionmodels.shinyapps.io/benefit_calculator/.

Benefit definition

Benefit was defined as the difference in survival with and without LT during the next five years. The endpoint of survival analysis was five years, because using 10-year or overall survival as outcome would give too much importance to variables that predict post-transplant survival.[4,29] Also, further increasing the prediction horizon made estimates less certain. At five years, the waiting list model showed an excellent AUC, also when compared to other similar analyses.21,23 Compared to recently reported and tested post-transplant survival models, our 5-year post-transplant survival model performed similar (LiTES) or better (HALT-HCC, Metroticket).[10]

Estimation of benefit

We choose our methods to estimate benefit from the moment of possible LT. These methods differed from previous clinical studies that modeled waiting list survival counted from first registration.[13,16,17,32,36] Our goal was to model future survival without LT, whereas counting from baseline gives survival before LT, see Figure 6.1. Also, patient states at first listing and transplantation should not be compared, as survival changes within each patient over waiting list time due to e.g., disease progression and possible non-LT treatments.[6,21,22,24–26] We therefore calculated counterfactual waiting list survival (without LT) through time-dependent analysis with additional correction for bias.[21,26] Others performed similar analyses over time, but averaged calculated benefit over waiting list follow-up,[21,23] which for us seemed suboptimal as possible transplantation and its benefit occurred at one moment in time per patient. Lastly, some previous studies calculated benefit using characteristics of a 'median donor' assigned to all patients.[13,37] Instead, we choose to use the actual transplantations between 2010-2019, with the aim to best evaluate reality, as the observed transplants indicate inequity between (non-)HCC patients.[5]

Non-HCC and HCC benefit

A competing risks study by Berry et al. showed that HCC patients in the US overall gained negative or little benefit from transplantation, i.e., that HCC patients wasted benefit.[13] This contrasts with our

findings that mean HCC benefit was positive across all MELD(-Na) scores, mainly because HCC survival without LT was low. Clinically, it makes sense that out of two otherwise identical patients, the patient with HCC will live shorter without LT because of the malignancy in situ.[38] It was suggested that Berry et al. overestimated HCC waiting list survival,[39] and that having HCC increased risk of waiting list mortality by factor 1.5.[21] Therefore, on the individual patient level, transplantation for HCC will add life years. However, on a population level, (over)prioritizing HCC patients can indirectly waste benefit, as non-HCC patients often will gain more survival from LT due to worse liver function. Interestingly, many HCC patients were transplanted at MELD(-Na) <10, which was considered harmful in previous study.[36] Moreover, resectable HCC may be regarded a contra-indication for LT,[4] especially when considering the limited number of available liver donors. Therefore, the selection of HCC patients for transplantation remains one of the most important parts of liver graft allocation.[29]

Using benefit scores

The LT benefit scores offer a continuous metric to stratify survival equally for non-HCC and HCC patients, as one single model is used for both groups. This abandons the use of waiting time, which is inherently flawed,[19] and binary criteria, which allow underreporting of HCC severity.[40,41] Current HCC criteria lack granularity, as patients that have the same waiting list priority can have very different survival with(out) LT.[10,13,17,21] Changing LT priority based on benefit scores could therefore prevent loss of life-years, as also shown in simulations.[21] Allocation policies like the HCC cap, HCC delay, and Median MELD at Transplant helped to reduce HCC LT access, but HCC patients are currently still better of regarding waiting time, transplantation rates, and death rates.[5,42] Clearly, there is a need for an equalizing principle for all eligible LT candidates. Still, consensus must be reached whether to consider benefit in allocation at all. Understandably, some feel uncomfortable to base treatment decisions on future posttransplant outcomes, which is in part why US policy first focused on improving regional disparities.[15,43,44] On the other hand, there is consensus on acceptable posttransplant outcomes,[45] and posttransplant survival can be accurately predicted. Interestingly, in the UK, a benefit-based allocation system was implemented in 2018.[46] The evaluation of this system

will be valuable for the debate on benefit and its role in liver allocation. However, it is most important that, regardless of the driving allocation principle, scarce liver grafts should be fairly distributed based on patient characteristics and disease severity, not arbitrary exception points.

Limitations

Our study has limitations. We excluded a minority of patients with exception points that did not have HCC, our findings might therefore not apply to the whole waiting list population. However, our goal was to compare non-HCC and HCC patients. Also, five-year post-transplant follow-up was not complete for all patients, as we compromised completeness and study period. Furthermore, we could only draw conclusions based on patients that were listed for transplantation. Therefore, selection bias exists, which is inherent to the analysis of registries. The UNOS also does not register HCC recurrence, which would be valuable as HCC recurrence rates can be up to 20%, after which median survival is less than a year.[41] Studying these data in HCC MELD>30 patients would be especially interesting. Still, overall mortality is considered free from bias, whereas disease-specific survival is not.[47] Also, due to the small number of transplantations in HCC patients with MELD(-Na)>30, estimates were less reliable for that group. Lastly, the presented time-dependent IPC-weighted analyses are complex and not intuitively interpreted. However, this complexity was needed to calculate future survival without LT and to best answer the clinical question of survival benefit. We attempted to translate complexity into an easy-accessible online benefit calculator.

Conclusion

In 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. Liver transplant benefit scores offer equal survival stratification for (non-)HCC patients. It is now possible online to calculate these scores based on individual patient characteristics. Considering benefit better resembles clinical reasoning and can optimize life years gained for the whole waiting list population.

Survival benefit scores could therefore serve to more equally allocate scarce liver grafts amongst patients eligible for transplantation.

References

1. 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
2. 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
3. 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. Published online 2020. doi:10.1111/ajt.16142
4. Vitale A, Cucchetti A, Qiao GL, et al. Is resectable hepatocellular carcinoma a contraindication to liver transplantation? A novel decision model based on "number of patients needed to transplant" as measure of transplant benefit. J Hepatol. 2014;60(6):1165-1171. doi:10.1016/j.jhep.2014.01.022
5. 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
6. Mazzaferro V, Citterio D, Bhoori S, et al. Liver transplantation in hepatocellular carcinoma after tumour downstaging (XXL): a randomised , controlled ,. Lancet Oncol. 2020;21(7):947-956. doi:10.1016/S1470-2045(20)30224-2
7. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Allen C, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2016;388(10053):1459-1544. doi:10.1001/jamaoncol.2016.5688
8. Galle PR, Forner A, Llovet JM, et al. Clinical Practice Guidelines OF HEPATOLOGY EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma q. J Hepatol. 2018;69(1):182-236. doi:10.1016/j.jhep.2018.03.019
9. Alver SK, Lorenz DJ, Marvin MR, Brock GN. Projected outcomes of 6-month delay in exception points versus an equivalent Model for End-Stage Liver Disease score for hepatocellular carcinoma

liver transplant candidates. Liver Transplant. 2016;22(10):1343-1355. doi:10.1002/lt.24503

10. Goldberg D, Mantero A, Newcomb C, et al. Predicting survival after liver transplantation in patients with hepatocellular carcinoma using the LiTES-HCC score. J Hepatol. Published online 2021:1-9. doi:10.1016/j.jhep.2020.12.021
11. 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
12. 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
13. 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
14. 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
15. OPTN/UNOS Liver and Intestinal Transplantation Committee. OPTN / UNOS Policy Notice Revisions to National Liver Review Board Policies. Published 2019. Accessed April 21, 2021. https://optn.transplant.hrsa.gov/media/2816/liver_nlrb-revised-policynotice-dsa_01252019.pdf
16. Toso C, Dupuis-Lozeron E, Majno P, et al. A model for dropout assessment of candidates with or without hepatocellular carcinoma on a common liver transplant waiting list. Hepatology. 2012;56(1):149-156. doi:10.1002/hep.25603
17. 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
18. Mehta N, Dodge JL, Roberts JP, Yao FY. A novel waitlist dropout

score for hepatocellular carcinoma - identifying a threshold that predicts worse post-transplant survival. J Hepatol. Published online 2020:1-9. doi:10.1016/j.jhep.2020.10.033
19. Persad G, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. Lancet. 2009;373(9661):423-431. doi:10.1016/S0140-6736(09)60137-9
20. 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
21. 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
22. Schaubel DE, Sima CS, Goodrich NP, Feng S, Merion RM. The survival benefit of deceased donor liver transplantation as a function of candidate disease severity and donor quality. Am J Transplant. 2008;8(2):419-425. doi:10.1111/j.1600-6143.2007.02086.x
23. 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.
24. 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
25. 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. Am J Transplant. 2021;(June):ajt.16730. doi:10.1111/ajt.16730
26. Gong Q, Schaubel DE. Estimating the average treatment effect on survival based on observational data and using partly conditional modeling. Biometrics. 2017;73(1):134-144. doi:10.1111/biom.12542
27. Mazzaferro V, REGALIA E, DOCI R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med. 1996;334(11):693-699.
28. Yao FY, Kerlan RK, Hirose R, et al. Excellent outcome following down-staging of hepatocellular carcinoma prior to liver transplantation: An intention-to-treat analysis. Hepatology.

2008;48(3):819-827. doi:10.1002/hep.22412
29. 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
30. 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
31. Llovet JM, Fuster J, Bruix J. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: Resection versus transplantation. Hepatology. 1999;30(6):1434-1440. doi:10.1002/hep.510300629
32. Lai Q, Vitale A, Iesari S, et al. Intention-to-treat survival benefit of liver transplantation in patients with hepatocellular cancer. Hepatology. 2017;66(6):1910-1919. doi:10.1002/hep.29342
33. Kaplan A. The Conduct of Inquiry : Methodology for Behavioral Science. Chandler ; Chandler; 1964.
34. Harrell FE. Regression Modeling Strategies. Vol 45.; 2003. doi:10.1198/tech.2003.s158
35. 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
36. Vitale A, Huo T La, Cucchetti A, et al. Survival Benefit of Liver Transplantation Versus Resection for Hepatocellular Carcinoma: Impact of MELD Score. Ann Surg Oncol. 2015;22(6):1901-1907. doi:10.1245/s10434-014-4099-2
37. 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
38. Vitale A, Volk ML, Senzolo M, Frigo AC, Cillo U. Estimation of Liver Transplant Related Survival Benefit: The Devil Is in the Details. Gastroenterology. 2016;150(2):534-535. doi:10.1053/j.gastro.2015.12.002
39. Mehta N, Heimbach J, Hirose R, Roberts JP, Yao FY. Minimal Transplant Survival Benefit for Hepatocellular Carcinoma: Is it Real or an Overestimation of Waitlist Life Expectancy? Gastroenterology. 2016;150(2):533-534. doi:10.1053/j.gastro.2015.08.059
40. Aufhauser DD, Sadot E, Murken DR, et al. Incidence of Occult Intrahepatic Metastasis in Hepatocellular Carcinoma Treated

with Transplantation Corresponds to Early Recurrence Rates after Partial Hepatectomy. Ann Surg. 2018;267(5):922-928. doi: 10.1097/SLA.0000000000002135

41. Mahmud N, Hoteit MA, Goldberg DS. Risk Factors and Center-Level Variation in Hepatocellular Carcinoma Under-Staging for Liver Transplantation. Liver Transplant. 2020;26(8):977-988. doi: 10.1002/lt.25787
42. 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
43. Kadry Z, Schaefer EW, Uemura T, Shah AR, Schreibman I, Riley TR. Impact of geographic disparity on liver allocation for hepatocellular cancer in the United States. J Hepatol. 2012;56(3):618-625. doi:10.1016/j.jhep.2011.08.019
44. Neuberger J, Heimbach JK. Allocation of deceased-donor livers – Is there a most appropriate method? J Hepatol. 2019;71(4):654-656. doi:10.1016/j.jhep.2019.07.013
45. Mehta N, Bhangui P, Yao FY, et al. Liver Transplantation for Hepatocellular Carcinoma. Working Group Report from the ILTS Transplant Oncology Consensus Conference. Transplantation. 2020;104(6):1136-1142. doi:10.1097/TP.0000000000003174
46. National Health Service Blood and Transplantat. Policy for Deceased Donor Liver Distribution and Allocation. Published online 2018:1-18. http://www.odt.nhs.uk/transplantation/tools-policies-and-guidance/policies-and-guidance/
47. Penston J, Steele R, Brewster D. Should we use total mortality rather than cancer specific mortality to judge cancer screening programmes? Yes/No. BMJ. 2011;343(7830):1-2. doi:10.1136/bmj.d6395

Part IV: Summary, general discussion, and future perspectives

"Zodat het dan net lijkt alsof u vanaf het begin al de meest verantwoorde gedachten had over uw variabelen (u weet wel, die dingen waar u achteraf altijd zo'n spijt van had), en over uw hypothesen (u weet wel, die dingen die u dan achteraf verzon om het nog ergens op te laten lijken)."

— Arno Goudsmit

Chapter 7

Summary

The persistent scarcity of donor liver grafts necessitates prioritization of patients based on expected future survival without transplantation. The goal of this thesis was to improve survival prediction models for patients on the LT waiting list. Through advancements in prediction models, liver grafts can be allocated in the best way possible.

In **Chapter 2**, the MELD-Na score (devised in the UNOS region) was validated for the Eurotransplant region. We investigated the relationship between serum sodium levels, MELD scores, and 90-day mortality. Hyponatremia of <135, <130, and <125 mmol/L was found in respectively 28.5%, 8.8%, and 2.6% of the patients. We found that between 140 and 125 mmol/L, the risk of 90-day death increased threefold (HR 2.9; 95% CI 2.30-3.53; p<0.001). Every point decrease in serum sodium levels increased 90-day mortality by 8% (HR 0.92; 95% CI 0.90-0.94; p<0.001). Concordance statistics of MELD and MELD-Na were 0.832 and 0.847, respectively. Predictions based on MELD-Na were also more accurate than MELD. Comparing the possible impact of using MELD-Na instead of MELD for allocation on the waiting list, we found that approximately 20% of patients would receive a significantly higher predicted risk of death with MELD-Na and therefore a better chance for timely LT.

In **Chapter 3**, the 20-year-old UNOS MELD score was refitted to the Eurotransplant population. We assessed the relation of each MELD(-Na) parameter to 90-day mortality. Based on the data, the lower and upper parameter bounds and coefficients with the best fit were established. Specifically: creatinine 0.7- 2.5 mg/dL, bilirubin 0.3- 27 mg/dL, INR 0.1- 2.6, and sodium 120- 139 mmol/L. The resulting reMELD(-Na) significantly improved fit, discrimination, and calibration compared to MELD(-Na). Compared to MELD, reMELD-Na could have prioritized patients with on average 1.6 times higher 90-day mortality, thus better effectuating the sickest-first principle.

In **Chapter 4**, we developed and validated joint models for the Eurotransplant (MELD-JM) and UNOS (MELDNa-JM) regions. Repeated MELD(-Na) measurements were modeled flexibly over time and joined with Cox proportional hazards models. It was found that both MELD(-Na) value and its rate of change were strongly associated with waiting list mortality. The JMs significantly improved AUCs and Brier scores for waiting list survival prediction in both regions. MELD(Na)-JM possibly could have prioritized patients with three to five times higher 90-

day waiting list mortality than MELD(-Na).

In **Chapter 5**, we constructed and validated the ACLF-JM for patients with ACLF on the waiting list. For the ACLF-JM, repeated MELD-Na scores were corrected for CLIF-C OF scores at baseline, age, sex, life-support dependency, presence of bacterial peritonitis, and presence of cirrhosis. ACLF-JM performance was compared to a landmark MELD-Na Cox model. ACLF grade 1 to 3 was present in respectively 16.4%, 10.4%, and 6.2% of the patients. ACLF-JM performance, measured through AUCs and prediction errors, was significantly better than landmark MELD-Na. The ACLF-JM identified patients with lower MELD-Na scores but four times higher 90-day mortality.

In **Chapter 6**, we studied the survival benefit that LT caused, by comparing 5-year survival with and without LT between patients with and without HCC in the US. HCC patients had lower waiting list survival than non-HCC patients. 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. Thus, 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.

Chapter 8

General discussion

Part I: Forms of MELD

In **Chapter 2**, we showed that hyponatremia increased 90-day waiting list mortality. We also found that MELD-Na was a significantly better predictor of waiting list survival than MELD. Prioritization based on MELD-Na survival predictions could therefore reduce waiting list mortality. However, in the Eurotransplant region, MELD-Na is still not used for liver allocation.

Sodium levels and post-transplant survival

One of the concerns in the Eurotransplant community was that prioritizing hyponatremic patients for LT could decrease post-transplant survival. This concern arose in part because pre-transplant sodium levels are associated with increased morbidity, complications, and hospital admission.[1] Some older European studies indeed showed decreased short-term post-transplant survival in hyponatremic LT recipients.[2,3] Still, in recent Eurotransplant data, we found no significant post-transplant survival differences between normo- and hyponatremic LT recipients (data not published). This is in agreement with the largest study on post-transplant sodium effects in the US.[4] MELD-Na evaluation also showed that after implementing the score for allocation, the negative effect of hyponatremia on waiting list mortality was greatly reduced.[5]

In the US, MELD-Na was implemented for MELD>11 patients after studying the effect of serum sodium levels on both survival with and without LT.[6] Ideally, such analyses would also have been done in Eurotransplant. However, the required longitudinal sodium data is not available, as sodium is not adequately registered. In our validation study, we had to exclude two-thirds of eligible patients at baseline due to missing sodium. This missingness forms the most important limitation and rationale of our MELD-Na validation study. MELD-Na implementation could further improve waiting list ranking than found in this study because missing data analysis suggested that hyponatremia likely was more prevalent in patients with missing sodium data, as these patients significantly more often had alcohol-induced cirrhosis and higher creatinine levels. The seminal validation study of MELD by Wiesner et al. also excluded 48% (n=3,214) of patients due to missing data.[7] This

illustrates that sometimes evidence of improvement is provided despite missing data.

Sodium levels and renal function

Another concern was that increasing priority based on serum sodium levels would increase LT access for patients with renal dysfunction. Liver cirrhosis leads to portal hypertension and pooling of blood in the splanchnic bed. This lowers effective circulating blood volume, which increases the risk of renal dysfunction and renal failure.[8] Hyponatremia in cirrhosis results from the renal compensation of the lowered effective circulating blood volume due to vasodilatation.[1] Considering lowered serum sodium levels could therefore increase waiting list priority and transplantation rates for patients with renal dysfunction over patients with liver failure alone.

However, (over)prioritization of patients with renal dysfunction is more likely caused by the high relative weight of creatinine in MELD than by the incorporation of serum sodium. MELD was developed in a cohort wherein patients with renal failure were excluded.[9] In these patients, high creatinine levels likely indicated hepatorenal syndrome (HRS). Treating HRS with LT can reverse renal dysfunction postoperatively. Therefore, creatinine received a high weight in MELD, i.e., an increase in creatinine levels greatly increases MELD scores and transplant access. After construction, subsequent MELD validations were done in LT waiting list populations where patients with renal dysfunction were included.[7,10] This resulted in increased prioritization and transplantation for all patients with renal dysfunction,[11,12] whereas the aim of creatinine's weight in MELD was to increase transplant rates for patients with HRS.

Interestingly, after MELD's implementation, the number of liver-kidney transplant candidates tripled.[12] Therefore, the concern of (over)prioritizing patients with renal dysfunction for LT is relevant, but argues mostly against the current form of MELD. For the Eurotransplant region, a possible clinical solution could be to optimize patient's renal function before transplantation. This would however also lower a patient's ranking on the waiting list. Perhaps a better statistical solution could be to reweigh MELD's parameters to decrease the importance of creatinine in LT allocation priority. It must be

kept in mind that measuring creatinine and estimating GFR tends to overestimate renal function in cirrhotic patients,[13,14] creatinine is however widely available.

The Eurotransplant region uses a form of MELD that was constructed 20 years ago in 231 US patients. In its current form, MELD therefore does not represent the Eurotransplant population. Moreover, the predictive power of MELD is decreasing, as shown in **Chapter 2** and in literature.[15] In **Chapter 3**, we aimed to investigate whether updating MELD's coefficients and bounds for the current Eurotransplant population would improve survival prediction for patients on the waiting list. We found that the refit models indeed significantly outperformed older non-Eurotransplant forms of MELD.

Beyond linearity

Refitted MELD and MELD-Na were based on the best fit in recent data to establish new parameter coefficients between new bounds. Refit MELD(-Na) is a linear model and splits continuous data into evidence-based categories, e.g., the proposed creatinine bounds of 0.7 and 2.5 mg/dL. The advantage of linear parameter relations to mortality is easy interpretation and computation. Some disadvantages are discussed below.

First, information was lost, as we forced linearity where the data showed non-linear parameter relations to mortality (e.g., sodium level relation to 90-day mortality). By categorizing continuous parameters, we assumed relations to be constant within each category but different between categories, which is not true. For example, we assumed that an 0.1-point creatinine increase from 0.7 to 0.8 mg/dL and from 2.4 to 2.5 mg/dL would give the same increase in risk of mortality. Then, for an increase from 2.5 to 2.6 mg/dL, a very different (constant) relation was assumed. This clearly is suboptimal. Still, these new bounds and resulting coefficients were a significantly better fit than those of UNOS-MELD. This implies that capturing the majority of patients with the right coefficient is most important.

Second, parameter lower and upper limits were set for the linear models. Beyond these limits, linearity broke down and parameter values were kept constant. Still, many patients had values beyond these limits. For example, 55% of Eurotransplant patients had a creatinine level below

1 mg/dL at listing, which was set to 1. Capping lower creatinine values might especially disadvantage female LT candidates, as measured creatinine overestimates their renal function,[16] which results in MELD underestimation of mortality and perhaps unequal transplant access. To counter this inequality, additional MELD points for women have been suggested.[17] Another possibility would be to express renal function through estimated glomerular filtration rate,[18] which is still based on creatinine. At the higher end of creatinine levels, a limit was set to 4 mg/dL, again without evidence based on mortality risks.[19] This upper limit also served to decrease the LT access for patients on dialysis, as all dialysis-dependent patients were set to this value. We proposed a new evidence-based upper limit for creatinine. Additionally, the need for dialysis could be incorporated in MELD as predictor, interacting with creatinine levels.

We especially argue against MELD's lower limits of 1 for creatinine, bilirubin, and INR, as these were chosen to prevent negative MELD scores after log-transforming values below 1.[19] Furthermore, we believe that survival probabilities should be used instead of MELD scores. Firstly, because this would eliminate the abovementioned arbitrary lower bounds of 1. Secondly, although clinicians have become used to communicating 90-day survival probabilities through MELD scores, they are an unintuitive and unnecessary translational step from actual probabilities to arbitrary scores. Currently, a 50% chance of being alive after 90 days is communicated to patients and clinicians as a MELD score of 30, which is arguably less easily understood. Primarily communicating survival probabilities would benefit both patients and clinicians.

MELD 3.0

Recently, MELD 3.0 was proposed, which refits MELD-Na and adds serum albumin, patient sex, and significant interactions.[20] Interestingly, MELD 3.0 improves none of the abovementioned limitations. Although non-linearity was present for sodium and albumin levels, a linear model was used. Lower bounds of 1 were kept. Reality is not linear, yet MELD is. Therefore, as alternative, in Supplement **Chapter 10.1** we proposed to use a flexible, non-linear waiting list model.[21] Such a spline-based model would capture non-linear relations and thus provide a better fit to the data. A concern could be that the model would

overfit. However, this seems unlikely given the large data sample and small number of predictors. A model best represents the population it was constructed in. Parameter relations to mortality will change over time within the same population, which for MELD resulted in decreased prediction performance.[15,22] The established model can be a bad fit to other independent datasets, which we confirmed by refitting the 20-year-old UNOS-MELD in a recent Eurotransplant dataset. This is why regular updates of prediction models are recommended.[23] The fear of overfitting therefore should not prevent updates that bring valuable improvements for patients on the waiting list.[15,24]

Part II: Disease over time

MELD's linearity reduces non-linear reality. MELD 6-to-40 scores are used instead of survival probabilities. Longitudinal data is registered but is currently ignored. Current prediction models are static but should be dynamically updated based on newly available data. Using MELD at one single moment does not acknowledge changes over time and how these changes are related to survival. Clinicians intuitively update estimates of patient life-expectancy with changing patient condition and measurements.

These formed the reasons to investigate LT candidate survival prediction models that could meet these demands. In **Part II** of this thesis, we aimed to better approximate a clinician who evaluates patient prognosis.

Approximation of disease severity over time

Current waiting list survival predictions are based on measurements at one moment in time, i.e., the last measurement available. However, previous data provide important information about the severity of disease and its rate of change over time.[25,26] The second part of this thesis therefore focuses on joint models (JMs), which combine longitudinal and survival analysis. This allowed investigation of the effect of changing MELD scores over time on patient survival.

Previously, time-dependent Cox (TDC) models have been used to model MELD scores and waiting list survival over time.[25–31] In TDC analysis, the changing temporal effect of a predictor is estimated based on follow-up time divided into intervals of measurement, e.g., 0-30 days and 31-

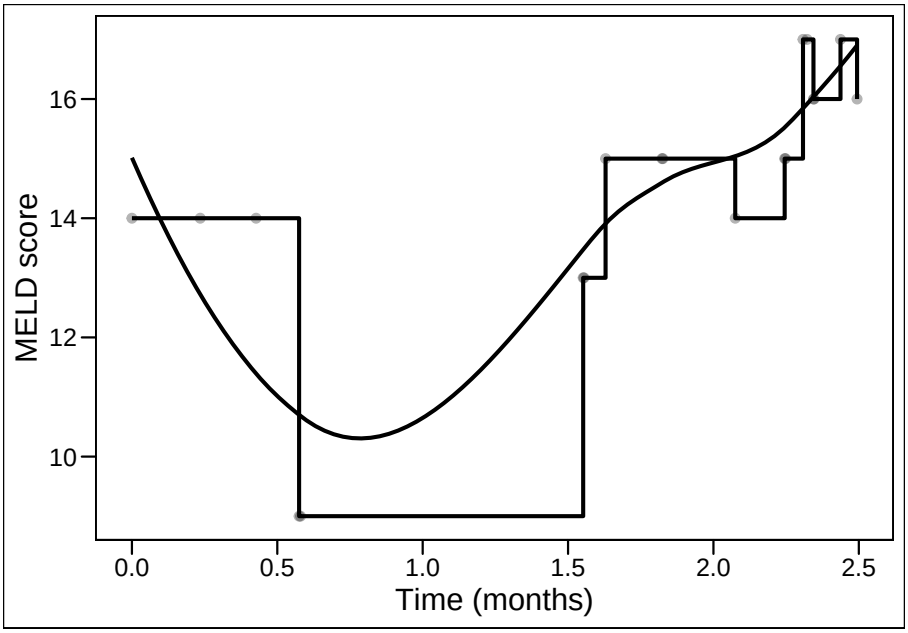


Figure 8.1: Staircase versus smooth representation of disease over time.

60 days. Within each time interval, TDC models assume that the last measured value is carried on forward. In the abovementioned example, creatinine values measured on day 0 and 31 would remain constant for the next 30 days. Crucially, there is no interpolation of values, thus a creatinine value at e.g. day 45 is not approximated.[32] In clinical terms, the TDC assumes that the disease state does not change until the next moment of measurement. This results in a 'staircase effect,' where the trajectory of disease over time is represented through rectangular steps, see Figure 8.1. Survival is then predicted based on this staircase trajectory.

However, it is clinically evident that the condition of a patient and the liver disease are not constant until a new moment of measurement. Instead, disease develops continuously, as a smooth and non-linear trajectory. We configured JMs to estimate disease as a smooth continuum over time with interpolation of trajectories between measurements. At and between measurements, not the last measured value was assumed, but the 'true' underlying value. For the abovementioned example, the JM would estimate values for each moment in time between day 0 and 60. Crucially, the model considered that measurements from the same

patient were more related than measurements between patients. TDC models ignore this correlation.

Another advantage of extrapolating the true underlying trajectory is that missing values are filled in. Missing values therefore have less effect on estimated survival. We therefore believe that JMs are better suited for real-life cohort data, where disease is continuously developing and where measurements are correlated and can be missing. The performance of JMs versus TDC models was previously assessed in small and theoretical simulation studies, where JMs showed significantly improved performance over TDC models.[33–35] However, JMs were never applied to large cohorts of patients nor to the field of LT. In **Chapter 4**, **Chapter 5** and supplement **Chapter 10.2**, we therefore investigated JMs as alternatives for survival prediction in LT candidates.

Joint modeling disease and survival over time

In **Chapter 4**, we fitted JMs to waiting list data of the Eurotransplant and UNOS regions. The JMs modeled average and individual MELD scores and considered both the value of MELD(-Na) and its rate of change at each moment in time. For the first time, liver disease was considered as developing entity within each patient on the waiting list.

Underlying MELD value

The observed, that is measured, values of MELD(-Na) scores have been used in liver graft allocation for 20 years. The JM uses these observed measurements to estimate the 'true' underlying disease trajectory. The model can therefore assume a different MELD(-Na) score than is observed for each patient. For example, for actively listed LT candidates in the Eurotransplant region between 2007-2018, the median measured value of MELD at the start of listing was 15 (Table 4.1). However, the JM assumed a baseline MELD value of 17.8, which is notably higher. In addition, for each patient, the JM considered the individual deviation from the average MELD(-Na) score at a given moment in time. This placed patients in context to the population average. The individual deviations from the average were also used as prognostic information in survival prediction. Naturally, the question arose whether this underlying disease severity should be used over actually measured disease.

Prediction performance based on underlying trajectories

Interestingly, when predicting waiting list survival, the JM outperformed MELD both at baseline and during follow-up. This implied that 1) the JM-estimated underlying disease severity better corresponded with survival than observed MELD values and 2) using individual deviation from the population average added prognostic information. The estimated underlying disease trajectory is less sensitive to missingness or errors. JMs can however be severely biased if they are mis-specified, particularly in the specification of longitudinal trajectories.[33] Therefore, we considered multiple configurations of spline-based and linear mixed effect models (longitudinal part of the JMs) and assessed their fit through Akaike information criterion (AIC) values. Most notably, the use of spline-based instead of linear-approximated patient trajectories greatly improved model fit.

Over time more waiting list data per patient typically becomes available. Therefore, after listing, the JM predictions became increasingly accurate within each patient as follow-up increased, which contrasts to MELD(-Na). Little attention is given in literature to the fact that MELD is a Cox model constructed and validated to first listing data.[7,9,10] However, most patients on the waiting list are months away from first listing. When assessing JM and MELD performance over time, a decline in discrimination and accuracy was shown. The patients who survived longest on the waiting list despite their MELD scores likely had a better condition beyond what MELD measured, or vice versa. However, since only MELD was measured, over time it became more difficult to predict survival in the resulting population. Still, JM performance was significantly better than MELD performance for most follow-up times. Also, in our analysis, all patients started from the same moment in time (first listing). However, on the actual waiting list, patients are constantly added and removed. In other words, survival prediction for liver graft allocation is based on cross-sections, not a cohort. In real waiting list data, MELD's discrimination is therefore likely to be lower, as the sickest patients are transplanted quickly and ranking the remaining less ill patients is more difficult. The JM accuracy increases with more available measurements over time. Because of this, we would not expect a similar decrease in performance if the JM would be applied to the actual waiting list.

Joint modeling acute-on-chronic liver failure

ACLF and MELD-Na

Liver disease is constantly changing. In clinical practice, the rate at which a patient changes directly influences medical urgency and possible intervention. This might be especially true for patients with acute-on-chronic liver failure (ACLF). ACLF is characterized by initially stable and chronic liver disease, which rapidly deteriorates after a predisposing event and leads to multi-organ failure and often death.[36] Timely transplantation can save a subset of these patients,[37] but MELD-Na underestimates ACLF mortality and therefore the need for transplantation.[38,39]

In supplement **chapter 10.2**, we hypothesized that JMs would be suited for predicting ACLF survival.[40] First, because each individual patient's condition can change rapidly. Therefore, it is relevant to predict survival based on both past and current data. It is also relevant to place the individual disease and survival in context to the population average. Second, by using both measured disease severity and its rate of change over time, the acceleration in ACLF severity is linked to future survival. Third, updating future predictions at each new measurement is relevant in patients with increasing disease severity.

In **Chapter 5**, we approximated liver disease severity in ACLF patients based on repeated MELD-Na values, corrected for baseline ACLF grade and other predictors (sex, age, presence of cirrhosis, life-support dependency, and presence of bacterial peritonitis). However, predicting ACLF survival based on MELD-Na measurements was suboptimal. This is because ACLF involves inflammation and multi-organ failure,[36] which are not captured by MELD-Na scores. Therefore, ACLF survival prediction could be improved further by modeling more organ system functions over time. Survival prediction based on simultaneous consideration of multiple organ systems is possible in multivariate JMs. It would make sense to separately consider the role of each organ system. Unfortunately, such data is not readily available for both the Eurotransplant and UNOS regions. Therefore, like others,[37,38,41] we could only correct for ACLF grade at baseline. However, within the European Foundation for the study of Chronic Liver Failure (EF CLIF) consortium data, longitudinal CLIF ACLF scores measurements per patients could be available. Therefore, future application of JMs in this data might result in JMs that better represent changes in ACLF and let failure of each

organ system correlate to mortality.

Underlying MELD rate of change

Despite using MELD-Na as basis, we still hypothesized that improvement was possible, mainly because baseline ACLF severity and MELD-Na rate of change would be considered. For the rate of change, the term 'slope' is often used, as the rate of change is the derivative of the function of MELD-Na values over time. The concept of MELD-Na's rate of change (or slope) over time is not new. Most notably, delta-MELD has been proposed previously.[25] However, the slope generated by the JM differs notably from delta-MELD. Firstly, the JM slope is based on the assumed true underlying disease development (see above). Secondly, the JM slope is the derivative of the measured value at one specific moment in time. In contrast, delta-MELD is defined as the difference between the current MELD score and the lowest MELD score in the previous 30 days, divided by the number of days between the current and lowest scores.[25] Thus, the obtained delta-MELD slopes are averaged over a varying number of days for different patients and time points. Also, using the lowest previously measured value overestimates the rate of change, unless the previous value actually is the lowest. This way, delta-MELD could indicate increasing disease severity even though a patient was in stable condition, see Figure 8.2 below. Basing treatment decisions on such estimates therefore seems inappropriate. In the example below, the JM slope would be approximately horizontal at 30 days and therefore the instantaneous slope at each moment is a better representation changing disease.

Lastly, the effect of delta-MELD on waiting list mortality depends on the number of previous MELD measurements.[42] This causes bias for survival prediction on the LT waiting list, as the severity of disease determines the number of measurements. For example, a clinician could increase the frequency of measurement after a patient's disease worsens, or vice versa. Thus, measuring MELD and delta-MELD in sick patients corresponds with death, but an improvement would likely be less easily observed. Therefore, Delta-MELD depends on the number of measurements, which causes bias that will increase its apparent usefulness.[26] Despite this bias, the concept of delta-MELD is used often.[5,28,43–48] In our view, this stipulates the need to adequately incorporate MELD's rate of change in survival prediction. The JM is not biased by the

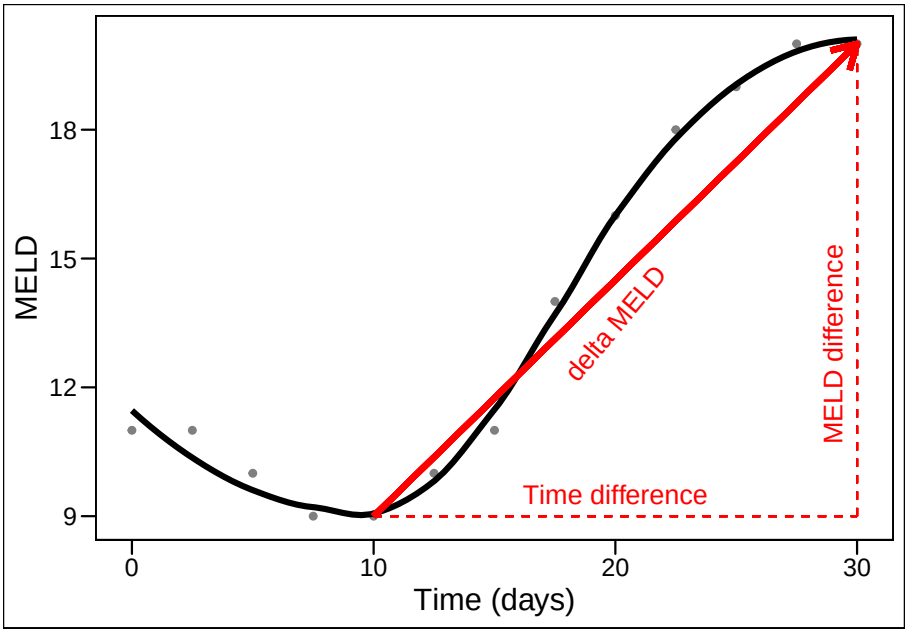


Figure 8.2: Illustration of delta-MELD slope overestimation of disease increase.

number of measurements, as it estimates a continuous underlying trajectory. Still, with increasing measurements available, the trajectory of the patient will be more accurately reflected.

Personalized predictions

Because JMs consider both the average and individual development of disease, survival predictions can be personalized. Consider that a Cox model uses coefficients, derived from a studied population, to predict outcome. These coefficients can be viewed as the average parameter-mortality relationships of a studied population. However, applying these coefficients on an individual level will give only an average prediction of survival. A patient could ask her physician: " *How long will I survive with my current disease?*" Based on a MELD score, e.g. 20, the physician could give a prognosis estimate based on population averages counted from baseline. In other words, a correct answer would be: " *If there would be 100 patients with your MELD score 20, we estimate that on average 10% will have died within three months after first waiting list registration.*" After this clarification, questions and uncertainty remain

for the patient and possibly also for the physician, because of several reasons.

First, the patient does not know how 'average' she is, that is how well the average parameter-mortality relation will apply to her. This is why considering individual patient trajectories through joint-modeling is valuable. Second, the patient is most likely not at the moment of first registration but beyond that, at some later point in time, which is why it is better to use accumulating data over time and update predictions accordingly. Third, clinicians could also miss that MELD's predictions were only validated on baseline populations.[7,9,10]

We believe that the personalized predictions can benefit both the patient and the clinician. The main reason being that the patient is recognized as unique entity and is not abstracted into population averages. The clinician can also be more confident that the predicted prognosis applies to the individual patient. Therefore, personalized JM predictions were made available at https://predictionmodels.shinyapps.io/meld-jm/.

Part III: Survival with and without transplantation

Benefit from liver transplantation

The final part of this thesis studies a simple question: *"does transplantation improve survival?"* In **Chapter 6**, we investigated whether LT *caused* survival improvement for patients on the waiting list. The difficulty is that such causal effects, that is the difference between transplanting and not transplanting, cannot be observed, as each patient is either transplanted or not. It would be considered unethical to conduct a randomized trial on LT survival benefit. Therefore, counterfactual waiting list survival of transplanted patients was estimated through inverse probability of censoring weighting (IPCW) analysis.[49] Benefit scores were calculated as the difference between survival with and without LT.

We used sequential stratification and IPCW to predict counterfactual waiting list survival, which is the waiting list survival of a transplanted patient if LT would not have been done. We applied these techniques because patients on the waiting list are transplanted after baseline and the donor graft is allocated depending on the severity of disease. See Figure 8.3 below, where four hypothetical patients on the waiting list

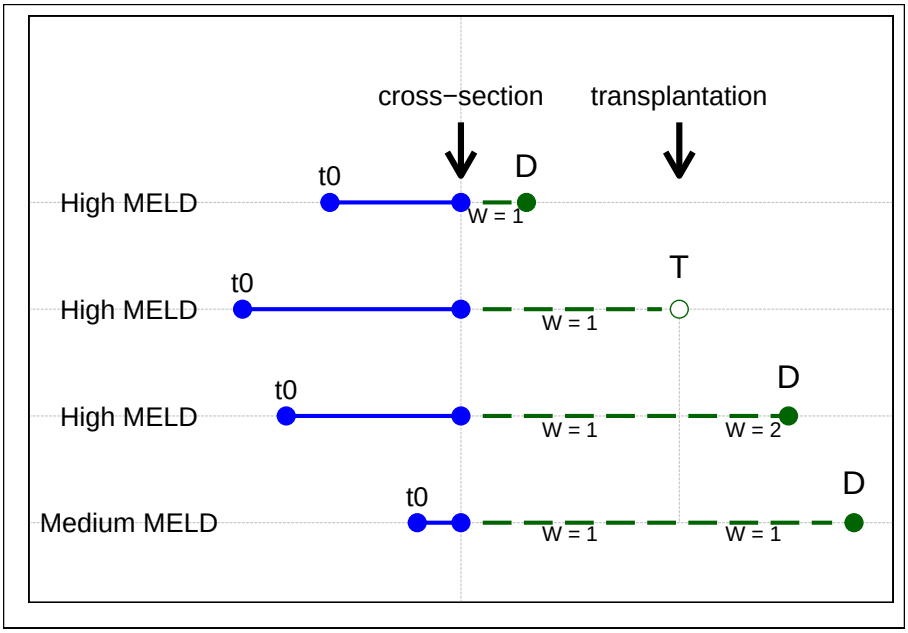


Figure 8.3: An illustration of IPCW. Note the change from w=1 to w=2 for patient 3. D: death, T: transplantation.

are shown (three severely ill, one less ill). In this example, patient 2 is dependently censored at transplantation and therefore survival of remaining and comparable patients is given more weight. Each patient was listed at a different point in time (t0) and therefore spent a variable amount of time waiting at the cross-section. Survival is counted from the moment of cross-section and all patients receive equal weights (w=1). Due to high disease severity, patient 1 died before a liver graft became available. Patient 2 survived long enough and was transplanted. After transplantation of patient 2, patient 3 received more weight (w=2) to compensate for the missing survival time of patient 2 after censoring. Patient 4 (medium MELD) did not receive higher weight as its condition was not comparable to patient 2.

Validity

In the literature, benefit is often defined as the difference between post-transplant survival and waiting list survival counted from baseline.[50–54] The idea is to match patients with a similar disease state (e.g., MELD

score) either at waiting list registration or transplantation. However, this definition of benefit assumes that two different patients at two different moments in time will yield survival curves that can be compared. We argue against these assumptions. Firstly, the fact that two patients have the same MELD score, perhaps with some more similarities like age and sex, does not make their state of disease comparable. We showed this to be true in **Part II**, where we showed that 1) previous disease development is different between two persons and 2) the rate of change in disease severity significantly influences future survival. This is perhaps best illustrated in Figure 4.1. Secondly, following from the previous arguments and the fact that liver disease typically progresses over time, survival predictions based on two different moments in time should not be compared to estimate benefit. Third argument is that the decision to transplant or not is made at the moment of liver graft offering, not baseline. Fourth, the fact that MELD and other predictors can be measured at baseline or at transplantation does not mean that it is right to use only these, which is the law of the instrument.[55] Therefore, by comparing survival within patients based on previous disease and slope, we provided a more precise and valid definition of patient disease. Still, the validity of the time-dependent Cox benefit estimates could have been improved further by using JMs to better define disease severity.

Reliability

The reliability of benefit estimates was also improved. Firstly, because we estimated survival from a certain calendar moment in time (cross-section). This is important, as liver grafts are offered to cross-sections of patients, where each patient has previously waited a variable amount of time and survival is predicted from that moment on. Counting survival from baseline instead makes all patients start from the same moment in time. Secondly, we used weighting to 1) correct waiting list survival for dependent censoring bias and 2) estimate waiting list survival as if LT was not available as treatment. Careful consideration of which question is answered by which statistical method is important when estimating benefit.[56] In clinical terms: an example is making a distinction between the survival before LT and survival without LT, which are very different (Figure 6.1).

Careful consideration argues against using competing risks (CR) anal-

yses to approximate waiting list survival without LT.[50–52] CR analysis estimates survival before LT and should be used to evaluate waiting list outcomes: transplantation, death, or removal.[56] It would be wrong to base allocation on CR-predicted future waiting list survival. To illustrate, consider a patient with MELD 40 (very ill) and a patient with MELD 20 (reasonably ill). A physician might predict correctly that the MELD 40 patient has a (much) higher chance of receiving a LT the next 90 days than the MELD 20 patient, as transplant chances increase with disease severity. In CR reasoning, we would then argue that the risk of death for the MELD 40 patient is lowered, as transplantation competes with death. However, it would be perverse to decrease allocation priority based on this reasoning, as the high chance of transplantation for the MELD 40 patient is a result of the high risk of death. Instead, priority should be based on the risk of death without LT,[49,56] which we properly modeled using censorship with adjustment for dependent censoring.

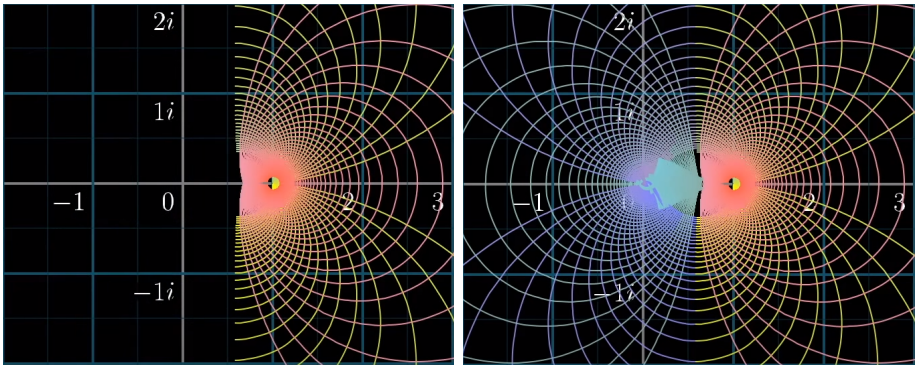
By correctly modeling waiting list survival without LT in the last part of this thesis, we must acknowledge, due to progressive insight, that the reliability of the previous survival prediction models in **Part I** and **II** could have been improved further. We modeled survival in a censorship framework but did not adjust dependent censoring bias through IPCW. This should have been done, as the priority for LT depends on MELD and the Cox model assumes that censored patients have the same chance of dying as patients who remain on the waiting list, which is not the case for transplanted patients. Since transplantation chances increase with MELD, the sickest patients typically spend the least time on the waiting list, because they are transplanted (and censored) more frequently and faster. Through IPCW, after a (high MELD) patient is transplanted, more weight is given to the remaining and comparable (high MELD) patients, who can survive some more time on the waiting list. Censoring without weights, which MELD(-Na) does, therefore leads to an increasing underestimation of mortality for patients with increasing disease severity, as death is more frequently prevented through transplantation (and after censoring outcomes and survival times are unknown). In other words, by using the unweighted MELD(-Na), current liver allocation is biased where it matters most, as it underestimates mortality in the sickest patients.

Logical continuation

Although the clinical relevance of causal models is evident, a problem is that their prediction performance cannot be assessed (yet).[57] Consider for example calibration, where predicted and observed risks are compared. This comparison cannot be done, as counterfactual waiting list survival is not observed. We are however confident about the obtained estimates. Firstly, because simulation studies showed that the used methods are valid.[49,58] Secondly, the future waiting list survival estimates of transplanted patients are a logical continuation of observed and corrected waiting list survival. An illustrative analogous example from mathematics is analytic continuation, where the domain of a function is extended in the only way possible that is preserving certain requirements. Consider Figure 8.4A below, where the lines in the right half of the plot represent a certain function (Riemann Zeta function, source: https://www.3blue1brown.com/lessons/zeta). Only the right half is shown, as only this side can be defined by the function. The left half of the plot in Figure 8.4B shows the analytically continued right half, which is continued from the right based on requirements such as line angles. However, the left half cannot be defined by the function that plots the right half, even though its continuation is logical and can be visualized. This is analogous to estimating without LT survival (left half) based on observed waiting list survival (right half), which is a logical continuation based on available data, but by definition cannot be observed nor validated.

Although causal models currently cannot be validated, benefit allocation policy has been based on these methods, most notably the MELD-Na implementation for patients with MELD>11 and UK benefit-based allocation.[6,30,59] We believe that the predicted survival without LT as possible treatment best serves as guide for transplantation assignment in future patients.[56,57,60]

Further improvements of causal liver allocation models are possible. We performed a retrospective study of benefit. However, for prospective use in allocation, IPCW could be replaced by IPTW, that is inverse probability *treatment* weighting. IPTW differs from IPCW in that both survival with and without LT are estimated as future hypothetical risks, whereas in the IPCW analysis of **Chapter 6** the with LT survival was retrospectively observed. IPTW further approximates clinical decision making based on expected outcomes with and without transplantation,



(a) Right half can be defined

(b) Left half cannot be defined, but it can be continued.

Figure 8.4: Analytic continuation of the Riemann zeta function.

as in reality both outcomes are hypothetical at the moment of liver graft offering. This requires clinicians to be comfortable with basing treatment decisions on hypothetical risks from models that cannot be validated (yet). However, this is what experienced clinicians do intuitively when evaluating an offered donor liver graft for a LT candidate. Indeed, the statistical machinery required to approximate clinical decision making is complex. This highlights the capabilities and intuition required from an experienced physician who is faced with the decision to transplant or not.

Two principles

We compared LT survival benefit of patients with and without HCC. This comparison is relevant because different allocation principles are applied to patients with and without HCC. The group of HCC patients is intended to be exemplary for other exception patients. With increasing HCC incidence,[61] already inequal LT access might be worsened further.[12] For non-HCC patients, LT listing is based on expected waiting list survival, or the principle of urgency (sickest first). HCC patients are listed based on Milan criteria, which represent acceptable post-transplant survival.[62] Considering post-transplant survival is the principle of utility, which ignores HCC waiting list survival and alternative pre-LT HCC treatment options.[63] Moreover, instead of patient char-

acteristics, artificial exception points are used to express HCC waiting list priority, which further worsened the already inequal LT access between non-HCC and HCC patients.[50,64,65] Lastly, HCC patients within Milan criteria and within one region are prioritized on waiting time, which is inherently flawed, as waiting longest does not equal to highest waiting list mortality.[66,67] To resolve these issues, we proposed the use of survival benefit as single equalizing metric. Previous simulation showed that benefit-based allocation resulted in more life-years gained from the same number of available liver grafts.[30]

However, if physicians and policy makers do not endorse benefit as metric, at least (non-)HCC waiting list survival should be estimated by a single pre-transplant survival model, which could be similar to our proposed weighted waiting list model. The use of actual patient characteristics to estimate both waiting list and post-transplant survival removes the need for the inherently flawed exception points. With the availability of HCC waiting list survival prediction models, there is no need for arbitrary and artificial inadequacy through exception points, as these are solely needed to compensate MELD(-Na)'s inability to predict waiting list survival in patients with preserved liver function. Policy and research should focus on collecting data and establishing models that adequately predict survival. This would remove the ongoing time-consuming arbitrary changes required for the exception point system. Survival prediction and liver graft allocation should be based on actual patient characteristics, not arbitrary points.

Chapter 9

Future perspectives

Simulation

Throughout this thesis, several methods were applied to estimate new model impact on the LT waiting list. We used reclassification tables, new-to-old score differences, or estimated changes in waiting list priority. These methods were used because reviewers and policymakers requested evidence of possible model impact on current waiting list outcomes. Although understandable, it is difficult and likely impossible to reliably estimate the impact of a new model on the allocation system. The best way to evaluate the effects of a new model is to implement it. The next best option is evaluation through simulation. For the Eurotransplant region, a simulation program is currently missing. An important future direction of research could therefore be the construction of what could be called the Simulation of the Eurotransplant Liver Allocation System (SELAS). SELAS would improve both Eurotransplant allocation research and policy. It would also help Eurotransplant regain its leading role in organ allocation and development. Realization of SELAS seems feasible given the existing collaboration between Eurotransplant International Foundation and the Technical University Eindhoven, as the latter has considerable experience with simulation models. The longstanding cooperation between Eurotransplant and the Leiden University Medical Center would then ensure integration of allocation, statistical methodology, and clinical knowledge.

In the U.S.A., a liver simulation program is available, that is the Liver Simulated Allocation Model (LSAM). LSAM lets users change existing allocation rules and simulate the effects in historical US data. Indeed, US allocation research is often complemented by simulation evidence. Still, simulated results should be interpreted with care. Evaluation of LSAM showed that although trends were adequately estimated, exact numbers of waiting list deaths and transplants were over- and underestimated, respectively.[68] Also, simulation performance was significantly worse for pediatric patients,[69] which indicates that simulations might be unreliable for yet undefined subgroups.

Even simulation programs have limitations. Therefore, researchers should rely on their methodology and clinical experience. Consider for example the refit coefficients in **Chapter 3**. We presented significant improvements in fit, discrimination, and accuracy. Although these metrics are important evidence, improvement was most intuitively shown through visual representation of new and old coefficients Figure

3.3). These clearly showed that reMELD(-Na) better represents the Eurotransplant population and therefore will likely better predict risk in future LT candidates. Simulation of evidence therefore has a role in the path of implementation, but sound methods and reasoning should be considered most important.

New model implementation

Possibilities are investigated to alleviate the shortage of available donor organs, such as more liberal donor criteria, living donation, machine perfusion, organoids, and xenotransplantation. Whatever improvements might be made, survival prediction will remain paramount to decide which patient should be treated. For example, with machine perfusion techniques, a larger number of liver grafts will likely become available and will be preserved longer outside the donor. This could imply more widespread allocation of organs to find the best match with the recipient. Also, with more time available, more complex calculations could be done to estimate outcomes of possible donor-recipient combinations. These calculations could be based on causal inference models, JMs, or ideally a combination of both.

For now, the shortage of donor organs persists. As mentioned, currently the principle of urgency is used for liver allocation, by prioritizing the sickest patients first. Eurotransplant has maintained this basis since 2006. In this thesis, we showed that significant improvements in survival prediction are possible. Understandably, reasons beyond clinical relevance and statistical significance determine model implementation. Because of (inter)national interests within Eurotransplant, changes in allocation are not easily implemented. Still, in our view, refit MELD (reMELD) would be relatively easy to implement, as no changes in the data structure of Eurotransplant would be required. We therefore urge Eurotransplant policy makers to consider that the refit models were a significantly better fit to the current Eurotransplant population, that ranking patients from most to least ill (discrimination) was significantly improved, and that refit model mortality risk estimates were more accurate. Implementation of (refit) MELD-Na would also not be very difficult, since sodium is a readily available laboratory measurement, that is almost always assessed in combination with creatinine. Again, the significant prediction improvements should form sufficient rationale for further allocation improvements.

Other additions to MELD could also be considered, such as serum albumin, von Willebrand factor and C-reactive protein.[18,20,70] A problem is that these variables are not collected within Eurotransplant. Several aspects of MELD, that are not evidence based, can however be improved without changing existing data registries.[19] Arguably one of the most important and counter-intuitive aspects is MELD's upper bound of 40, which means that patients with MELD>40 receive a score of 40. Therefore, allocation stops considering disease severity in the sickest patients. Already in the first validation study of MELD, MELD's relation to 90-day risk of death was plotted and showed an increasing waiting list mortality above MELD 40.[7] Recent evaluation confirmed this finding, without increased post-transplant mortality for recipients with MELD>40.[71] It therefore makes clinical sense to remove the upper border of MELD in order to improve allocation for the sickest patients. Other suggestions to improve MELD were mentioned previously in this thesis, like removing arbitrary lower and upper bounds and using survival probabilities as primary metric.

The implementation of JMs for allocation would require more effort. Eurotransplant would need to ensure that longitudinal data of each listed patient is available every time a liver graft is offered. However, if using one measurement per patient is possible, it should also be possible to use multiple, as these longitudinal data are stored by Eurotransplant. The computation of JM survival predictions would require notably more time than calculating MELD, as simulations are done for each patient. However, we believe that the advantages of correctly specified JMs are convincing. Also, although the JMs were trained in large patient cohorts, their practical application for the Eurotransplant waiting list would mean calculating survival for several hundred patients, which is done within minutes. Considering previous and current data for each patient on the waiting list would be a major improvement.

From urgency to benefit

Deciding how to allocate scarce medical interventions is relevant, as the recent COVID pandemic has shown for vaccines and ICU beds. The COVID pandemic also showed that with increasing resource scarcity, a shift in allocation principle could be warranted, that is from a 'first come first served' to a benefit-based approach.[72]

In the field of LT, organ demand persistently exceeds supply, which argues against sickest-first allocation.[67] This is because prioritizing the sickest ignores currently less ill patients that might gain more from treatment or who could be worse off in the future as disease progresses. Therefore, sickest-first allocation can only be just if the scarcity is temporary, which is not the case. This does evoke questions on how to handle high-urgency patients, as these patients are the pinnacle of urgency-based allocation and receive priority over other patients that have higher waiting list mortality.[31,49,73] Another extreme of urgency are multi-organ transplants. These possibly save only one life, whereas each of the organs could have saved a patient. Saving more lives is arguably more just. Finally, re-transplantations would require similar reconsideration of urgency and benefit,[73] as the highest priority is given to patients who might gain little and, perhaps more importantly, the liver is then denied to another recipient. Although benefit will not resolve all allocation issues, it is an inherently more just and therefore a better principle than urgency alone.[67]

We devised methods that predict survival benefit from LT. This opens the possibility for the change from urgency- to benefit-based allocation. It is however important to recognize that US data were used for the calculation of benefit. These US data encompass more LT candidate variables, that allow better estimation of future waiting list survival. Currently, Eurotransplant registers fewer LT candidate variables. It is easy to see that this will cause delay in allocation development, especially compared to other regions. This is arguably already the case, as No major revision of MELD allocation has been done by Eurotransplant since 2006. During this period, survival prediction models in US liver graft allocation were investigated and significantly improved. In our view, Eurotransplant should strive for a data registry structured much like UNOS, which allows researchers easy access to anonymized data. This in turn generates evidence upon which policy can be based. In our view, Eurotransplant should also provide a central platform where professionals and patients can gain insight in allocation policy and evidence. Transparency created through interactive statistics and accessible prediction models would greatly improve Eurotransplant's scientific basis and would perhaps place more trust in the organization. Most importantly, patients deserve to know their estimated prognosis of waiting for or accepting an organ. To this end, in this thesis, we provided several prediction models in interactive online applications. The aim was to

increase insight for both clinicians and patients.

Another possible solution for the advancement of liver allocation, despite the missing data across Eurotransplant, could be detailed national allocation based on more detailed hospital data. This allocation could be either benefit- or urgency-based, as long as one model is used to calculate future waiting list survival, preferably corrected for dependent censoring. Most organs are allocated nationally, that is 83.4% of MELD-allocated liver grafts in Belgium, Germany, and The Netherlands (*data not published*), which also ignores possibly sicker recipients abroad. Therefore, it seems feasible to abandon the sickest-first principle and to implement benefit-based allocation on a national level. This way, each country would be responsible for the method and accuracy of its survival prediction and subsequent allocation. International organ exchange would then be based on Eurotransplant standards.

Conclusion

In conclusion, this thesis investigated survival prediction models in the setting of LT, where organ scarcity and allocation necessitates continuous development of such methods. Statistically significant and clinically relevant advancements were demonstrated that could improve liver allocation through better survival prediction for patients on the waiting list.

References

1. Alukal JJ, John S, Thuluvath PJ. Hyponatremia in Cirrhosis: An Update. Am J Gastroenterol. 2020;115(11):1775-1785. doi:10.14309/ajg.0000000000000786
2. 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
3. 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
4. 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
5. 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
6. 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.
7. Singal AK, Ong S, Satapathy SK, Kamath PS, Wiesner RH. Simultaneous liver kidney transplantation. Transpl Int. 2019;32(4):343-352. doi:10.1111/tri.13388
8. 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
9. 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
10. 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
11. Sharma P, Schaubel DE, Guidinger MK, Goodrich NP, Ojo AO, Merion RM. Impact of MELD-based allocation on end-stage renal disease after liver transplantation. Am J Transplant.

2011;11(11):2372-2378. doi:10.1111/j.1600-6143.2011.03703.x
12. 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
13. Gonwa TA, Jennings L, Mai ML, Stark PC, Levey AS, Klintmalm GB. Estimation of glomerular filtration rates before and after orthotopic liver transplantation: Evaluation of current equations. Liver Transplant. 2004;10(2):301-309. doi:10.1002/lt.20017
14. Sherman DS, Fish DN, Teitelbaum I. Assessing renal function in cirrhotic patients: Problems and pitfalls. Am J Kidney Dis. 2003;41(2):269-278. doi:10.1053/ajkd.2003.50035
15. 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
16. Cholongitas E, Marelli L, Kerry A, et al. Female liver transplant recipients with the same GFR as male recipients have lower MELD scores - A systematic bias. Am J Transplant. 2007;7(3):685-692. doi:10.1111/j.1600-6143.2007.01666.x
17. Allen AM, Heimbach JK, Larson JJ, et al. Reduced Access to Liver Transplantation in Women: Role of Height, MELD Exception Scores, and Renal Function Underestimation. Transplantation. 2018;102(10):1710-1716. doi:10.1097/TP.0000000000002196
18. Asrani SK, Jennings LW, Kim WR, et al. MELD-GRAIL-Na: Glomerular Filtration Rate and Mortality on Liver-Transplant Waiting List. Hepatology. 2020;71(5):1766-1774. doi:10.1002/hep.30932
19. 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
20. Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era. Gastroenterology. Published online 2021. doi:10.1053/j.gastro.2021.08.050
21. Goudsmit BFJ, Putter H, van Hoek B. The Model for End-stage Liver Disease 3.0: an update without proven accuracy. Gastroenterology. Published online 2021. doi:10.1053/j.gastro.2021.09.047
22. 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. Published online

2020. doi:10.1111/ajt.16142
23. Jenkins DA, Sperrin M, Martin GP, Peek N. Dynamic models to predict health outcomes: current status and methodological challenges. Diagnostic Progn Res. 2018;2(1):1-9. doi:10.1186/s41512-018-0045-2
24. D'Amico G, Maruzzelli L, Airoldi A, et al. Performance of the model for end-stage liver disease score for mortality prediction and the potential role of etiology. J Hepatol. Published online 2021. doi:10.1016/j.jhep.2021.07.018
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. Therneau T, Crowson C, Atkinson E. Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. 2019;33(3):1-8. doi:10.1093/jpepsy/jsn055

33. 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
34. 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
35. 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
36. Arroyo V, Moreau R, Jalan R. Acute-on-Chronic Liver Failure. N Engl J Med. 2020;(382):2137-2145. doi:10.1056/NEJMra1914900
37. 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
38. 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
39. 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. Published online 2020. doi:10.1016/j.jhep.2020.06.005
40. 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
41. 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
42. 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

43. 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
44. Huo TI, Wu JC, Lin HC, et al. Evaluation of the increase in model for end-stage liver disease (delta MELD) score over time as a prognostic predictor in patients with advanced cirrhosis: Risk factor analysis and comparison with initial MELD and Child-Turcotte-Pugh score. J Hepatol. 2005;42(6):826-832. doi:10.1016/j.jhep.2005.01.019
45. 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
46. Györi GP, Silberhumer GR, Zehetmayer S, et al. Dynamic changes in MELD score not only predict survival on the waiting list but also overall survival after liver transplantation. Transpl Int. 2012;25(9):935-940. doi:10.1111/j.1432-2277.2012.01519.x
47. 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
48. Brock GN, Washburn K, Marvin MR. Use of rapid Model for End-Stage Liver Disease (MELD) increases for liver transplant registrant prioritization after MELD-Na and Share 35, an evaluation using data from the United Network for Organ Sharing. PLoS One. 2019;14(10):1-17. doi:10.1371/journal.pone.0223053
49. Gong Q, Schaubel DE. Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. Biometrics. 2013;69(2):338-347. doi:10.1111/biom.12023
50. 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
51. Toso C, Dupuis-Lozeron E, Majno P, et al. A model for dropout assessment of candidates with or without hepatocellular carcinoma on a common liver transplant waiting list. Hepatology.

2012;56(1):149-156. doi:10.1002/hep.25603
52. 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
53. Lai Q, Vitale A, Iesari S, et al. Intention-to-treat survival benefit of liver transplantation in patients with hepatocellular cancer. Hepatology. 2017;66(6):1910-1919. doi:10.1002/hep.29342
54. Vitale A, Huo T La, Cucchetti A, et al. Survival Benefit of Liver Transplantation Versus Resection for Hepatocellular Carcinoma: Impact of MELD Score. Ann Surg Oncol. 2015;22(6):1901-1907. doi:10.1245/s10434-014-4099-2
55. Kaplan A. The Conduct of Inquiry: Methodology for Behavioral Science. Chandler; Chandler; 1964.
56. 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
57. Sperrin M, Diaz-Ordaz K, Pajouheshnia R. Invited Commentary: Treatment Drop-in—Making the Case for Causal Prediction. Am J Epidemiol. 2021;190(10):2015-2018. doi:10.1093/aje/kwab030
58. Gong Q, Schaubel DE. Estimating the average treatment effect on survival based on observational data and using partly conditional modeling. Biometrics. 2017;73(1):134-144. doi:10.1111/biom.12542
59. National Health Service Blood and Transplantat. Policy for Deceased Donor Liver Distribution and Allocation. Published online 2018:1-18. http://www.odt.nhs.uk/transplantation/tools-policies-and-guidance/policies-and-guidance/
60. Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: A simulation study. BMC Med Res Methodol. 2017;17(1):1-12. doi:10.1186/s12874-017-0375-8
61. Fitzmaurice C, Akinyemiju T, Abera S, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level results from the global burden of disease study 2015. JAMA Oncol. 2017;3(12):1683-1691. doi:10.1001/jamaoncol.2017.3055
62. Mazzaferro V, REGALIA E, DOCI R, et al. Liver transplantation

for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med. 1996;334(11):693-699.
63. 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
64. 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
65. 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
66. 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
67. Persad G, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. Lancet. 2009;373(9661):423-431. doi:10.1016/S0140-6736(09)60137-9
68. Goel A, Kim WR, Pyke J, et al. Liver Simulated Allocation Modeling: Were the Predictions Accurate for Share 35? Transplantation. 2018;102(5):769-774. doi:10.1097/TP.0000000000002079
69. Wood NL, Mogul DB, Perito ER, et al. Liver simulated allocation model does not effectively predict organ offer decisions for pediatric liver transplant candidates. Am J Transplant. 2021;21(9):3157-3162. doi:10.1111/ajt.16621
70. Starlinger P, Ahn JC, Mullan A, et al. The Addition of C-Reactive Protein and von Willebrand Factor to Model for End-Stage Liver Disease-Sodium Improves Prediction of Waitlist Mortality. Hepatology. 2021;74(3):1533-1545. doi:10.1002/hep.31838
71. 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
72. FMS. Draaiboek Triage Op Basis van Niet-Medische Overwegingen Voor IC-Opname Ten Tijde van Fase 3 in de COVID-19 Pandemie Criteria Voor Fase 3 Stap C Aansluitend Op Het NVIC

Draaiboek Pandemie Versie 2.0-November 2020.; 2020.

73. de Boer J, Braat A, Putter H, et al. Outcome of Liver Transplant Patients with High Urgent Priority. Are We Doing the Right Thing? Transplantation; 2018. doi:10.1097/tp.0000000000002526

Chapter 10

Appendix

10.1 Supplement Chapter: The Model for End-stage Liver Disease 3.0: an update without proven accuracy

Goudsmit BFJ, Putter H, van Hoek B. The Model for End-stage Liver Disease 3.0: an update without proven accuracy. *Gastroenterology*, 2021; doi: 10.1053/j.gastro.2021.09.047.

Letter

With great interest we read the study by Kim et al.[1] In this work, the authors showed that MELD-Na performance is improved by including serum albumin levels, LT candidate sex, a creatinine cap set to 3 mg/dL, and significant interactions. Most notably, the MELD 3.0 concordance statistic (c-index) was 0.869, versus a MELD-Na c-index of 0.862. However, we have some concerns regarding this study.

First, the authors report only discrimination (c-index) as model performance indicator. Indeed, high discrimination is important when ranking patients for LT, as it ensures that the model prioritizes the sickest patients. However, when basing treatment decisions on estimated mortality risks, it is vital to assess and report how accurate risks are estimated, i.e., model calibration. This is because a badly calibrated model can still have a high c-index, but treatment decisions should not be based on such a model.[2] Model calibration is typically reported with calibration plots, that give insight in possible over- or underestimation of risk. Previous work showed that MELD-Na overestimated risks for the sickest patients.[3,4] More importantly, recent study found that MELD predicted risks inaccurately.[5] Therefore, the authors cannot conclude that "MELD 3.0 affords more accurate mortality prediction," as calibration was not reported. It would be interesting to assess and report MELD 3.0 calibration, especially for male versus female LT candidate sex.

Second, the authors report net 8.8% reclassification of deceased patients from a lower MELD-Na stratum to a higher MELD 3.0 stratum, for women this number was 14.9%. The idea is that higher MELD 3.0 scores thus better reflect mortality risks. The first important concern with proving MELD 3.0 prediction improvement through reclassification methods is that a poorly calibrated model can show improved prediction performance, even when this is not possible.[6] These false effects can be found both in actual cohorts and simulated data. In part, this is due to the fact that the actual waiting list population cannot be separated into the suggested MELD strata (6-9, 10-19, etc.). Instead, when evaluating added biomarkers, measures like the Brier score, that simultaneously assess discrimination and calibration, should be used in independent validation data.[6] A second concern is that reclassification allows for 'stage migration bias,'[7] i.e., assigning patients to new strata improves strata-specific survival, even though survival of individual patients has

not changed. The sickest patients from a lower MELD-Na stratum are moved to a higher MELD 3.0 stratum and survival is better in both strata. Therefore, stating that MELD 3.0 will lower deaths on the waiting list based on reclassification tables must be done cautiously, as this can inflate within-strata survival rates.

Third, the authors keep the lower borders of bilirubin, creatinine, and INR set to 1. These borders were chosen 20 years ago, to prevent negative logarithm transformation in the linear MELD formula. The more pressing clinical fact is that a substantial number of patients on the waiting list had creatinine (55%) and bilirubin (24%) values below 1 mg/dL at first registration.[8] Including these lower measurements when predicting survival would be a better representation of the actual waiting list and would place the higher values in a more appropriate context, especially considering the lower creatinine values for women. Also, even though linear models are more easily understood and used, non-linear effects are clearly present (creatinine, sodium, and albumin). Therefore, flexible models could be considered to model more measurements and their non-linear effect on mortality.

In conclusion, MELD 3.0's accuracy must be proven before it can be considered as new allocation model, e.g., with calibration plots and Brier scores. Reclassification cannot be used alone to prove clinical improvement. We agree with the authors that efforts should be made to continuously improve MELD and liver graft allocation, but appropriate evidence must be presented.

References

1. Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era. Gastroenterology. Published online 2021. doi:10.1053/j.gastro.2021.08.050
2. 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
3. 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

4. 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. Published online 2020. doi:10.1111/ajt.16142
5. D'Amico G, Maruzzelli L, Airoldi A, et al. Performance of the model for end-stage liver disease score for mortality prediction and the potential role of etiology. J Hepatol. Published online 2021. doi:10.1016/j.jhep.2021.07.018
6. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: Do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2014;33(19):3405-3414. doi:10.1002/sim.5804
7. Feinstein AR, Sosin DM, Wells CK. The New England Journal of Medicine Downloaded from nejm.org at BOSTON UNIVERSITY on September 19, 2013. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2010 Massachusetts Medical Society. All rights reserved. N Engl J Med. 1985;312(12):1604-1608.
8. 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

10.2 Supplement Chapter: 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.

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. *Journal of Hepatology.* 2021;74(2):475-476. doi: 10.1016/j.jhep.2020.08.032.

Letter

With great interest we read the article by Hernaez et al.[1] The authors showed that predicted survival by the Model for End-stage Liver Disease sodium (MELD-Na) score underestimated the observed survival in acute-on-chronic liver failure (ACLF) patients. As a result, ACLF patients might be underserved in the MELD-Na-based allocation of donor livers. We agree with the authors that the MELD-Na score is not optimal for ACLF patients. However, we suggest several considerations for this paper.

First, the authors state that "it is unclear whether MELD-Na captures clinical severity" in ACLF patients. Considering the available literature, it is clear that the disease course of ACLF is not captured by MELD-Na, especially for ACLF-3 patients.[2] In their large UNOS analysis, Sundaram et al. already showed ACLF death and removal rate to be independent of MELD-Na score, as mortality rates were highest in MELD-Na <25 and ACLF-3 patients.

Second, the MELD-Na accuracy of mortality prediction in ACLF patients is questioned. The CLIF score, specifically developed for ACLF patients, achieved a 90-day mortality concordance statistic (c-index) of 0.76, whereas the MELD-Na had a c-index of 0.67.[3] The c-index shows how accurate the model can discern between life and death, by pairwise patient comparisons in the given data. The discrimination of both scores is not optimal. Given that the MELD-Na was not developed for ACLF patients, but for chronically-ill patients at listing for liver transplantation (LT), its discrimination seems respectable. The current allocation system is based on MELD-Na because, for the majority of patients with chronic liver disease, MELD-Na offers excellent performance.[4,5] Still, the authors showed that MELD-Na and thus transplant chances increased with higher ACLF grades, with median MELD scores of 24, 27 and 32 for ACLF grade 1-3 respectively. The authors do not focus on the c-index as the main model performance indicator but assess the calibration instead. The expected and observed mortality rates in ACLF patients were compared. One could question the assessment and main focus of calibration if the model captures few relevant factors in these patients. Even in cirrhotic patients, for whom MELD-Na was designed, the MELD-Na becomes less reliable with increasing disease severity.[4,5]

Third, the authors showed that LT was not often considered/performed

in ACLF patients. Many patient-specific and center-level factors influence the evaluation for LT. Still, ACLF showed a positive association with LT, which was higher than for non-ACLF patients. Patient exclusion from transplantation is most likely due to expected futile efforts. The fact that the allocation system is MELD-Na based, does not change that. As Nadim et al. stated: "while scoring systems for ACLF may help centers decide who to transplant, the scores do not affect organ allocation; it is still the MELD score that ultimately determines organ allocation in most countries, including the US."[6] Granting exception points or status 1 may be the best option for the small number of ACLF patients listed for LT.

Finally, Hernaez et al. note that "future research should also focus on developing and validating prognostic scores that incorporate dynamic changes in patients clinical course" and that they "did not capture longitudinal changes of ACLF scores over time." Traditional Cox models, like the MELD-Na, make assumptions that often do not hold in the data and use only one measurement in time for survival prediction. Thus, dynamic changes are not modeled and longitudinal data is ignored. For dynamic prognostic modeling of longitudinal data, joint models (JM) present an appropriate method of capturing changing disease severity.[7] The JM adequately links longitudinal measurements to survival analysis by combining mixed-effect and Cox models. It considers all past measurements, changes in values and the rate of change at every point in time and uses this for patient-specific predictions that are updated based on every new available measurement. This is valuable for ACLF patients. In simulation studies, the JM outperformed Cox models with less biased results.[8–10]

In conclusion, the MELD-Na underestimates survival in ACLF patients because it uses only some of the relevant prognostic factors for ACLF patient survival. Joint models should be considered to dynamically predict patient-specific survival based on repeated measurements.

References

1. 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

2. 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
3. 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
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. 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
7. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat Med. 1996;15(August 1995):1663-1685.
8. 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
9. 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
10. 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

Nederlandse samenvatting (Summary in Dutch)

Verbetering van voorspellingsmodellen voor levertransplantatiekandidaten

Een levertransplantatie is levensreddend voor patiënten met een leverziekte. Omdat niet iedere patiënt (direct) kan worden geholpen, worden patiënten op een wachtlijst geplaatst. Op deze wachtlijst wordt de volgorde bepaald door de ernst van de ziekte: de ziekste patiënten gaan eerst. De ziekte-ernst wordt ingeschat door de toekomstige wachtlijstoverleving te berekenen. Hoe lager de toekomstige wachtlijstoverleving, hoe hoger de prioriteit. De methode van het inschatten van overleving is dus van levensbelang voor deze patiënten. Dit proefschrift onderzoekt nieuwe modellen voor het voorspellen van de overleving rond levertransplantatie.

Deel I

In **Hoofdstuk 2** werd een verbetering onderzocht van het huidige model dat de wachtlijstvolgorde bepaalt: de 'Model for End-stage Liver Disease' (MELD) score. Specifiek werd onderzocht of het uitbreiden van de MELD score met natrium (MELD-Na) een verbetering zou geven van de sterftevoorspelling op de wachtlijst. We vonden dat een laag natrium (hyponatriëmie) de kans op wachtlijststerfte verhoogt. Patiënten met een natrium van 125 mmol/L hebben een 2.9 (95%CI 2.30-3.53; $p<0.001$) keer grotere kans op sterfte binnen 90 dagen dan patiënten met een normaal (140 mmol/L) natrium. Vergeleken met de MELD score was de MELD-Na score een significant betere voorspeller van overleving, met een een c-index van respectievelijk 0.832 en 0.847. Een c-index waarde dichter bij de 1.0 is beter en betekent dat een model beter patiënten kan rangschikken op de wachtlijst van meest naar minst ziek. Waarschijnlijk zal het gebruik van de MELD-Na score voor leverallocatie de wachtlijststerfte verlagen omdat de mate van hyponatriëmie wordt meegenomen en dus wachtlijststerfte preciezer wordt ingeschat.

Aangezien de huidige vorm van de MELD score 20 jaar geleden werd ontworpen in de Verenigde Staten, werd in **Hoofdstuk 3** onderzocht of het herwegen van de MELD score in de Eurotransplant regio een betere overlevingsvoorspelling zou geven. Het lijkt gek om een Amerikaans model te gebruiken om Europese patiënten te prioriteren. We

onderzochten de relatie van de MELD parameters (serum kreatinine, bilirubine en de INR) en het natrium met de 90-daagse overlijdingskans op de wachtlijst. We vonden dat nieuwe afkapwaardes voor de MELD parameters en het natrium resulteerden in significant betere modellen: de refit MELD en refit MELD-Na. De nieuwe modellen waren preciezer in het rangschikken van patiënten op de wachtlijst. Vergeleken met de MELD, prioriteerde de refit MELD-Na score patiënten met een 1.6 keer hogere 90-daagse wachtlijststerfte. Op basis van de refit modellen zouden donorlevers dus beter verdeeld kunnen worden omdat de ziekste patiënten beter geïdentificeerd kunnen worden. Hierdoor zou sterfte op de wachtlijst kunnen worden voorkomen.

Deel II

In het tweede deel werden van dit proefschrift werden metingen over de tijd gebruikt om tegelijkertijd ziekte en overleving te modelleren. Het idee was om een betere benadering te geven van de manier waarop een arts de prognose van een patiënt inschat. Een arts zal altijd het ziekteverloop uit het verleden meenemen om de prognose in te inschatten. Het is daarom onlogisch dat de huidige modellen die wachtlijstvolgorde bepalen, alle voorgaande beschikbare metingen negeren. Net als een arts die niet meer weet wat er gisteren is gebeurd. Met de techniek van joint models (JMs) namen we alle beschikbare data over de tijd mee in voorspellingen van overleving. Hierbij werd gekeken naar zowel de gemeten ziekte-ernst als de mate van verandering. Een analogie voor ziekte-ernst en verandering is hardlopen. Je kunt met een bepaalde snelheid rennen (bijvoorbeeld 3 m/s) en daarbij versnellen (bijvoorbeeld met 1 m/s^2) of vertragen. De verandering geeft dus mogelijk belangrijke informatie over ziekte.

Hoofdstuk 4 toont de eerste toepassing van JMs in levertransplantatiekandidaten. De analyse van MELD(-Na) metingen over de tijd werd gecombineerd met overlevingsanalyse. Hierdoor kon het effect van ziekteverandering over de tijd op overleving worden bestudeerd. We vonden dat zowel de gemeten MELD(-Na) score als de mate van verandering over de tijd een belangrijke invloed hadden op wachtlijstoverleving. De JMs waren significant beter in het voorspellen van wachtlijststerfte dan de huidige modellen die wachtlijstvolgorde bepalen. De JMs zijn een belangrijke verbetering omdat alle beschikbare metingen werden gebruikt, zowel de ziekte-ernst als mate van verandering werd meegenomen en

dat persoonlijke voorspellingen konden worden gedaan. Ook werden de voorspellingsmodellen in een online applicatie geplaatst, waarmee gebruikers data van individuele patiënten kunnen uploaden om JM voorspellingen te krijgen voor wachtlijstoverleving.

In **Hoofdstuk 5** onderzochten we hoe de JMs, die nieuwe voorspellingen maken voor elke nieuwe meting over de tijd, overleving voorspelden in patiënten met Acute-on-Chronic Liver Failure (ACLF). ACLF is een dodelijke ziekte die snel verandert over de tijd. Daarom is het belangrijk dat een voorspellingsmodel meeverandert, hetgeen een JM kan. We vonden dat een aanzienlijk deel van de patiënten op de leverwachtlijst een vorm van ACLF had. Het huidige model dat overleving voorspelt (MELD-Na score) had een slechte c-index (capaciteit tot rangschikken op de wachtlijst) met oplopende ziekte-ernst. Hierdoor wordt de huidige wachtlijstprioriteit minder nauwkeurig in ziekere patiënten. Dit is ongewenst. De JMs waren nauwkeuriger en bleven dat ook in de ziekste patiënten. Met de JMs konden nauwkeurigere voorspellingen worden gegeven, voor zowel de populatie als het individu.

Deel III

In het laatste deel en **Hoofdstuk 6** onderzochten we hoeveel levenswinst patiënten kregen door levertransplantatie. Het verschil in overleving met en zonder levertransplantatie werd berekend en vergeleken tussen patiënten met en zonder hepatocellulair carcinoom (HCC). We vonden dat patiënten met HCC een hogere wachtlijststerfte hadden en meestal bij lagere MELD(-Na) scores werden getransplanteerd dan niet-HCC patiënten. Doordat HCC patiënten bij lagere MELD(-Na) werden getransplanteerd, haalden ze minder levenswinst uit levertransplantatie dat niet-HCC patiënten, die vooral bij hogere MELD(-Na) scores werden getransplanteerd. Leverfunctie was de belangrijkste voorspeller van overlevingswinst en daarom kregen patiënten zonder HCC gemiddeld meer levensjaren van transplantatie. Gezien de schaarste van donor levers zou men dus kunnen overwegen om HCC patiënten zoveel als mogelijk zonder levertransplantatie te behandelen.

In conclusie werden er in dit proefschrift modellen onderzocht die overleving voorspellen in levertransplantatie. De statistisch significante en klinisch relevante verbeteringen kunnen worden gebruikt om de huidige leverallocatie te verbeteren.

Dankwoord

Tot slot een woord van dank. De afgelopen jaren heb ik me mogen verdiepen in het onderzoek naar voorspellingsmodellen in levertransplantatie. Dit is mogelijk geweest door een groot aantal mensen, maar ik wil met name de volgende personen bedanken:

Doctor Braat, beste Dries, door jou is dit alles begonnen. Je hebt me enthousiast gemaakt voor de complexe wereld van het transplantatie-onderzoek en je hebt me geholpen aan mijn eerste baan. Je geduldige begeleiding, visie, koppigheid en humor zijn onmisbaar geweest voor de totstandkoming van dit proefschrift.

Professor van Hoek, beste Bart, vanaf het begin was het voor mij duidelijk dat je een betrokken arts en begeleider bent. Je oprechte interesse en enthousiasme voor mijn onderzoek zijn een grote steun geweest.

Professor Putter, beste Hein, je bent een tovenaar. Ik ben begonnen met R doordat ik de magie wilde leren die ik je wekelijks zag programmeren. Dit heeft mijn onderzoek ongetwijfeld verrijkt en verdiept. Door jouw wekelijkse sturing is dit alles mogelijk geworden.

Doctor Tushuizen, beste Maarten, je geloofde in de potentie van mijn ideeën en was altijd bereid om snel en nauwkeurig mee te denken en mijn manuscripten te lezen. Alleen PhD'ers kunnen volledig waarderen hoe waardevol dat is.

Geachte leden van de leescommissie en de oppositie. Dank voor jullie instemmen met en het verdiepen in mijn proefschrift. Het is een eer de gelijktijdige aandacht te krijgen van zo veel scherpe geesten.

Beste collega PhD'ers, dank voor de (on)zinnige praat, de mooie reizen en avonden die we samen hebben gehad. Met name dank aan Jaap en Fenna voor alle extra begeleiding.

The author thanks the Eurotransplant Liver and Intestine Advisory Committee (ELIAC) members for their critical appraisal and approval of several study protocols.

De medische staf van Eurotransplant, dank voor alle lessen die jullie me geleerd hebben.

De allocatiemedewerkers, Olga, en alle anderen bij Eurotransplant, dank voor alle mooie gesprekken en goede sfeer.

Mijn vrienden van de trip, vanaf het begin zat het goed, het is mooi zo'n breed palet aan mensen te kennen binnen ons vakgebied.

Hamëz, je werklust, talent en humor zijn onmisbare motivatoren geweest.

Mijn goede vrienden van Nexus, jullie karakters zijn goud waard, dank voor alle mooie tijden die waren en die nog komen.

Daan, vriend van het eerste uur, dank voor al je steun tijdens deze PhD. Samen hetzelfde pad doorlopen doet me nog steeds iedere dag deugd.

Mijn huisgenoten de laatste jaren, Sjors en Leon, dank voor de afleiding en verdieping die jullie me gaven.

Mijn ouders. Jullie trots en geloof in mij vormen een ononderbroken steun.

Mijn zussen, we zijn allemaal andere paden ingeslagen en daardoor is mijn leven zoveel rijker geworden, dank voor jullie geduld en warmte.

Fleur, licht van mijn leven. Je viert elke dag en leert me te genieten. Samen met jou kijk ik uit naar alles wat nog komt.

Curriculum vitae

Ben Goudsmit was born in Geleen on January 2nd, 1993. He grew up in Maastricht, where he graduated from gymnasium. From 2012 on, he studied medicine in Leiden.

During his studies, he gained an interest in research. In 2018, he graduated from medical school and started a combined function of PhD student at the Leiden University Medical Center and medical staff member at Eurotransplant International Foundation.

Ben currently lives and works in Den Haag.

List of publications

1. Alons IME, Goudsmit BFJ, Jellema K, van Walderveen MAA, Wermer MJH, Algra A. Response to Letter to the Editor Regarding" Yield of Computed Tomography (CT) Angiography in Patients with Acute Headache, Normal Neurological Examination, and Normal Non Contrast CT: A Meta-Analysis.". J Stroke Cerebrovasc Dis. 2018;27(7):2044-2045.

2. Alons IME, Goudsmit BFJ, Jellema K, van Walderveen MAA, Wermer MJH, Algra A. Prediction of vascular abnormalities on CT angiography in patients with acute headache. Brain Behav. Published online 2018. doi:10.1002/brb3.997

3. Alons IME, Goudsmit BFJ, Jellema K, van Walderveen MAA, Wermer MJH, Algra A. Yield of Computed Tomography (CT) Angiography in Patients with Acute Headache, Normal Neurological Examination, and Normal Non Contrast CT: A Meta-Analysis. J Stroke Cerebrovasc Dis. Published online 2017. doi: 10.1016/j.jstrokecerebrovasdis.2017.11.016

4. Goudsmit BFJ, Langeveld APM. Behandeling van een larynxfractuur met een combinatie van intraluminale stenting en externe fixatie. Ned Tijdschr voor Keel-neus-oorheelkd. 2018;24(2):54-59.

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. 2021;21(1):229-240.

6. Goudsmit BFJ, Putter H, Tushuizen ME, et al. Invited response to" MELD calibration". Am J Transplant. Published online 2020.

7. 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

8. 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.

9. 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. Am J Transplant. 2021;21(11):3583-3592.

10. Goudsmit BFJ, Putter H, van Hoek B. The Model for End-stage Liver Disease 3.0: an update without proven accuracy. Gastroenterology. Published online 2021.

11. 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;3(6):100369.

12. Collaborative C, Collaborative G. Timing of surgery following SARS-CoV-2 infection: an international prospective cohort study. Anaesthesia. 2021;76(6):748-758. doi:10.1111/anae.15458

13. CHOLECOVID Collaborative. An international multi-centre appraisal of the management and outcomes of acute CHOLEcystitis during the COVID-19 pandemic: The CHOLECOVID Study. Br J Surg. 2021;108(7). doi:10.11164/jjsps.5.2_381_2

14. Mes SD, Hendriksma M, Heijnen BJ, et al. Long-term voice outcomes of laryngeal framework surgery for unilateral vocal fold paralysis. Eur Arch Oto-Rhino-Laryngology. Published online 2021:1-9.