Quality in liver transplantation: perspectives on organ procurement and allocation

Boer, J.D. de

Chapter 7

Predictive capacity of risk models in liver transplantation

Jacob D. de Boer MD, Hein Putter PhD, Joris J. Blok MD PhD, Ian P.J. Alwayn MD PhD, Bart van Hoek MD PhD, Andries E. Braat MD PhD.

Transplantation Direct. 2019; 5(6):e457.

Abstract

Background
Several risk models to predict outcome after liver transplantation (LT) have been developed in the last decade. This study compares the predictive performance of 7 risk models.

Methods
Data on 62,294 deceased donor LTs performed in recipients ≥18 years old between January 2005 and December 2015 in the UNOS region were used for this study. The balance of risk (BAR), donor risk index (DRI), Eurotransplant-DRI, donor-to-recipient model (DRM), simplified recipient risk index (sRRI), Survival Outcomes Following Liver Transplantation (SOFT) and donor Model for End-stage Liver Disease (d-MELD) scores were calculated, and calibration and discrimination were evaluated for patient, overall graft and death-censored graft survival. Calibration was evaluated by outcome of high-risk transplantations (>80th percentile of the respective risk score) and discrimination by concordance index (c-index).

Results
Patient survival at 3 months was best predicted by the SOFT (c-index: 0.68) and BAR score (c-index: 0.64) while the DRM and SOFT score had the highest predictive capacity at 60 months (c-index: 0.59). Overall graft survival was best predicted by the SOFT-score at 3-months (c-index: 0.65) and by the SOFT and DRM at 60-months follow-up (c-index: 0.58). Death-censored graft survival at 60-months follow-up is best predicted by the DRI (c-index: 0.59) and ET-DRI (c-index: 0.58). For patient- and overall graft survival, high-risk transplantations were best defined by the DRM. For death-censored graft survival, this was best defined by the DRI.

Conclusions
This study shows that models dominated by recipient factors have best performance for short-term patient survival. Models that also include sufficient donor factors have better performance for long-term graft survival. Death-censored graft survival is best predicted by models that predominantly included donor factors.

Introduction

Nearly 14,000 patients are currently on the liver transplantation (LT) waiting list in the US, and each year >10% of these patients die without a transplantation[1]. Optimal use and allocation of livers available for transplantation is therefore essential. Such 'optimal' allocation is however difficult to define. Currently, the majority of livers in the US and Europe are allocated according to the Model for End-stage Liver Disease (MELD) or models derived from the MELD score (e.g. MELD-Na)[2,3]. MELD is an objective score that includes 3 laboratory values of the recipient (creatinine, bilirubin and International Normalized Ratio (INR)), validated for the prediction of 3-month waiting list mortality[4,5]. Studies showed that it is less suitable to accurately predict outcome after transplantation[6].

A model to predict outcome after transplantation should include all relevant characteristics of the donor, the recipient and other relevant data relating to the transplantation. It would enable to objectify and quantify the impact of several risk factors and could have numerous other applications. Over the last decade, several models for donor quality, recipient quality or the combination have been developed. To predict outcome after LT, the Survival Outcomes Following Liver Transplantation (SOFT)[6], donor MELD (D-MELD)[7], Balance of Risk (BAR) score[8] have been developed. While these models incorporate donor, recipient and transplant characteristics, the Donor Risk Index (DRI)[9] and Eurotransplant-Donor Risk Index (ET-DRI)[10] include solely donor and transplant characteristics to measure donor and organ quality. The ET-DRI was developed and validated for the Eurotransplant region in 2012. Later on, the simplified Recipient Risk Index (sRRI) was developed[11]. Both the donor model (ET-DRI) and recipient model (sRRI), were combined to predict outcome based on the combination of significant donor, transplantation and recipient factors; the Donor to Recipient Model (DRM)[11]. Although all models predict 'outcome' after LT, there are several differences between them[12]. Most importantly, the considered endpoint varies.

This study aims to compare the predictive capacity of seven models on patient-, overall graft- and death-censored graft survival at different post-transplant follow-up periods after LT.

Methods

Data selection

This study used data on LTs from January 1st, 2005 till December 31st, 2015 from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the US, submitted by

the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. No ethical statement was required according to European guidelines and Dutch law. Follow-up data were available up to March 2017.

Study population

In the study period, 71,429 LTs were performed. All LTs in recipients <18 years old were excluded (n=6,201) as well as those performed with livers from living donors (n=2,347) and auxiliary transplanted livers (n=37). Any combinations of organs other than liver and kidney were also excluded (n=550). This resulted in 62,294 transplantations included in the analysis.

Calculation of the BAR, SOFT, DRI, DRM, D-MELD and maximum C-statistic

Variables incorporated in the respective models are shown in Table 1. Cold ischemic times were missing in 3% (n=1562) and were singly imputed with the median cold ischemic time (6.3h). Recipient body mass index (BMI) was missing in 1,552 cases and set at reference (BMI<30) for calculation of the SOFT score. Gamma-glutamyl transpeptidase (GGT) and 'rescue allocation' are required for calculation of the ET-DRI[10] but were not available in the dataset. Rescue allocation can be considered a fast-track allocation that is used in the Eurotransplant region for a "center-oriented" allocation after organs have not been accepted in "patient-oriented" allocation for medical or logistical reasons[13]. They were therefore set at reference (GGT<50 U/L and Rescue allocation 'no'). Then, BAR score, SOFT score, DRI, ET-DRI, sRRI, DRM and D-MELD scores were calculated for all transplantations as described before[6–11]. The maximal c-statistic was calculated for a *dynamic* model including all factors that were incorporated in either the BAR, SOFT, DRI, ET-DRI, sRRI, DRM or

D-MELD score. The model is considered *dynamic* because the effects of each factor were estimated for each timepoint (per month follow-up period) separately.

Table 1. Overview of all variables per risk model

Factor	D-MELD	BAR	DRI	ET-DRI	sRRI	DRM	SOFT
Donor							
Age	X	X	X	X		X	X
GGT				X (n/a)		X (n/a)	
Race			X				
Height			X				
Cause of death			X	X		X	X
Donation after circulatory death (DCD)			X	X		X	
Partial or Split			X	X		X	
Serum creatinin							X
Recipient							
Age		X			X	X	X
MELD-score at transplantation	X	X			X	X	X
Retransplantation		X			X	X	X
Life support pre-transplant		X					X
Sex					X	X	
Etiology of disease					X	X	
BMI							X
Encephalopathy pre-transplant							X
Portal vein thrombosis							X
Portal bleed within 48h pretransplant							X
Previous abdominal surgery							X
Ascites pre-transplant							X
Dialysis pretransplant							X
Pre-transplant status (IC, hospital, home)							X
Albumin							X
Transplant							
Location (local, regional, national)			X	X		X	X
Cold ischemia time		X	X	X		X	X
Rescue allocation				X (n/a)		X (n/a)	
Number of factors	**2**	**6**	**8**	**8**		**13**	**18**

Definitions

Primary outcomes were patient (1), overall graft (2) and death-censored graft survival (3) at follow-up periods of 3 months, 1 year and 5-year after transplantation. Patient survival (1) was defined as the time period between transplantation and patient death. Overall graft survival (2) was evaluated as non-death censored graft survival and was defined as the time period between transplantation and either date of graft failure or patient death, whichever occurred first. Death-censored graft survival (3) was defined as the time period between transplantation and date of graft failure (note that patients were censored when deceased). Graft failure was, as specified in the OPTN follow-up forms, not entered for patients that died as a result of some other factor unrelated to graft failure. The individual scores were used to define risk groups of transplantations using increments of 20% in the quantiles of risk scores. High-risk transplantations were arbitrarily defined as scores above 80th percentile according to the respective risk models.

Statistical analysis

Clinical characteristics were summarized by median and 25% and 75% interquartile ranges (IQR) and number and percentage (N/%) for respectively continuous and categorical variables. Numerical and categorical factors between groups were compared using Kruskall-Wallis and Chi-square tests. Predictive performance of all models was compared by the area under the ROC curve or 'c-statistic'[14]. This c-statistic was calculated monthly up to 5 years for all three considered endpoints. Calculated c-statistics of individual models were compared in a boot-strapped 1000-fold database. Subsequently, transplantations were stratified by risk groups per score to evaluate the discriminative ability. Outcome of transplantations was stratified by risk groups using increments of 20% in the quantiles of risk scores in Kaplan-Meier analyses. Survival rate and rate of graft loss in the high-risk transplantations (above 80th percentile) were compared per endpoint between the several scores at 5-year follow-up. For death-censored graft survival, censoring by death was accounted for as a competing risk when calculating cumulative incidences[15].

All analyses were performed with SPSS version 24 and R version 3.3.2. A p-value below 0.05 was considered statistically significant. All analyses were performed in collaboration with the Department of Biomedical Data Sciences, Leiden University Medical Center.

Results

Study population

In the study period, 62,294 performed LTs were included. Mean transplant follow-up was 5.5 years for patient survival. Demographics of donors, patients and transplantations

are shown in Table 2. Most notably, donors had a median age of 42 years old (IQR 26-54) and were transplanted with a median cold ischemic time of 6.3 hours (IQR 5-8). Approximately 10% of all donors had diabetes mellitus (DM) and about a third of all livers was shared either regionally (24%) or nationally (5%). Recipients had a median age of 56 years old and a median laboratory MELD score of 21 (IQR 14-30). Most recipients were transplanted for hepatitis C related disease (28%), followed by alcoholic cirrhosis (20%) or other causes of cirrhosis (17%).

Table 2. Study demographics (n= 62,294)

Donor factor	**Mean**	**Median**	**IQR**
Age (years)	41	42	(26-54)
Height (cm)	171	173	(165-180)
Weight (kg)	80	78	(67-91)
BMI	27	26	(23-30)
Cold ischemic time	6.8	6.3	(5-8)
	N	**%**	
Sex (Male)	37202	60%	
Donortype (DCD)	3262	5%	
Cause of death			
Anoxia	14452	23%	
CVA/Stroke	24226	39%	
Head trauma	22036	35%	
CNS Tumor	327	1%	
Other	1253	2%	
Donorrace			
White	49078	79	
Black	11232	18	
Other	1984	3	
Split (yes)	788	1	
Share			
Local	44402	71	
Regional	14968	24	
National	2924	5	
Diabetes			
0-5 years	2445	4	
6-10 years	1242	2	
>10 years	2400	4	

Table 2. Continued.

Donor factor	**Mean**	**Median**	**IQR**
Yes, duration unknown	701	1	
No or unknown	55506	89	
Recipient factor	**Mean**	**Median**	**IQR**
Age (years)	54	56	50-61
Height (cm)	172	173	165-180
Weight (kg)	84	82	70-96
BMI	28	28	24-32
Lab-MELD	22	21	14-30
	N	%	
Sex (Male)	41968	67	
Primary disease			
Metabolic	1331	2%	
Acute	2795	5%	
Cholestatic	4695	8%	
Alcoholic	12514	20%	
Malignant	7006	11%	
HBV	1673	3%	
HCV	17696	28%	
Other cirrhosis	10590	17%	
Other/unknown	3994	6%	
Race (SRTR)			
Asian	2810	5%	
Black	6264	10%	
White	52468	84%	
Other	752	1%	
Pre-transplant life support (yes)	5102	8%	
Ever approved for HCC exception (yes)	16764	27%	
Retransplantation (Yes)	4080	7%	
Last encephalopathy			
Grade 1-2	32586	52%	
Grade 3-4	7365	12%	
Previous Upper Abdominal Surgery (Yes)	24241	39%	
History of Portal Vein Thrombosis (Yes)	2733	4%	
Diabetes type (present)			

Table 2. Continued.

Donor factor	Mean	Median	IQR
1	1442	2%	
2	12418	20%	
Other	160	0.3%	
Type unknown	2625	4%	
Risk scores	**Mean**	**Median**	**IQR**
DRI	1.4	1.3	(1.1-1.6)
sRRI	2.4	2.2	(1.8-2.6)
ET-DRI	1.3	1.3	(1.0-1.5)
DRM	2.8	2.6	(2.1-3.4)
SOFT score	9.4	7.0	(4-13)
D-MELD score	901	782	(480-1218)
BAR score	8.9	8	(4-13)

Discrimination

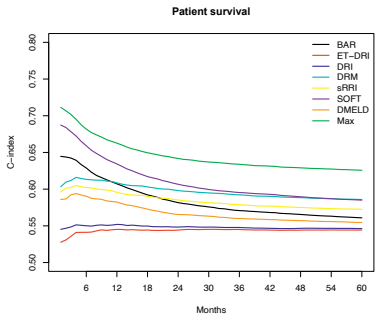
For the BAR, ET-DRI, DRI, DRM, sRRI, SOFT and D-MELD scores, the change in predictive capacity (c-index) is demonstrated over time and per outcome type. For patient survival this is shown in Figure 1a. In general, the ability to predict outcomes accurately, decreases over time. Therefore, outcome at short-term follow-up can be more accurately predicted than at longer follow-up. Patient survival at 3 months follow-up was best predicted by the SOFT score (c-index: 0.68, p<0.001) followed by the BAR (c-index: 0.64, p<0.001) and DRM-score (c-index: 0.61, p<0.001). From 3-year follow-up onwards, the SOFT score has a comparable performance to the DRM. The initial high performance of the BAR score decreases rapidly to below 0.6 at 18 months follow-up. Patient survival at 60 months follow-up was best predicted by the DRM and SOFT score (c-index: 0.59 for both, p=0.60). The maximal c-statistic for patient survival was higher at each time period than all other models (p<0.001). The model with all factors included, calibrated monthly, reached a c-statistic of 0.70 at 3 months follow-up and decreased gradually to 0.66 and 0.63 at 12- and 60-month follow-up, respectively.

To predict overall graft survival at short-term follow-up, highest predictive value at 3 months was also achieved by the SOFT score (c- index of 0.65, p<0.001), as is shown in Figure 1b. The BAR score and DRM performed reasonably when predicting overall graft survival at 3-month follow-up with c-indexes of 0.61 and 0.59, p=<0.001, respectively. Overall graft survival at 60-month follow-up, was again best predicted by the SOFT score and by the DRM with a similar c-index of 0.58 (p=0.22). A notable difference between these two models is the performance at short term; the SOFT score had an optimal performance at approximately 2 months post-transplantation whereas the DRM reached

a peak after 6 months. Performance of the other risk scores for overall graft survival stabilizes around a c-index of 0.56 after approximately 2 years. The maximal c-statistic for overall graft survival was 0.67 at 3-month follow-up and decreased to 0.65 and 0.62 at 12- and 60-months follow-up, respectively. These c-statistics were significantly higher than all other models at 3-, 12- and 60-month follow-up (p<0.001).

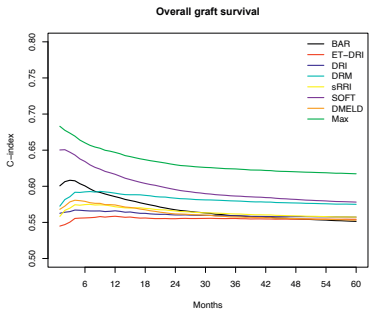
Death-censored graft survival showed a different picture; models that are dominated by donor factors like the DRI as well as the ET-DRI, had best predictive capability as from one year onwards, shown in Figure 1c. The DRI and ET-DRI achieved c-indexes at 12 months of 0.60 and 0.59 (p=0.01), respectively and at 60 months of 0.59 and 0.58 (p=0.16). The maximal c-statistic for death-censored graft survival was significantly higher as compared to each other model at the respective time points (p<0.001); it varied from 0.68 to 0.66 and 0.65 at 3-, 12- and 60-month follow-up, respectively.

A. Patient survival



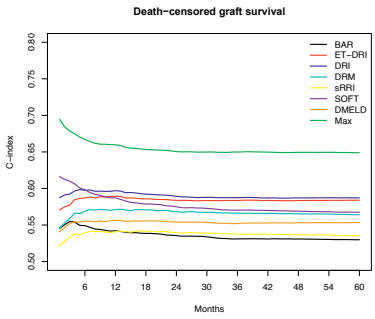
	3 months	1 year	5 years
BAR	0.64	0.61	0.56
ETDRI	0.54	0.55	0.54
DRI	0.55	0.55	0.55
DRM	0.61	0.61	**0.59**
sRRI	0.60	0.60	0.57
SOFT	**0.68**	**0.63**	0.59
DMELD	0.60	0.58	0.56
C-maximum	*0.70*	*0.66*	*0.63*

B. Overall graft survival



	3 months	1 year	5 years
BAR	0.61	0.59	0.55
ETDRI	0.55	0.56	0.54
DRI	0.57	0.57	0.56
DRM	0.59	0.59	0.58
sRRI	0.57	0.57	0.56
SOFT	**0.65**	**0.62**	**0.58**
DMELD	0.58	0.57	0.55
C-maximum	0.67	0.65	0.62

C. Death-censored graft survival



	3 months	1 year	5 years
BAR	0.56	0.54	0.53
ETDRI	0.58	0.59	0.58
DRI	0.59	**0.60**	**0.59**
DRM	0.56	0.57	0.56
sRRI	0.53	0.54	0.54
SOFT	**0.61**	0.59	0.57
DMELD	0.55	0.56	0.55
C-maximum	0.68	0.66	0.65

Figure 1. Performance of risk models

Calibration

As a measure of calibration, outcome of transplantations was stratified by risk groups defined by increments of 20% of the several risk models (Table 3). Lowest patient survival rate in high-risk transplantations was observed in the group defined by the DRM with a survival rate of 64% at 5-year follow-up (Figure 2). Patient survival stratified by other risk models is shown in supplementary figures 1A-F.

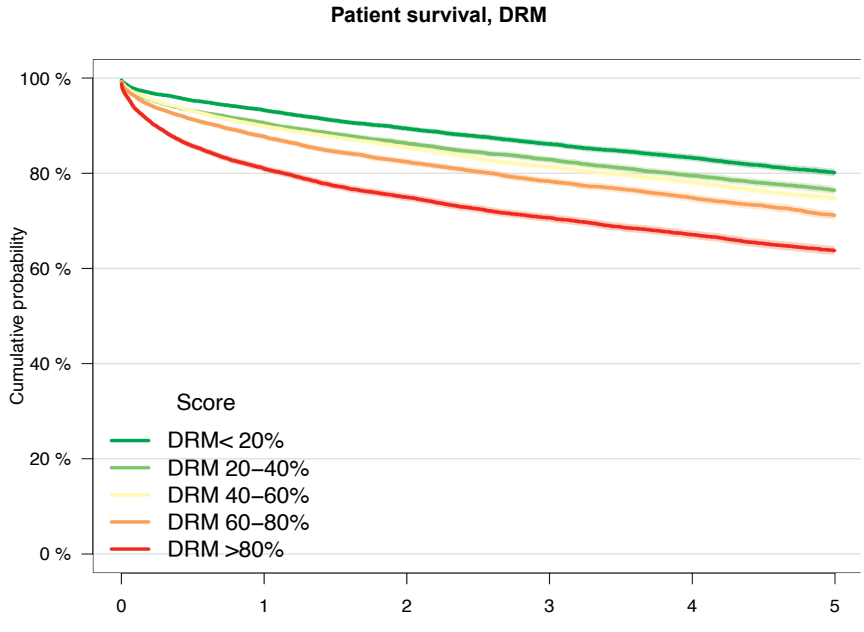
Table 3. Outcome by risk groups at 5-year follow-up.

	Patient survival (%)	**N at risk**	**Overall graft survival (%)**	**N at risk**	**Graft loss (%)**	**N at risk**
DRI						
<20%	77.7%	5432	76.4%	5320	6.9%	5320
20-40%	76.5%	5085	74.7%	4943	8.3%	4943
40-60%	72.9%	4839	70.5%	4655	10.2%	4655
60-80%	71.0%	4801	68.0%	4557	12.3%	4557
>80%	68.2%	4841	63.7%	4462	**14.9%**	**4462**
sRRI						
<20%	78.8%	5736	75.1%	5434	10.3%	5434
20-40%	76.2%	5219	73.6%	5000	9.3%	5000
40-60%	73.8%	5146	71.3%	4933	9.8%	4933
60-80%	71.5%	4876	68.9%	4677	11.4%	4677
>80%	66.0%	4021	64.3%	3893	11.7%	3893
ET-DRI						
<20%	77.5%	5529	75.9%	5394	7.5%	5394

Table 3. Continued.

	Patient survival (%)	N at risk	Overall graft survival (%)	N at risk	Graft loss (%)	N at risk
20-40%	76.4%	4724	74.7%	4590	7.7%	4590
40-60%	73.4%	5100	71.2%	4922	10.3%	4922
60-80%	70.6%	4774	67.3%	4522	12.4%	4522
>80%	68.6%	4871	64.4%	4509	14.5%	4509
DRM						
<20%	80.1%	5813	77.4%	5585	8.5%	5585
20-40%	76.4%	5227	73.5%	4984	9.7%	4984
40-60%	74.8%	5107	72.2%	4897	9.5%	4897
60-80%	71.1%	4728	68.6%	4540	11.2%	4540
>80%	**63.8%**	**4123**	**61.5%**	**3931**	13.7%	3931
SOFT						
<20%	77.7%	4297	75.4%	4139	8.6%	4139
20-40%	76.7%	4958	73.9%	4744	9.3%	4744
40-60%	75.6%	4987	72.7%	4760	10.1%	4760
60-80%	73.2%	6468	70.5%	6190	10.9%	6190
>80%	64.5%	4288	62.1%	4104	13.1%	4104
BAR						
<20%	77.0%	3461	74.3%	3319	9.3%	3319
20-40%	73.5%	5711	71.0%	5474	10.0%	5474
40-60%	75.9%	6748	72.5%	6401	11.2%	6401
60-80%	73.7%	4648	71.3%	4465	10.4%	4465
>80%	67.7%	4430	65.8%	4278	11.1%	4278
D-MELD						
<20%	76.8%	5225	74.8%	5071	8.0%	5071
20-40%	75.2%	5357	72.6%	5144	9.8%	5144
40-60%	74.5%	5164	71.9%	4942	10.4%	4942
60-80%	72.6%	4992	69.4%	4728	11.7%	4728
>80%	67.3%	4260	64.6%	4052	12.6%	4052

Values in bold indicate highest rate per outcome.

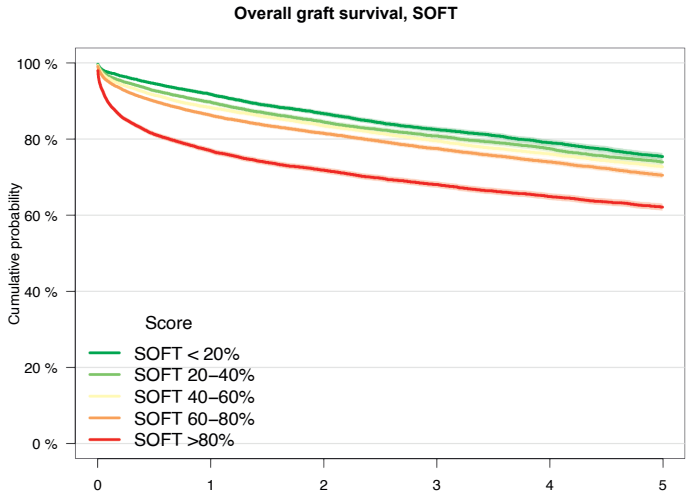


Score						
DRM< 20%:	12458	10997	9455	8104	6898	5813
DRM 20–40%:	12461	10663	8927	7551	6287	5227
DRM 40–60%:	12457	10592	8840	7389	6187	5107
DRM 60–80%:	12458	10327	8485	7042	5789	4728
DRM >80%:	12460	9546	7667	6186	5070	4123

Figure 2. Patient survival by DRM risk groups, Kaplan-Meier analysis

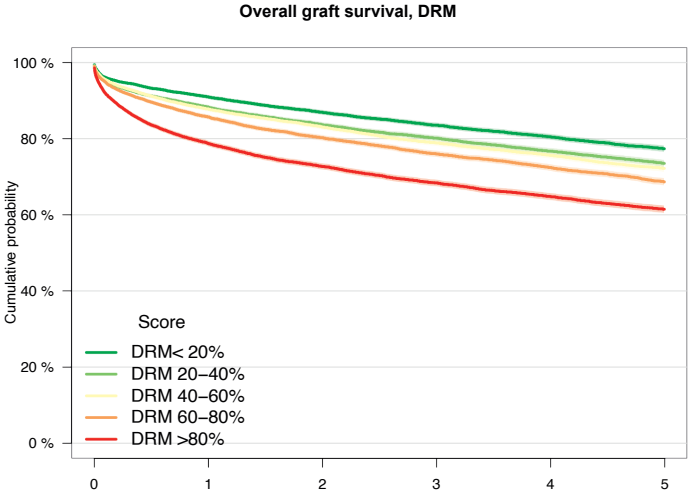
Also, for overall graft survival, lowest survival rate in high-risk transplantations was observed in the group defined by the SOFT (Figure 3) and by the DRM score with a survival rate of 62% (Figure 4).

Overall graft survival stratified by other risk models is shown in supplementary figures 2A-E. Death-censored graft survival was best predicted by models that were dominated by donor characteristics as the DRI and ET-DRI. In high-risk transplantations defined by these models, a graft loss rate of 15% was observed (Figure 5 and 6). Death-censored graft survival stratified by other risk models is shown in supplementary figures 3A-E.



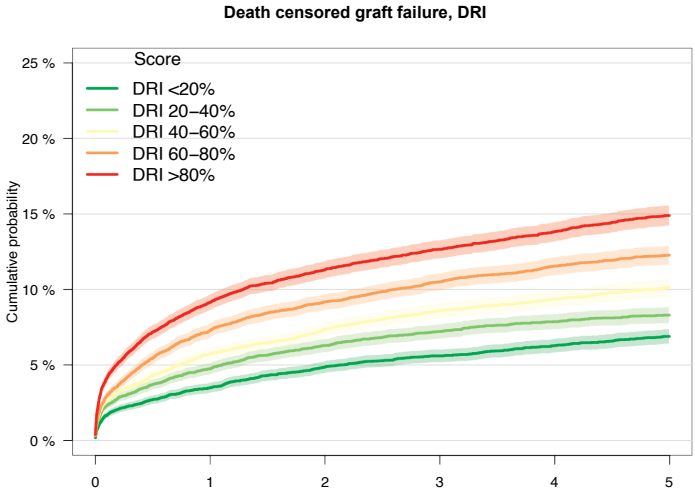
Score					
SOFT < 20%: 11149	9559	7806	6408	5173	4139
SOFT 20–40%: 11378	9659	8084	6838	5760	4744
SOFT 40–60%: 11386	9570	8018	6804	5691	4760
SOFT 60–80%: 15496	12701	10645	8914	7456	6190
SOFT >80%: 12885	9318	7496	6052	4989	4104

Figure 3. Overall graft survival by SOFT risk groups, Kaplan-Meier analysis



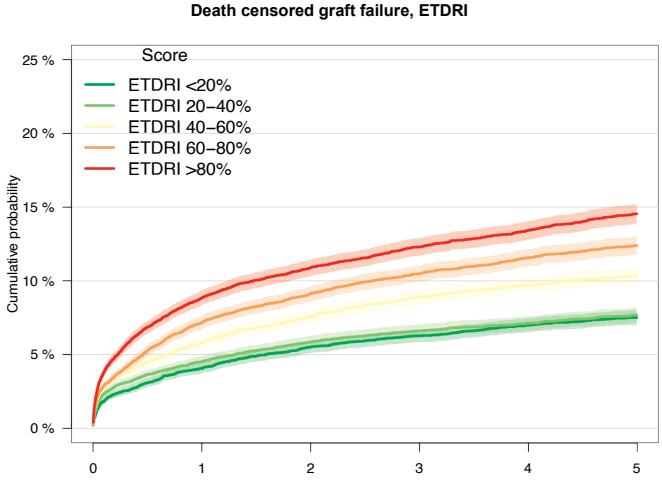
Score					
DRM< 20%: 12458	10722	9178	7838	6639	5585
DRM 20–40%: 12461	10390	8635	7268	6030	4984
DRM 40–60%: 12457	10330	8574	7131	5954	4897
DRM 60–80%: 12458	10091	8254	6824	5583	4540
DRM >80%: 12460	9274	7408	5955	4863	3931

Figure 4. Overall graft survival by DRM risk groups, Kaplan-Meier analysis



Score					
DRI <20%: 12456	10637	9018	7661	6421	5320
DRI 20–40%: 12463	10320	8526	7171	5973	4943
DRI 40–60%: 12470	10159	8376	6909	5698	4655
DRI 60–80%: 12447	9992	8179	6716	5513	4557
DRI >80%: 12458	9699	7950	6559	5464	4462

Figure 5. Death-censored graft survival by DRI risk groups, Kaplan-Meier analysis



Score					
ETDRI <20%: 12457	10588	9010	7688	6491	5394
ETDRI 20–40%: 12465	10228	8303	6811	5594	4590
ETDRI 40–60%: 12450	10182	8456	7105	5940	4922
ETDRI 60–80%: 12485	10023	8216	6786	5536	4522
ETDRI >80%: 12437	9786	8064	6626	5508	4509

Figure 6. Death-censored graft survival by ET-DRI risk groups, Kaplan-Meier analysis

Discussion

Predicting outcome after LT is important for issues varying from quality control to decision-making for liver offers. It could even be important for improving allocation algorithms. Therefore, several prediction models have been proposed in the last decade. This study has evaluated their performance with SRTR data, when applied to patient-, overall graft- and death-censored graft survival at different post-transplant follow-up periods. Our results show that models that pre-dominantly constitute of recipient characteristics, have best performance at predicting (short-term) patient survival. Models that include a combination of donor and recipient characteristics, like the SOFT and DRM, have a better performance for predicting overall graft survival. Death-censored graft survival, is best predicted by a model that predominantly constitutes of donor factors, as in the DRI and ET-DRI.

To evaluate the efficacy of LT, overall graft survival might be the most suitable outcome measure. This endpoint covers patient mortality as well as survival of the graft, which is as important in the light of the current organ shortage. Both the DRM and SOFT score, that both include donor- and recipient characteristics, have the highest predictive value for this outcome at long-term follow-up (c-index of 0.60). However, highest overall predictive performance was observed for short-term patient survival. Both the SOFT and BAR score achieved c-indexes of 0.68 and 0.64, respectively, for predicting patient survival at 3-month follow-up.

Our results show that when the follow-up period increases, the accuracy of the prediction of post-transplant outcome decreases. This increasing uncertainty is most likely the result of the input for the models; the prediction is based on factors that are defined at the time of or just prior to the transplantation. The initial strong relation with short-term complications or early mortality after transplantation decreases rapidly after the transplantation. Issues like changes in therapy, unexpected events or medical compliance are therefore not taken into account. Models that predict short-term outcomes are therefore more likely to achieve higher c-indexes as compared to models that focus on long-term survival[16]. Our results also show that the performance of post-transplant outcome decreases when used for other endpoints than they were developed for. This applies to the respective outcome as well as the considered follow-up period.

The maximal c-indexes that can be achieved by incorporating all factors of the respective models are promising and indicate that current models may be further improved. It is to be noted that in these *maximum* models, the effects of each factor are calibrated for each timepoint separately. The SRTR has made an effort to do so by analyzing their entire dataset and all variables[17]. They have developed models for patient- and

overall graft survival at 1 and 3-year follow-up. These four models include between 40 and 48 factors and incorporate between 165 and 204 coefficients[17]. They are updated periodically and can be used to correct center-specific outcomes[18]. Although the extent of the data and analyses are impressive, the number of coefficients and the required data pose challenges for other transplant organizations to use them. The 1-year SRTR models for patient- and graft survival in adults achieved a c-indexes at 1-year follow-up of 0.677 and 0.664, respectively (data SRTR)[19].

Our results are in line with published results on the performance of all models when they are applied to their initial endpoints. For patient survival at 90 days follow-up, the SOFT score has a reported predictive capacity of 0.7[6,8] (c-index of 0.68 in this study) and the BAR score of 0.66-0.74[8,20–25] (c-index of 0.64 in this study). In one study a c-index of 0.8 was reported for both the BAR and SOFT score[26]. The D-MELD was also developed for patient survival. It has a relatively low reported predictive capacity, most likely because of its simplicity and because it is often applied on short term outcomes[8,23,24,27–29]. To predict graft survival at long term follow-up, the DRM model has been developed in the Eurotransplant region. It has a reported c-index of 0.62 to predict 5-year graft survival[11] in the Eurotransplant database (c-index of 0.58 in this study). In calculating the DRM, GGT and rescue allocation were not available and were therefore set at reference in this study. Most likely, the c-index would higher if these factors had been available to get a more accurate DRM value. Models that solely include donor factors like the ET-DRI and DRI provide a suboptimal predictive capacity for long-term overall graft survival when used without adjustment for recipient characteristics as indicated by a c-index below 0.6 [8,23,24,30–32]. These models however, have the best performance for predicting death-censored graft survival. Such donor models can therefore be considered as a measure for the quality of the organ itself.

We have chosen to validate the risk models in the UNOS database because it is the most complete and extensive database available. Therefore, most risk models could be calculated correctly except for the ET-DRI. The ET-DRI, also used for the DRM, contains two factors (Rescue allocation and GGT) that were not available. While most studies focus on patient survival at short-term follow-up, this study has analyzed patient-, overall graft- and death-censored graft survival with the follow-up period as a continuous variable. The findings from this study -an objective comparison of models in a large dataset - may be used as a reference to choose an appropriate model.

In comparing center-specific outcomes, risk models may be used to take potential differences in donor and recipient characteristics (case-mix) into account[18,33]. When outcomes of individual transplant centers are not adjusted for donor quality, available “high-risk” liver allografts are likely less used. Effects of a focus on absolute outcomes seems to be already more present in the US than in Europe; although utilization rates of available livers seem to be similar between both, the quality of transplanted livers

is not[34–36]. European transplant centers tend to accept livers that have a higher mean donor age and have more co-morbidities on average[37,38].

Besides an application in evaluating center-specific outcomes, risk models could also have great value for improving allocation algorithms. The modest discriminative accuracy of risk prediction models is currently the most important concern[22,39]. It is important to note that c-statistics represent the accuracy of a model to predict in what order individual patients will experience an event. Models may therefore have limited use for individual patients but might define risk factor strata very well. Such findings have been published for the widely used Gail model for breast cancer. It is reported to have a modest discriminatory accuracy (c-index of 0.58) but a good fit in the dataset[40,41]. Currently, liver allocation in the US and Europe is performed using the (Na-)MELD score[3]. This algorithm does not take into account outcome after transplantation. Models for outcome after LT could therefore increase the overall survival benefit[42] by balancing the estimated post-transplantation outcome with the expected outcome on the waiting list by the MELD score[43]. For LT, the risk models may not be perfect but they might represent the most accurate objective prediction of outcome that is currently available. Therefore, incorporating estimated survival at 3 months follow-up (with a c-statistic over 0.7) might provide a good start. We should however strive to further improve the performance of these models. This might be done by including more direct (bio) data. Such data may become available with the introduction of machine perfusion[20,44]. Also a more detailed characterization of patients may be incorporated, for example by including the frailty index or the degree of sarcopenia[45–48].

Conclusions

This study has validated the performance of 7 risk models in perspective of different LT endpoints. The accuracy of predicting posttransplant outcome decreases when the follow-up period increases. Models dominated by recipient variables, have best performance for predicting short-term patient survival. Overall graft survival is best predicted by the DRM and SOFT score, models that combine donor and recipient characteristics. The DRI and ET-DRI best predict death-censored graft survival and can therefore best describe donor quality.

Acknowledgements

The authors would like thank Mr. Bryn Thompson for all his help with the SRTR database. The authors of this manuscript have no conflicts of interest to disclose as described by the journal. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

References

1. Kim WR, Lake JR, Smith JM, et al. OPTN/SRTR 2016 Annual Data Report: Liver. *Am J Transplant*. 2018;18:172-253. doi:10.1111/ajt.14559.
2. 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.
3. 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.
4. Suzuki H, Bartlett ASR, Muiesan P, Jassem W, Rela M, Heaton N. High model for end-stage liver disease score as a predictor of survival during long-term follow-up after liver transplantation. *Transplant Proc*. 2012;44(2):384-388. doi:10.1016/j.transproceed.2011.11.013.
5. Desai NM, Mange KC, Crawford MD, et al. Predicting outcome after liver transplantation: utility of the model for end-stage liver disease and a newly derived discrimination function. *Transplantation*. 2004;77(1):99-106. doi:10.1097/01.TP.0000101009.91516.FC.
6. Rana A, Hardy MA, Halazun KJ, et al. Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation. *Am J Transpl*. 2008;8(12):2537-2546. doi:10.1111/j.1600-6143.2008.02400.x.
7. Halldorson JB, Bakthavatsalam R, Fix O, Reyes JD, Perkins JD. D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching. *Am J Transplant*. 2009;9(2):318-326. doi:10.1111/j.1600-6143.2008.02491.x.
8. Dutkowski P, Oberkofler CE, Slankamenac K, et al. Are There Better Guidelines for Allocation in Liver Transplantation? *Ann Surg*. 2011;254(5):745-754. doi:10.1097/SLA.0b013e3182365081.
9. Feng S, Goodrich NP, Bragg-Gresham JL, et al. Characteristics associated with liver graft failure: The concept of a donor risk index. *Am J Transplant*. 2006;6(4):783-790. doi:10.1111/j.1600-6143.2006.01242.x.
10. Braat AE, Blok JJ, Putter H, et al. The eurotransplant donor risk index in liver transplantation: ET-DRI. *Am J Transplant*. 2012;12(10):2789-2796. doi:10.1111/j.1600-6143.2012.04195.x.
11. Blok JJ, Putter H, Rogiers X, et al. Combined Effect of Donor and Recipient Risk on Outcome After Liver Transplantation: Research of the Eurotransplant Database. *LIVER Transplant*. 2015;21(12):1486-1493. doi:10.1002/lt.24308.
12. Blok, J.J.; Putter, H; Metselaar, H.J. ; Porte R.J.; Gonella, F.; De Jonge, J.; Van den Berg, A.P.; De Boer, J.D. ; Van Hoek, B; Braat, AE; Van der Zande J. Identification and validation of the predictive capacity of risk factors and models in liver transplantation over time. *Transplant Direct*. 2018.
13. Eurotransplant. Chapter 5 - ET Liver Allocation System (ELAS). *Eurotransplant Man*. 2018;February(Liver allocation). https://www.eurotransplant.org/cms/index.php?page=et_manual.
14. van Houwelingen, H.C. and Putter H. *Dynamic Prediction in Clinical Survival Analysis*. 1st ed. CRC Press; 2012.
15. Putter H, Fiocco M, Gekus RB. Tutorial in biostatistics: Competing risk and multi-state models. *Stat Med*. 2007;26(11):2389-2430. doi:10.1002/sim.2712.
16. Ma Y, Wang Q, Yang J, Yan L. Comparison of Different Scoring Systems Based on Both Donor and Recipient Characteristics for Predicting Outcome after Living Donor Liver Transplantation. *PLoS One*. 2015;10(9):e0136604. doi:10.1371/journal.pone.0136604.
17. SRTR. Risk adjustment models posttransplant outcomes. www.srtr/org/reports-tools/risk-adjustment-models-posttransplant-outcomes/. Accessed October 23, 2018.
18. Kasiske BL, Wey A, Salkowski N, et al. Seeking new answers to old questions about public reporting of transplant program performance in the United States. *Am J Transplant*. 2018. doi:10.1111/ajt.15051.
19. Wey A, Salkowski N, Kasiske BL, et al. The relationship between the C-statistic and the accuracy of program-specific evaluations. *Am J Transplant*. 2018. doi:10.1111/ajt.15132.

20. Golse N, Guglielmo N, El Metni A, et al. Arterial Lactate Concentration at the End of Liver Transplantation is an Early Predictor of Primary Graft Dysfunction. *Ann Surg*. 2018;XX(Xx):1. doi:10.1097/SLA.0000000000002726.
21. de Campos Junior ID, Stucchi RSB, Udo EY, Boin I de FSF. Application of the BAR score as a predictor of short- and long-term survival in liver transplantation patients. *Hepatol Int*. 2015;9(1):113-119. doi:10.1007/s12072-014-9563-3.
22. Schrem H, Platsakis AL, Kaltenborn A, et al. Value and limitations of the BAR-score for donor allocation in liver transplantation. *Langenbeck's Arch Surg*. 2014;399(8):1011-1019. doi:10.1007/s00423-014-1247-x.
23. 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.
24. Åberg F, Nordin A, Mäkisalo H, Isoniemi H. Who is too healthy and who is too sick for liver transplantation: External validation of prognostic scores and survival-benefit estimation. *Scand J Gastroenterol*. 2015;50(9):1144-1151. doi:10.3109/00365521.2015.1028992.
25. Jochmans I, Monbaliu D, Pirenne J. The Balance of Risk Score for Allocation in Liver Transplantation. *Ann Surg*. 2013;259(2):25902. doi:10.1097/SLA.0b013e3182a18086.
26. Conjeevaram Selvakumar PK, Maksimak B, Hanouneh I, Youssef DH, Lopez R, Alkhouri N. Survival outcomes scores (SOFT, BAR, and Pedi-SOFT) are accurate in predicting post-liver transplant survival in adolescents. *Pediatr Transplant*. 2016;20(6):807-812. doi:10.1111/petr.12770.
27. Schrem H, Reichert B, Frühauf N, et al. The Donor-Risk-Index, ECD-Score and D-MELD-Score all fail to predict short-term outcome after liver transplantation with acceptable sensitivity and specificity. *Ann Transplant*. 2012;17(3):5-13.
28. Briceño J, Cruz-Ramírez M, Prieto M, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: Results from a multicenter Spanish study. *J Hepatol*. 2014;61(5):1020-1028. doi:10.1016/j.jhep.2014.05.039.
29. Costabeber AM, Lionço LC, Marroni C, Zanotelli ML, Cantisani G, Brandão A. D-meld does not predict post-liver transplantation survival: A single-center experience from Brazil. *Ann Hepatol*. 2014;13(6):781-787.
30. Reichert B, Kaltenborn A, Goldis A, Schrem H. Prognostic limitations of the Eurotransplant-donor risk index in liver transplantation. *J Negat Results Biomed*. 2013;12(1):1-8. doi:10.1186/1477-5751-12-18.
31. Salgia RJ, Goodrich NP, Marrero JA, Volk ML. Donor factors similarly impact survival outcome after liver transplantation in Hepatocellular carcinoma and non-Hepatocellular carcinoma patients. *Dig Dis Sci*. 2014;59(1):214-219. doi:10.1007/s10620-013-2883-7.
32. Winter A, Féray C, Audureau E, et al. External validation of the Donor Risk Index and the Eurotransplant Donor Risk Index on the French liver transplantation registry. *Liver Int*. 2017;37(8):1229-1238. doi:10.1111/liv.13378.
33. Blok JJ, de Boer JD, Putter H, et al. The center effect in liver transplantation in the Eurotransplant region. *Transpl Int*. 2018. doi:10.1111/tri.13129.
34. Halazun KJ, Rana AA, Fortune B, et al. No country for old livers? Examining and optimizing the utilization of elderly liver grafts. *Am J Transplant*. 2018;18(3):669-678. doi:10.1111/ajt.14518.
35. Rana A, Sigireddi RR, Halazun KJ, et al. Predicting Liver Allograft Discard. *Transplantation*. 2018:1. doi:10.1097/TP.0000000000002151.
36. Blok JJ, Braat AE, Adam R, et al. Validation of the donor risk index in orthotopic liver transplantation within the Eurotransplant region. *Liver Transpl*. 2012;18(1):112-119. doi:10.1002/lt.22447.
37. Ghinolfi D, Lai Q, Pezzati D, De Simone P, Rreka E, Filipponi F. Use of Elderly Donors in Liver Transplantation. *Ann Surg*. 2017;XX(Xx):1. doi:10.1097/SLA.0000000000002305.
38. de Boer J., Blok JJ, Putter H, et al. Optimizing the use of geriatric livers for transplantation in the Eurotransplant region. How to deal with an ageing donor population? *Liver Transplant*. doi:10.1002/lt.25353.

39. Avolio AW, Halldorson JB, Burra P, Dutkowski P, Agnes S, Clavien PA. Balancing utility and need by means of donor-to-recipient matching: A challenging problem. *Am J Transplant*. 2013;13(2):522-523. doi:10.1111/ajt.12031.
40. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-1886. doi:10.1093/jnci/81.24.1879.
41. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. Model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358-366. doi:10.1093/jnci/93.5.358.
42. 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.
43. Magder LS, Regev A, Mindikoglu AL. Comparison of seven liver allocation models with respect to lives saved among patients on the liver transplant waiting list. *Transpl Int*. 2012;25(4):409-415. doi:10.1111/j.1432-2277.2012.01431.x.
44. Faitot F, Besch C, Battini S, et al. Impact of real-time metabolomics in liver transplantation: Graft evaluation and donor-recipient matching. *J Hepatol*. 2018;68(4):699-706. doi:10.1016/j.jhep.2017.11.022.
45. Englesbe MJ, Patel SP, He K, et al. Sarcopenia and mortality after liver transplantation. *J Am Coll Surg*. 2010;211(2):271-278. doi:10.1016/j.jamcollsurg.2010.03.039.
46. Lai JC, Feng S, Terrault NA, Lizaola B, Hayssen H, Covinsky K. Frailty predicts waitlist mortality in liver transplant candidates. *Am J Transplant*. 2014;14(8):1870-1879. doi:10.1111/ajt.12762.
47. Kahn J, Wagner D, Homfeld N, Müller H, Kniepeiss D, Schemmer P. Both sarcopenia and frailty determine suitability of patients for liver transplantation—A systematic review and meta-analysis of the literature. *Clin Transplant*. 2018;32(4). doi:10.1111/ctr.13226.
48. Hamaguchi Y, Kaido T, Okumura S, et al. Impact of quality as well as quantity of skeletal muscle on outcomes after liver transplantation. *Liver Transplant*. 2014;20(11):1413-1419. doi:10.1002/lt.23970