OPEN

Post–Liver Transplant Outcomes: A Comparative Study of 6 Predictive Models

Christof Kaltenmeier, MD,[1] Eishan Ashwat, BA,[1] Hao Liu, MD, PhD,[1] Charbel Elias, MD,[1] Amaan Rahman, BA,[1] Jason Mail-Anthony, BA,[1] Isabel Neckermann, BA,[1] Stalin Dharmayan, MD,[2] Andrew Crane, MD,[1] Godwin Packiaraj, MD,[1] Subhashini Ayloo, MD, MPH,[3] Armando Ganoza, MD, MBA,[1] Vikraman Gunabushanam, MD,[1] and Michele Molinari, MD, MHS[1]

Background. We compared the performance of the Liver Transplant Risk Score (LTRS) with the survival outcomes following liver transplantation (SOFT), pretransplant SOFT (P-SOFT), Balance of Risk Score (BAR), donor-age and model for end-stage liver disease (D-MELD), and Organ Procurement and Transplantation Network Risk Prediction Score (ORPS) for the prediction of 90-d mortality, 1-y mortality, and 5-y survival after first-time liver transplantation (LT). **Methods.** A retrospective analysis of the Scientific Registry of Transplant Recipients was conducted using data collected between 2002 and 2021. **Results.** A total of 82 696 adult LT recipients with a median age of 56 y were included. The area under the curve for 90-d mortality were 0.61, 0.66, 0.65, 0.61, 0.58, and 0.56 for the LTRS, SOFT, P-SOFT, BAR, D-MELD, and ORPS, respectively (all pairwise comparisons: $P < 0.05$). The area under the curve for 1-y mortality were 0.60, 0.63, 0.62, 0.59, 0.60, 0.57, and 0.59 for the LTRS, SOFT, P-SOFT, BAR, D-MELD, and ORPS, respectively (all pairwise comparisons: $P < 0.05$). The c-statistics for 5-y survival were not statistically significant among the models. For 90-d mortality, 1-y mortality, and 5-y survival, the correlation coefficients between the LTRS and P-SOFT (the 2 models requiring only preoperative parameters) were 0.90. 0.91, and 0.81, respectively ($P < 0.01$). **Conclusions.** None of the predictive models demonstrated sufficient precision to reliably identify LT recipients who died within 90 d and 1 y after LT. However, all models exhibited strong capabilities in perioperative risk stratification. Notably, the P-SOFT and LTRS models, the 2 models that can be calculated using only preoperative data, proved to be valuable tools for identifying candidates at a significant risk of poor outcomes.

(*Transplantation Direct* 2024;10: e1724; doi: 10.1097/TXD.0000000000001724.)

The aging of the general population,[1] the rising prevalence of obesity, and the implementation of the model for end-stage liver disease (MELD) have significantly increased the complexity of patients referred for liver transplantation (LT).[2-11] Despite these changes, the process of approving or rejecting patients for LT has remained unchanged and continues to rely mainly on the clinical intuition of transplant specialists.

The complex decision-making necessary to determine whether a patient is fit for LT is subjective, and the lack of consensus among transplant centers/providers results in significant heterogeneity in patient selection. The absence of easy-to-use instruments to stratify patients into clear risk categories represents a significant limitation that is well recognized by transplant providers.[12] Recent research has highlighted this issue by revealing that a significant proportion of patients excluded from LT as considered to be too high risk were subsequently transplanted with good outcomes at another center with more aggressive inclinations.[13]

Received 7 May 2024. Revision received 12 August 2024.

Accepted 4 September 2024.

[1] Department of Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA.

[2] Department of Surgery, Leicester General Hospital, Leicester, United Kingdom.

[3] Department of Surgery, Brown University, Providence, RI.

The authors declare no funding or conflicts of interest.

C.K. participated in study design, data analysis, and writing the article. E.A. participated in study design, data collection and analysis, and writing the article. H.L. participated in study design and writing the article and editing. C.E., A.R., J.M.-A., I.N., S.A., A.C., G.P., A.G., and V.G. participated in writing and editing the article. S.D. participated in study design and editing the article. M.M. participated in study design, data collection, data analysis, writing, and editing the article.

Supplemental digital content (SDC) is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.transplantationdirect.com).

Correspondence: Michele Molinari, MD, MHS, Department of Surgery, University of Pittsburgh Medical Center, 3459 Fifth Ave, N758, Pittsburgh, PA 15213. (molinarim@upmc.edu).

ISSN: 2373-8731

DOI: 10.1097/TXD.0000000000001724

Ideally, the decision to list or decline patients for LT should be based on clinical expertise and objective criteria to guarantee consistency and accountability in the decision-making process within the same group of transplant specialists and among different centers. Although the transplant community recognizes that there is a need for better ways to select patients for LT,[12,13] advances in this field have stalled, primarily due to the absence of simple and accurate instruments that could be applied before patients are listed for LT.[12,13]

Most existing models to predict perioperative mortality after LT are complex and require data that are usually not available during the referral phase or when patients are presented at selection meetings (eg, the quality of the donors, the duration of cold ischemia time [CIT], operative parameters).[5,6,8,12-18] Consequently, despite being well-known and validated models to predict posttransplant mortality, the survival outcomes following liver transplantation (SOFT), pretransplant SOFT (P-SOFT), Balance of Risk Score (BAR), donor age and MELD (D-MELD), and Organ Procurement and Transplantation Network Risk Prediction Score (ORPS)[6,15,17-19] are rarely used in clinical settings due to the above-mentioned limitations.

To address the lack of easy-to-use instruments to assist clinicians in stratifying patients into different risk groups before their listing, our group used machine learning methods to develop the Liver Transplant Risk Score (LTRS), a predictive model that requires only 5 parameters readily available when patients are referred for LT.[16,20,21] The LTRS includes the age of the patients, their body mass index (BMI), the MELD score calculated at the time of assessment, the need for dialysis, and history of diabetes, irrespective of their duration. The LTRS was validated for the prediction of 90-d mortality, 1-y mortality, and 5-y post-LT survival in a large cohort of recipients transplanted in the United States,[20] and more recently in Europe.[21]

During the development and validation of the LTRS, we observed that the values of the area under the curve (AUC) obtained using the new instrument were comparable with those of better known but more complex predictive models. However, a comparison of the performance of the LTRS with other models is lacking.[6,15,17-19] In the current study, our primary aim was to test the hypothesis of noninferiority of the LTRS versus other models. To do so, we compared the performance of the LTRS with the performance of the SOFT, P-SOFT, BAR, D-MELD, and ORPS for the prediction of 90-d and 1-y mortality, and 5-y patient survival using data collected on adults who underwent first-time LT in the United States.

PATIENTS AND METHODS

The data necessary for this study were extracted from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes parameters collected on all donors, waitlisted candidates, and transplant recipients in the United States and submitted by the members of the Organ Procurement and Transplantation Network. The Department of Health and Human Services provides oversight of the activities of the Organ Procurement and Transplantation Network and SRTR contractors. Due to the anonymity of the variables, the need for individual recipient consent was waived by the institutional review board that approved this study (protocol No.: PRO 13060220). All the procedures and methods used followed the Declaration of Helsinki regarding ethical principles for medical research involving human participants and the Declaration of Istanbul on Organ Trafficking and Transplant Tourism,[22] and reporting followed the Strengthening of the Reporting of Observational Studies in Epidemiology guidelines.[23]

Study Design, Setting, Inclusion, and Exclusion Criteria

This is a retrospective study of consecutive adult recipients (aged 18 y and older) who underwent LT between January 1, 2002, and December 31, 2021, in the United States with at least 1 y follow-up. Exclusion criteria were the history of previous transplantation, the use of partial liver grafts, simultaneous multivisceral transplants, malignancies except for hepatocellular carcinoma, and ABO incompatibility.

Outcomes

The main aim was to compare the predictive function of the LTRS versus other models for 90-d postoperative mortality. Secondary aims were comparisons among the models for 1-y mortality and 5-y survival.

Data Collection and Model Calculation

The parameters necessary for the computation of the LTRS, P-SOFT, SOFT, BAR, ORPS, and D-MELD are outlined in Figure 1. The methods used for the calculation of the LTRS, P-SOFT, SOFT, BAR, D-MELD, and ORPS have been described in previous publications.[6,15-17,19]

Of all the predictive models, only the LTRS and P-SOFT could be calculated in the pretransplant settings as they do not require parameters associated with the quality of the donors or intraoperative measurements.

The P-SOFT necessitates recipient age, BMI, history of previous LT, previous abdominal surgeries, serum albumin level, need for dialysis, intensive care unit stay before LT, hospital admission before LT, MELD score at transplantation, need of life support before LT, history of hepatic encephalopathy, presence of portal vein thrombosis, and presence of ascites.

SOFT integrates the P-SOFT value with additional factors, including donor age, donor cause of death, donor serum creatinine, history of bleeding from portal hypertension during the 48 h preceding LT, CIT, and the use of nationally allocated grafts. BAR is calculated using recipient age, donor age, MELD score at transplantation, history of redo LT, need for life support before transplantation, and CIT. The D-MELD is determined by multiplying the donor age by preoperative recipient MELD scores. ORPS computation involves donor age, recipient age, history of diabetes, recipient positive hepatitis C viral serology, serum creatinine levels, and history of mechanical ventilation before surgery.

MELD scores were calculated using the formula by Kamath et al[10]: MELD = 3.78 × ln [serum bilirubin (mg/dL)] + 11.2 × ln [international normalized ratio] + 9.57 × ln [serum creatinine (mg/dL)] + 6.43[24] without exception points for patients with hepatocellular carcinoma. Recipient and donor BMIs were estimated using the World Health Organization formula BMI = body weight (kg)/height^2 (m^2).[25]

	LTRS	P-SOFT	SOFT	BAR	ORPS	D-MELD
Number of patients	82,696	82,482	76,757	82,696	76,983	82,696
Number of parameters needed for the model	5	13	19	6	6	3
A – Donor Parameters						
Age	-	-	x	x	x	x
Cause of death	-	-	x	-	-	-
Creatinine	-	-	x	-	-	-
B – Recipient Parameters						
Age	x	x	x	x	x	-
BMI	x	x	x	-	-	-
MELD at listing	x	x	-	-	-	-
MELD at transplant	-	-	x	x	-	x
Previous liver transplant	-	x	x	x	-	-
Previous abdominal surgery	-	x	x	-	-	-
Portal vein thrombosis	-	x	x	-	-	-
Ascites pretransplant	-	x	x	-	-	-
Albumin	-	x	x	-	-	-
Dialysis	x	x	x	-	-	-
Admitted to hospital pretransplant	-	x	x	-	-	-
Admitted to ICU pretransplant	-	x	x	-	-	-
Life support pretransplant	-	x	x	x	-	-
Encephalopathy	-	x	x	-	-	-
Diabetes	x	-	-	-	x	-
HCV	-	-	-	-	x	-
Creatinine	-	-	-	-	x	-
Portal bleed 48 hours before transplant	-	-	x	-	-	-
Mechanical ventilation	-	-	-	-	x	-
C- Transplant Parameters						
Cold ischemia time	-	-	x	x	-	-
National allocation	-	-	x	-	-	-
Risk categories	**n. points**	**n. points**	**n. points**	**n. points**	**n. points**	**n. points**
low risk	**0**	**0-5**	**0-5**	**0-5**	**0**	**0-399**
Intermediate risk	**1**	**6-15**	**6-15**	**6-10**	**1**	**400-799**
	2	**16-35**	**16-35**	**11-15**	**2**	**800-1199**
	3	**36-40**	**36-40**	**16-20**	**3**	**1200-1599**
High risk	**≥4**	**>40**	**>40**	**>20**	**≥4**	**≥1600**

FIGURE 1. Parameters necessary for the computation of different predictive models. BAR, Balance of Risk Score; BMI, body mass index; D-MELD, donor-age and model for end-stage liver disease; HCV, hepatitis C virus; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

Time-dependent variables were the date of transplantation, the date of retransplantation, the date of last follow-up or death, and CIT measured in hours. CIT was defined as the duration between the cross-clamp of the donors' aorta and the removal of the liver graft from cold storage.

Donor characteristics included age, sex, BMI, ethnicity/race, primary cause of death, and type of donation categorized as donation after cardiocirculatory death or donation after brain death.

Risk Groups

The study cohort was divided into different risk categories as described in previous studies.[6,15-17,19] Low-risk individuals were patients with LTRS ≤1, P-SOFT/SOFT or BAR = 0–10, ORPS ≤1, and D-MELD ≤799. High-risk individuals were patients with LTRS ≥4, P-SOFT/SOFT or BAR ≥16, ORPS ≥4, and D-MELD ≥1200. LT recipients were categorized as the intermediate risk group if they did not meet the low-risk or high-risk criteria.

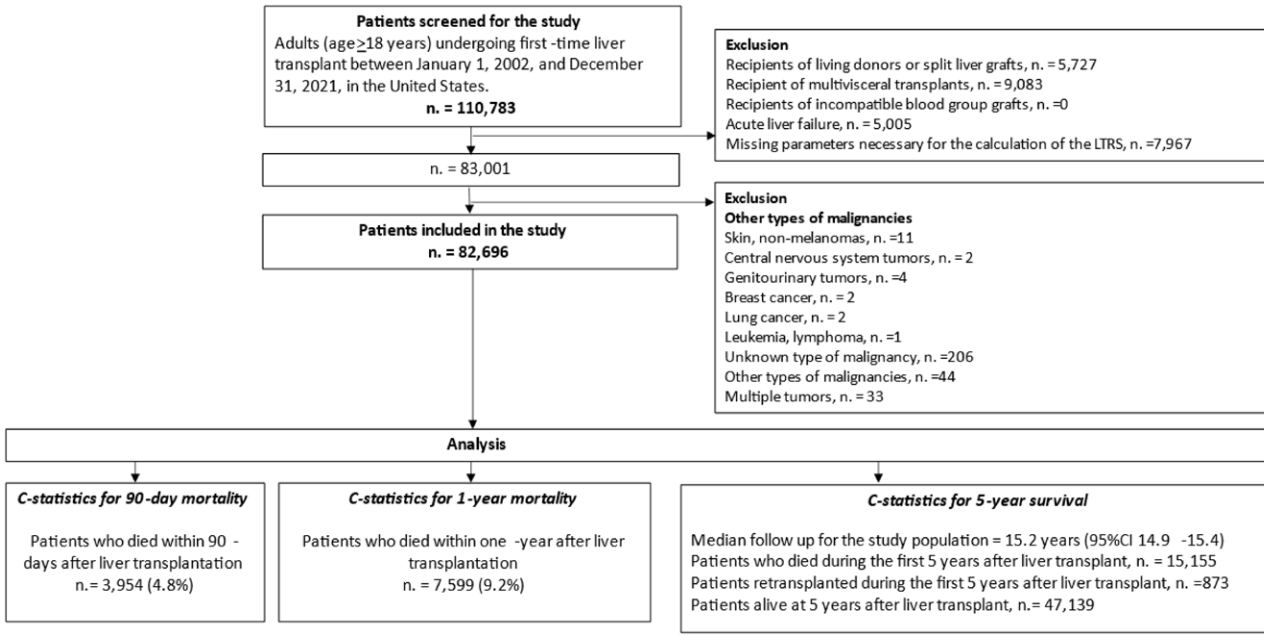


FIGURE 2. Flowchart of the study population. CI, confidence interval.

Statistical Analysis

Continuous variables are reported using the mean and SD, whereas data from nonnormally distributed variables are presented using the median and interquartile range (IQR). Categorical variables are expressed as frequencies and percentages. Univariate comparisons among risk groups were performed using the chi-square test for categorical variables and ANOVA or the Kruskal-Wallis test for continuous variables, as appropriate.

The predictive performance of each model was assessed using the area under the receiver operating characteristics curve with 95% confidence intervals (CIs; c-statistics).[26] To compare AUCs, we used the DeLong test for pairwise comparisons. This nonparametric statistical method assesses the difference between AUCs by accounting for the correlation between the curves. The DeLong test computes the standard errors based on the provided 95% CIs. Time-dependent c-statistics were assessed using Harrell concordance statistics.[27] To assess how well the predictive probabilities matched with the observed 90-d and 1-y mortality probabilities, calibration curves for each model were plotted. Positive predictive values, negative predictive values, and accuracies for 90-d and 1-y mortality were also calculated.

Survival analyses were conducted using the Kaplan-Meier method and Cox regression, with log-rank tests for comparison among risk groups. Censoring was applied for patients undergoing retransplantation, those lost to follow-up, or those alive at the end of the study.

Further analyses were performed to compare the performance of the LTRS and P-SOFT because they were the only 2 predictive models that could be calculated using only preoperative parameters. The correlation between the LTRS and the P-SOFT for 90-d, 1-y mortality, and 5-y survival were evaluated using scatter plots with respective correlation coefficients. Correlation coefficients R^2 ≥0.8 were deemed indicative of a high correlation.

Sensitivity analyses using different transplant eras were performed to assess whether there were significant differences in the performance of the models over time.

All analyses were conducted using SPSS Statistics for Windows, version 26 (IBM Corporation, Armonk, NY) or SAS version 9.4 (SAS Institute, Cary, NC) with Bonferroni corrected *P* values when appropriate. Two-tailed *P* values of <0.05 were considered statistically significant.

RESULTS

Cohort Characteristics

A total of 82 696 patients were included (Figure 2). The demographic and clinical characteristics of the study population are outlined in Table 1. The median age at the time of listing was 56 y (IQR, 49–62), with men comprising 66.8% of the cohort. The median MELD score was 19 (IQR, 13–28), and the primary indications for LT included viral hepatitis B or C (25.6%), alcoholic cirrhosis (20.1%), hepatocellular carcinoma (17.9%), and nonalcoholic steatohepatitis (8.9%). Regarding donor characteristics, the median donor age was 43 y (IQR, 27–55), 59.6% were men, and 94.4% were declared brain dead. Local sharing of organs occurred in 69% of cases, and the median CIT was 6.2 h (IQR, 4.9–8.0).

Performance of the Models for 90-d Mortality

The overall 90-d mortality of the study population was 4.8%. Patients who died within 90 d exhibited significantly higher mean scores across all predictive models: LTRS (1.4 versus 1.0; $P < 0.01$), P-SOFT (11.9 versus 7.8; $P < 0.01$), SOFT (13.2 versus 8.5; $P < 0.01$), BAR (10.5 versus 8.4; $P < 0.01$), ORPS (1.2 versus 1.0; $P < 0.01$), and D-MELD (1033 versus 870; $P < 0.01$). The AUC for the prediction of 90-d mortality was 0.61 for the LTRS compared with 0.66 for the SOFT ($P < 0.01$), 0.65 for the P-SOFT ($P < 0.01$), 0.61 for the BAR ($P = 0.10$), 0.58 for the D-MELD ($P = 0.02$), and 0.56 for ORPS ($P < 0.01$; Figure 3). Within each predictive model, a statistically significant difference was noted in 90-d mortality among low-, intermediate-, or high-risk groups (all

TABLE 1.

Patient, donor, and transplant characteristics of the study population

Characteristics	All patients N = 82 696	Patients who died within 90 d N = 3954 (4.8%)	Patients alive after 90 d N = 78 742 (95.2%)
Recipient			
Age, y, median (IQR)	56 (49–62)	57 (51–63)	56 (49–62)
Sex, n (%)			
Female/male	28 868 (33.2)/58,043 (66.8)	1442 (36.5)/2,512 (63.5)	25 442 (32.3)/53,300 (67.7)
BMI, kg/m^2, median (IQR)	27.6 (24.2–31.9)	28.2 (24.5–33.1)	27.7 (24.3–31.9)
MELD score at transplantation, median (IQR)	19 (13–28)	24 (15–35)	19 (13–29)
Indication for liver transplantation, n (%)			
Alcoholic cirrhosis	16 580 (20.1)	778 (19.7)	15 802 (20.1)
Hepatocellular carcinoma	14 766 (17.9)	552 (14.0)	14 214 (18.1)
Viral hepatitis B or C	21 128 (25.6)	938 (23.7)	20 190 (25.6)
Nonalcoholic fatty liver disease	7360 (8.9)	406 (10.3)	6954 (8.8)
Other indications	22 862 (27.6)	1280 (32.3)	21 582 (27.4)
Diabetes, n (%)	21 584 (24.8)	1114 (28.2)	19 550 (24.8)
Need for dialysis, n (%)	7258 (8.4)	647 (16.4)	6539 (8.3)
Portal vein thrombosis, n (%)	8588 (10.3)	587 (14.8)	8001 (10.1)
Previous abdominal surgeries, n (%)	38 939 (47.1)	2096 (53.0)	36 843 (46.8)
Admitted to hospital pretransplant, n (%)	14 655 (17.7)	780 (19.7)	13 875 (17.7)
Admitted to ICU pretransplant, n (%)	11 182 (13.5)	1199 (30.3)	9983 (12.6)
Life support pretransplant, n (%)	6459 (7.8)	832 (21.0)	5627 (7.1)
Encephalopathy	53 086 (64.2)	2877 (72.8)	50 209 (63.8)
Donor			
Age, y, median (IQR)	43.0 (27–55)	44.0 (28–56)	43 (27–55)
Sex, n (%)			
Female/male	33 438 (40.4)/49 258 (59.6)	1661 (42.0)/2293 (58.0)	31 777 (40.4)/46 965 (59.6)
Donation after circulatory death, n (%)	4687 (5.6)	274 (6.9)	4413 (5.6)
Cause of death, n (%)			
Anoxia	21 518 (26.0)	918 (23.2)	20 600 (26.2)
Cerebrovascular	31 353 (37.9)	1745 (44.1)	29 608 (37.6)
Trauma	27 608 (33.4)	1188 (30.0)	26 420 (33.6)
Other	2217 (2.6)	103 (2.6)	2114 (2.6)
Transplant			
Cold ischemia time, h, median (IQR)	6.2 (4.9–8.0)	6.8 (4.9–8.0)	6.1 (5.0–8.5)
Local share, n (%)	57 753 (69.8)	2590 (65.5)	55 163 (70.1)
Regional share, n (%)	24 934 (30.2)	1364 (34.5)	23 579 (29.9)

BMI, body mass index; ICU, intensive care unit; IQR, interquartile range; MELD, model for end-stage liver disease.

$P < 0.01$). In contrast, there was no statistically significant difference in the 90-d mortality rate for patients belonging to the same risk group across different predictive models (Figure 4).

Performance of the Models for 1-y Mortality

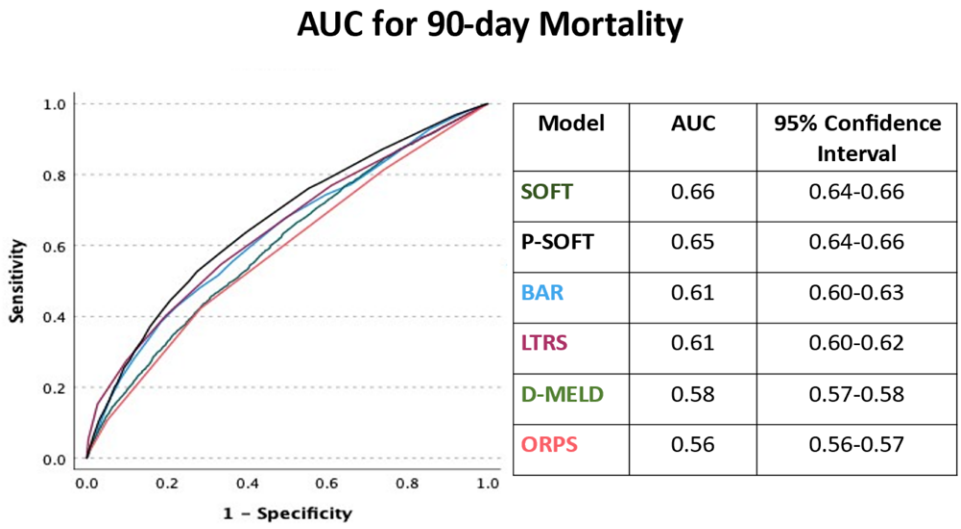
The overall 1-y mortality of the study population was 9.2%. Patients who died within the first year had significantly higher mean risk scores across all models: LTRS (1.3 versus 1.0; $P < 0.01$), P-SOFT (10.5 versus 7.7; $P < 0.01$), SOFT (11.7 versus 8.4; $P < 0.01$), BAR (9.7 versus 8.3; $P < 0.01$), ORPS (1.3 versus 1.0; $P < 0.01$), and D-MELD (1004 versus 865; $P < 0.01$). The AUC for the prediction of 1-y mortality was 0.60 for the LTRS compared with 0.63 for the SOFT ($P < 0.01$), 0.62 for the P-SOFT ($P < 0.01$), 0.59 for the BAR ($P = 0.10$), 0.59 for the ORPS ($P = 0.03$), and 0.57 for the D-MELD ($P < 0.01$; Figure 5). Within each predictive model, a statistically significant difference was noted in 1-y mortality among low-, intermediate-, or high-risk groups (all $P < 0.01$). In contrast, there was no statistically significant difference in 1-y mortality for patients belonging to the same risk group across all models, except for the high-risk group identified by the ORPS as they experienced a significantly higher mortality rate in comparison with the high-risk groups identified using other models ($P < 0.01$; Figure 6).

Performance of the Models for 5-y Survival

No statistically significant differences in the c-statistics were found for the prediction of 5-y survival among different models (Table 2). Within each model, there was a statistically significant difference in 5-y survival between low-, intermediate-, and high-risk groups (all P values <0.01; Figure 7). For patients belonging to the same risk group, 5-y survival rates were not statistically different among models, except for the high-risk group in the D-MELD because patients had significantly lower survival in comparison with high-risk groups identified using other models ($P < 0.01$).

LTRS Versus P-SOFT

Out of the 6 predictive models analyzed, the LTRS and the P-SOFT were the only 2 that could stratify patients into distinct risk categories based solely on their preoperative



Model	AUC	95% Confidence Interval
SOFT	0.66	0.64-0.66
P-SOFT	0.65	0.64-0.66
BAR	0.61	0.60-0.63
LTRS	0.61	0.60-0.62
D-MELD	0.58	0.57-0.58
ORPS	0.56	0.56-0.57

FIGURE 3. AUC for 90-d mortality and respective 95% confidence intervals for all the predictive models analyzed in this study. AUC, area under the curve; BAR, Balance of Risk Score; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

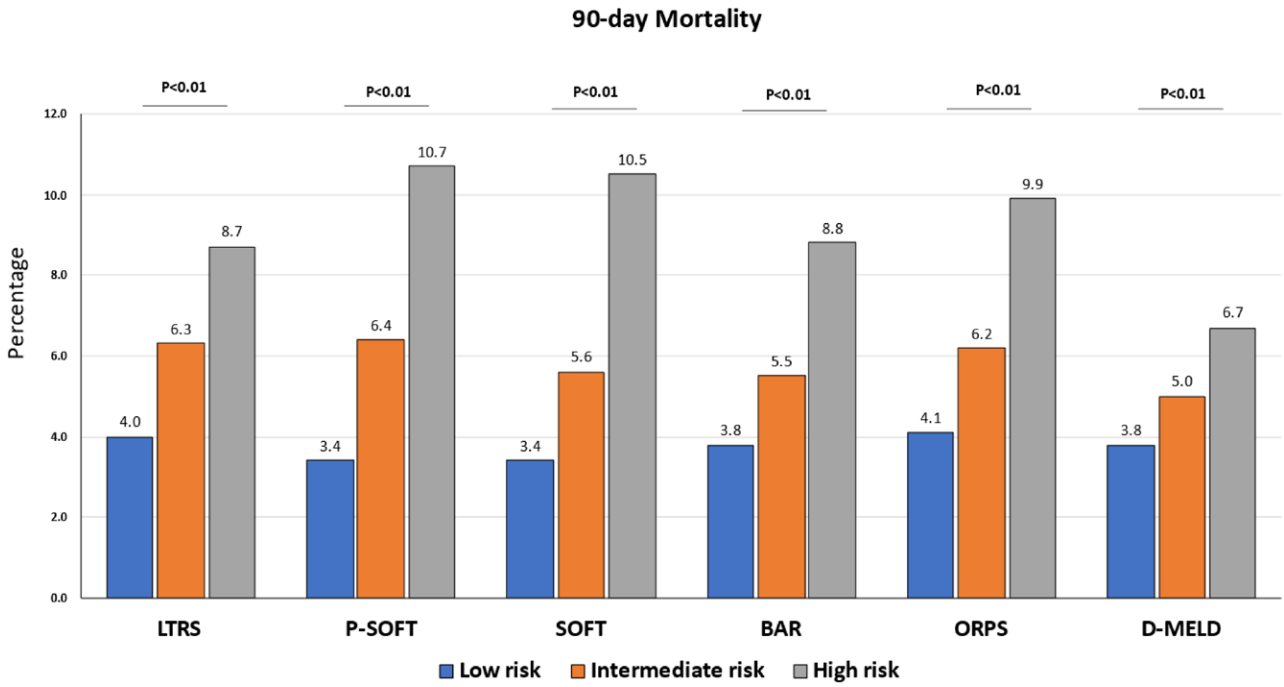
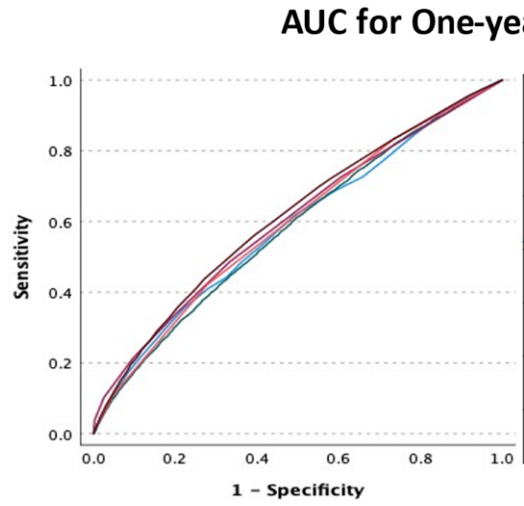


FIGURE 4. Observed 90-d mortality for different risk groups stratified by different predictive models. There were no statistically significant differences in observed 90-d mortality among patients within the same risk groups stratified by different predictive models. BAR, Balance of Risk Score; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

characteristics. The same risk classification proposed by Rana et al[19] during the development of P-SOFT was applied in the current study. Therefore, patients were classified as low-risk if their P-SOFT scores were ≤5, low-moderate risk when their scores were between 6 and 15, high-moderate risk if their scores were between 16 and 35, high risk when scores were between 36 and 40, and futile for those with a P-SOFT score >40. For the LTRS, patients were classified as low risk if their scores were 0 or 1, low-moderate risk when scores were 2, high-moderate risk when scores were 3 to 4, high risk when scores were 5–6, and futile when scores were ≥7.

90-d and 1-y Mortality Risk

The positive predictive value and sensitivity of the LTRS versus P-SOFT are reported in Table 3. The P-SOFT performed better than the LTRS in predicting patients who died within 90 d and 1 y after LT as shown by the higher precision-recall curves reported in Figure 8. However, the 2 models showed comparable performance in risk stratification with a high correlation coefficient for 90-d and 1-y mortality among patients belonging to the same risk group (correlation coefficient R^2 for 90-d mortality = 0.90; 95% CI, 0.88-0.91; $P < 0.01$; correlation coefficient R^2 for 1-y mortality = 0.91; 95% CI, 0.90-0.92; $P < 0.01$; Figure 9).



Model	AUC	95% Confidence Interval
SOFT	0.63	0.64-0.66
P-SOFT	0.62	0.61-0.62
BAR	0.59	0.59-0.60
LTRS	0.60	0.60-0.61
D-MELD	0.57	0.56-0.58
ORPS	0.59	0.59-0.59

FIGURE 5. AUC for 1-y mortality and respective 95% confidence intervals for all the predictive models analyzed in this study. AUC, area under the curve. AUC, area under the curve; BAR, Balance of Risk Score; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

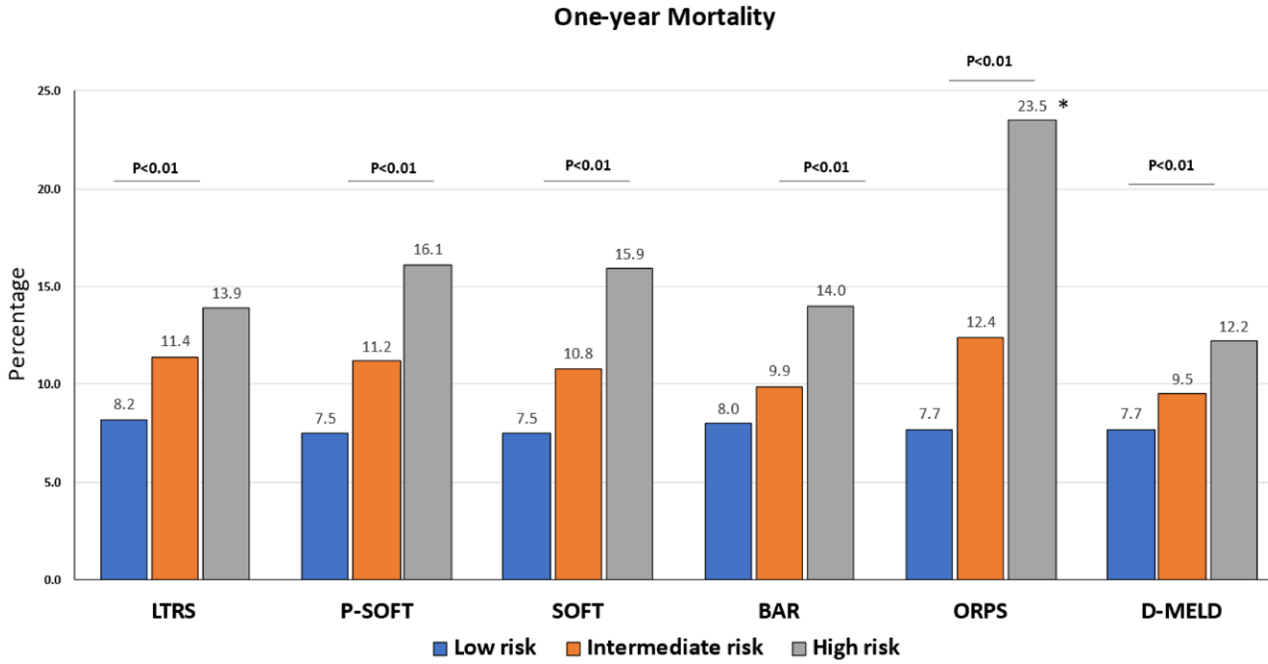


FIGURE 6. Observed 1-y mortality for different risk groups stratified by different predictive models. There were no statistically significant differences in observed 1-y mortality among patients within the same risk groups stratified by different predictive models except for patients identified as at high risk by the ORPS. ORPS, Organ Procurement and Transplantation Network Risk Prediction Score. BAR, Balance of Risk Score; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

Stratification of 5-y Survival

The 5-y survival rates for low-risk, moderate low-risk, moderate high-risk, high-risk, and futile groups were 79% versus 81%, 75% versus 76%, 73% versus 71%, 64% versus 36%, and 36% versus 24% for the P-SOFT and LTRS, respectively (Figure 10). There were no statistically significant differences in 5-y survival between similar risk groups identified by the 2 models except for patients belonging to the high-risk or futile categories. The correlation coefficient R^2 between the 2 models for the 5-y survival of patients within the same risk groups was 0.81 (95% CI, 0.79-0.83; $P < 0.01$). Notably, both the LTRS and P-SOFT identified patients belonging to the futile group as those with a 5-y survival rate of <50%, a value proposed by the scientific community as the threshold to define LT futile.

Sensitivity Analysis and Calibration

A sensitivity analysis was performed to evaluate the temporal stability of the model's predictive performance. The study period was stratified into 3 distinct eras: era I (January 1, 2002–December 31, 2007), era II (January 1, 2008–December 31, 2014), and era III (January 1,

TABLE 2.

Harrell concordance index for 5-y survival of all predictive models tested in the study

Model	Estimate	95% CI
LTRS	0.55	0.55-0.56
P-SOFT	0.56	0.55-0.56
SOFT	0.56	0.55-0.56
BAR	0.53	0.52-0.53
ORPS	0.58	0.57-0.58
D-MELD	0.54	0.54-0.55

Pairwise comparisons indicated that there were no statistically significant differences among the models.
Pairwise comparisons: all $P > 0.05$.
BAR, Balance of Risk Score; CI, confidence interval; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

2015–December 31, 2021). Within each era, we compared the area under the receiver operating characteristic curve (AUC) for each model to determine whether significant differences in predictive accuracy emerged over time. The results indicated no clinically significant variations in model performance across the eras, as detailed in **Tables S1–S4, SDC,** http://links.lww.com/TXD/A711). Furthermore, we assessed the calibration of the models by generating calibration curves for 90-d and 1-y mortality. These analyses revealed that all 6 models exhibited high correlation coefficients between predicted and observed mortality rates, demonstrating that the models were robustly calibrated across both time intervals (**Figures S1 and S2, SDC,** http://links.lww.com/TXD/A711).

DISCUSSION

In the current study, we compared the performance of 6 models designed to predict the outcomes of patients undergoing LT. Our findings revealed that both the SOFT and P-SOFT had the highest c-statistics for both 90-d and 1-y mortality compared with other existing models. However, the AUC values for all models consistently remained <0.7, indicating that their precision in identifying patients who die prematurely after LT is inadequate.

There are several explanations for the low c-statistics observed in the current study. One significant limitation is the evaluation of these predictive models exclusively on LT recipients rather than on all patients referred for LT. The exclusion of patients who were declined for LT introduces selection bias, which has certainly attenuated the value of the predictive models.

Another critical factor to consider is that the development of predictive models depends on the frequency of events. Because both 90-d and 1-y mortality rates after LT are relatively rare, occurring in only 4.8% and 9.2% of patients, respectively, predictive models were developed using data from a small subset of LT recipients, as the majority of patients survive postsurgery. This limited data set can hinder the accuracy and robustness of the models.

Furthermore, postoperative outcomes after LT are influenced by a myriad of factors, making it challenging to

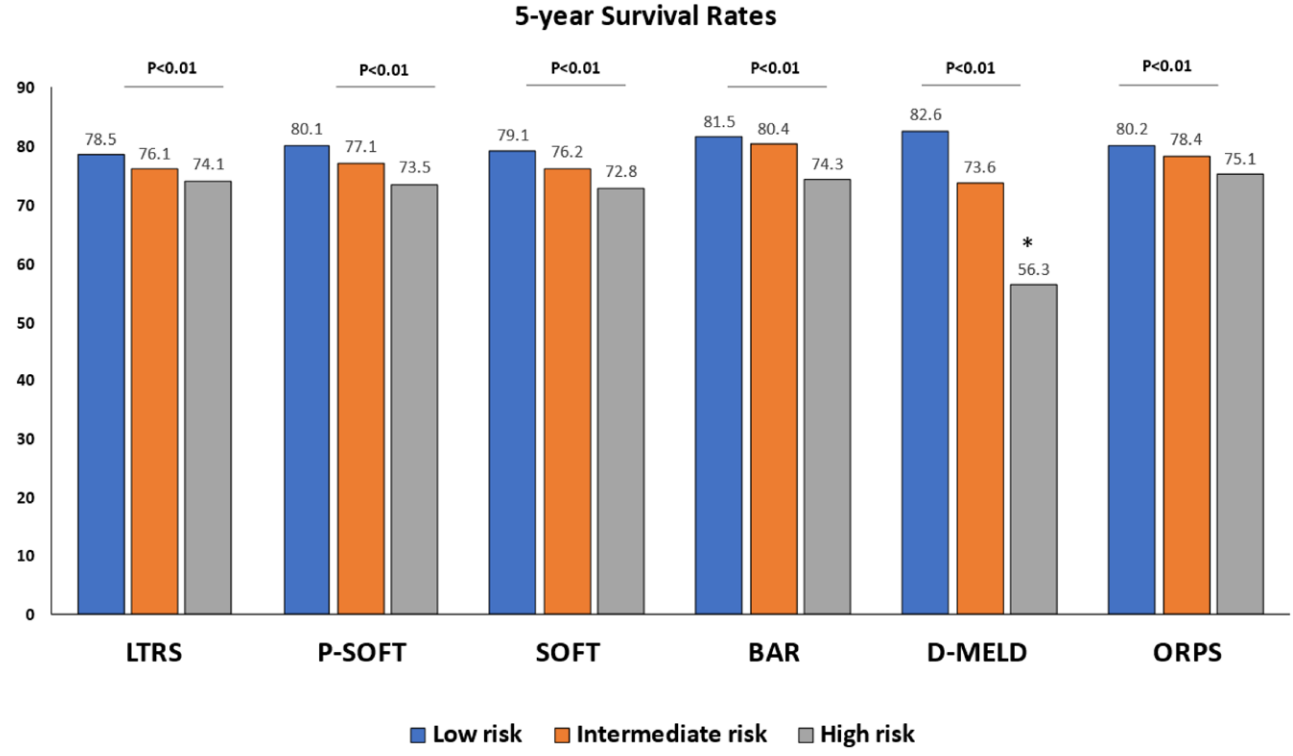


FIGURE 7. Five-year survival rates of patients belonging to distinct risk groups identified by different predictive models. No significant differences in 5-y survival rates were found among patients belonging to the same risk groups except for high-risk patients identified by the D-MELD. D-MELD, donor-age and model for end-stage liver disease. BAR, Balance of Risk Score; D-MELD, donor-age and model for end-stage liver disease; LTRS, Liver Transplant Risk Score; ORPS, Organ Procurement and Transplantation Network Risk Prediction Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

TABLE 3.

PPV, NPV, and sensitivity of the LTRS compared with the P-SOFT for 90-d and 1-y mortality after liver transplantation for patients belonging to different risk groups

		90-d mortality			1-y mortality		
Predictive model	**Risk group**	**PPV**	**NPV**	**Sensitivity**	**PPV**	**NPV**	**Sensitivity**
P-SOFT	Low risk	6.3%	97.2%	75.1%	11.2%	93.4%	83.0%
	Low-moderate risk	10.7%	95.0%	30.1%	16.1%	90.3%	23.2%
	High-moderate risk	17.1%	89.4%	13.0%	19.8%	83.9%	32.1%
	High risk	16.6%	83.3%	0.2%	25.0%	81.3%	0.5%
	Futile	18.0%	81.2%	0.1%	28.6%	75%	0.1%
LTRS	Low risk	6.8%	95.9%	38.4%	12.0%	91.8%	64.8%
	Low-moderate risk	8.7%	94.2%	23.0%	12.7%	89.2%	19.3%
	High-moderate risk	10.8%	92.5%	4.0%	17.5%	87.9%	9.4%
	High risk	16.1%	89.3%	1.0%	25.8%	82.7%	0.7%
	Futile	0.0%	83.8%	0.1%	50.0%	74.1%	0.1%

LTRS, Liver Transplant Risk Score; NPV, negative predictive value; PPV, positive predictive value; P-SOFT, pretransplant survival outcomes following liver transplantation.

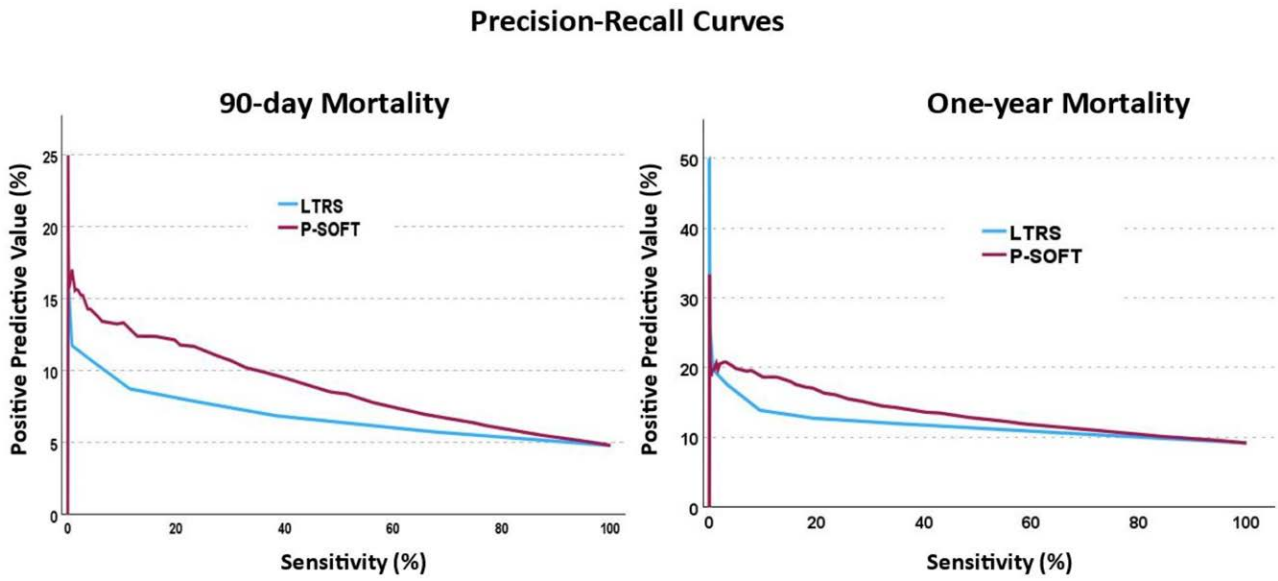


FIGURE 8. Precision-recall curves for 90-d and 1-y mortality observed using the LTRS and P-SOFT. The P-SOFT had a significantly higher precision-recall function in comparison with LTRS for both 90-d and 1-y mortality (all $P < 0.01$).). LTRS, Liver Transplant Risk Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

precisely identify patients who will experience poor outcomes. To address this complexity, more sophisticated predictive instruments are needed. These instruments should be capable of processing vast amounts of data, similar to the capabilities of current artificial intelligence technologies.

In situations where the outcomes of interest are unbalanced, conventional assessment metrics, such as the AUC, may not be the optimal instrument for evaluating predictive performance. Other metrics, such as precision-recall curves, are more suitable in those circumstances. In the current study, however, comparisons of the models were performed using AUCs because most providers are familiar with this metric. We also used precision-recall curves to assess the 2 P-SOFT and the LTRS, the 2 models that can be calculated in the preoperative settings. The head-to-head comparison of the P-SOFT and LTRS demonstrated that P-SOFT exhibited superior predictive capacity, as reflected by its higher precision-recall curves.

Despite their modest c-statistics, all predictive models provided a consistent stratification of patients into different risk groups. Therefore, we think that despite their limitations, these models should not be completely dismissed as clinically irrelevant. Their consistent performance in risk stratification suggests that current predictive instruments, such as the P-SOFT and the LTRS, could offer important information, especially in the preoperative settings, and may assist healthcare providers in determining more objectively the likelihood that a patient might experience undesired postoperative outcomes.

Although experienced healthcare providers can intuitively identify patients at elevated risk, heuristic approaches to selecting patients for LT have several limitations. Relying solely on clinical intuition introduces subjectivity and variability, leading to inconsistencies across transplant centers and providers.[13] Predictive models, although imperfect, offer a more systematic and data-driven approach, compelling providers to evaluate subtle yet crucial factors associated with postoperative outcomes, thus ensuring a more comprehensive assessment of patients' risk profiles. Furthermore,

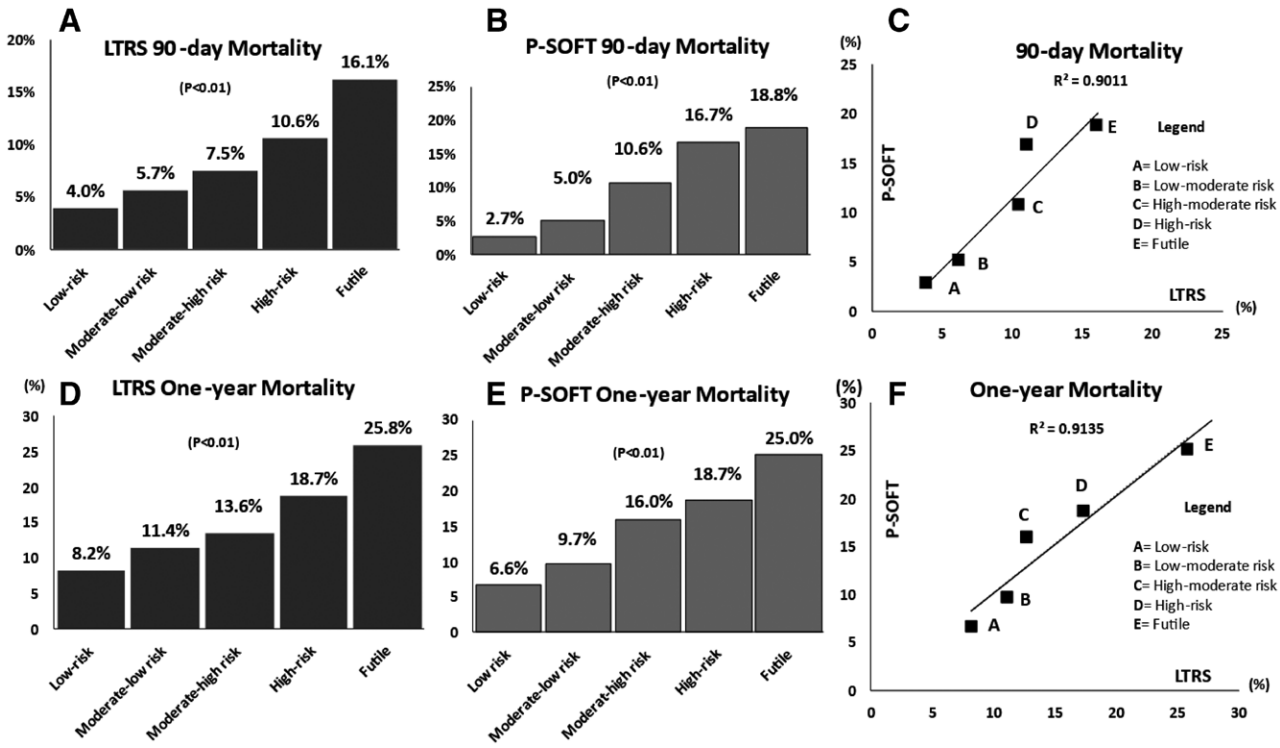


FIGURE 9. Observed 90-d mortality for patients belonging to different risk groups stratified by the LTRS (A). Observed 90-d mortality for patients belonging to different risk groups stratified by the P-SOFT (B). The correlation coefficient between the LTRS and P-SOFT for 90-d mortality (C). Observed 1-y mortality for patients belonging to different risk groups stratified by the LTRS (D). Observed 1-y mortality for patients belonging to different risk groups stratified by the P-SOFT (E). The correlation coefficient between the LTRS and P-SOFT for 1-y mortality (Panel F). LTRS, Liver Transplant Risk Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

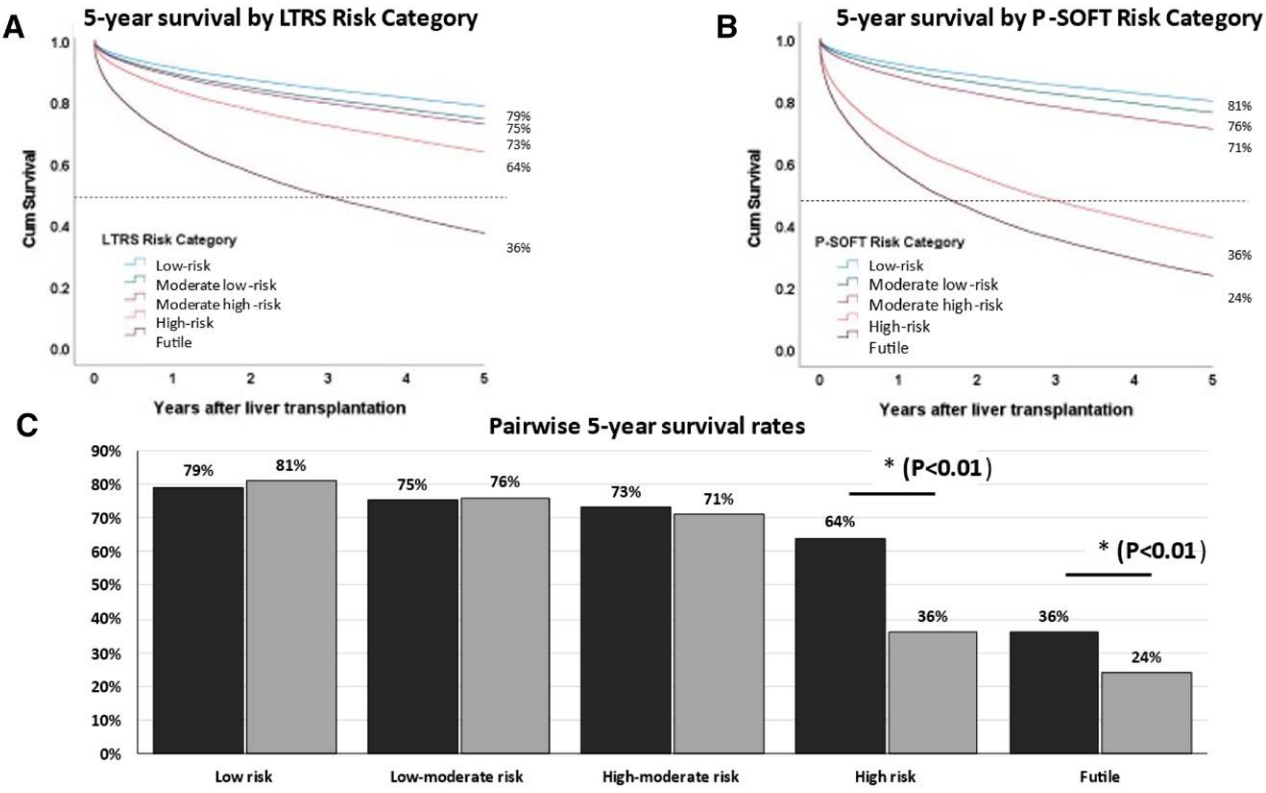


FIGURE 10. Five-year survival rates for different risk groups stratified by the LTRS (A) and by the P-SOFT (B). No statistically significant differences were found between patients belonging to the same risk group by the LTRS and by the P-SOFT except for the high-risk and futile groups (C). LTRS, Liver Transplant Risk Score; P-SOFT, pretransplant survival outcomes following liver transplantation.

the systematic use of predictive models in clinical settings enhances transparency and mitigates the effects of implicit biases.[13]

Both the P-SOFT and LTRS models have successfully identified LT recipients projected to have a 5-y survival rate <50%, a benchmark historically used to characterize LT as futile.[6]

Therefore, the application of predictive models in LT extends beyond individual patient care, providing valuable insights for healthcare resource planning and utilization. These models can inform quality improvement efforts, aiding transplant centers in optimizing their practices.

Importantly, predictive models can address the issue of unjustified risk aversion in some centers, which may lead to the unnecessary exclusion of patients from undergoing LT while also minimizing the risk of futile transplantations in others. On a broader scale, these models contribute to a comprehensive understanding of LT outcomes at regional or national levels. This knowledge can drive the identification of best practices, ultimately improving the efficiency and effectiveness of transplant programs to benefit patients and healthcare systems.

Strengths and Limitations

To the best of our knowledge, the current study is the only one to simultaneously analyze the performance of the LTRS, P-SOFT, SOFT, BAR, D-MELD, and ORPS for the prediction of the short- and long-term outcomes of LT recipients. Although this study plays an important role in determining which predictive models have the best performance, it is important to recognize its limitations. The retrospective study design introduces inherent biases and limitations associated with data collection. Additionally, the study primarily relies on data from the United States, potentially limiting the generalizability to patients transplanted in other countries. Furthermore, the focus of the study is on specific time points after LT, and it may not capture the dynamic nature of patient outcomes over an extended period.

In contrast, the current study has different strengths. First, the simultaneous analysis of several well-established predictive models applied to the same cohort of patients allowed us to assess their predictive profile without the risk of introducing confounders due to changes in clinical practices over time. In addition, the use of established metrics such as c-statistics provides a standardized and easily interpretable measure that can help clinicians when appraising the performance of these models. The large data set drawn from the United States also ensured a substantial sample size, enhancing the statistical power and reliability of our findings. Additionally, the focus on clinically relevant outcomes and the exploration of the futile transplantation category contributed valuable insights into the practical implications of these predictive models.

Future Developments

Future research will be necessary to refine current predictive models because these instruments have an overall limited ability to identify with precision which patients develop perioperative complications leading to their premature death after LT. With the increasing complexity of patients referred for LT, there is a need to improve our ability to consistently select patients using objective measures applicable to all transplant centers and providers. The development of dynamic models to accommodate evolving parameters and adapting to changing risk profiles has the potential to significantly improve our predictive capacity. One of the areas with a promising future is the integration of emerging technologies, such as artificial intelligence and machine learning, as they allow the processing in real time of very large data sets to predict post-LT outcomes. Yet, it is important to recognize that recent research[28] has indicated that even with the use of these more advanced methods, the c-statistics of predictive models obtained using machine learning algorithms remain in the range of 0.6–0.7, values that are comparable with the c-statistics observed using more "traditional" instruments compared in our study.

CONCLUSIONS

This study evaluated 6 models predicting 90-d mortality risk after LT. All models showed modest c-statistics, indicating limited precision in identifying patients at risk of early postoperative mortality. Nevertheless, they effectively stratified patients into distinct risk groups with clear short- and long-term outcomes. Despite its simplicity, the LTRS performed comparably to more complex models. These findings suggest that both P-SOFT and LTRS are valuable tools for enhancing objectivity and accountability in patient risk stratification during LT evaluations and multidisciplinary selection meetings.

ACKNOWLEDGMENTS

The authors acknowledge the SRTR for providing the data used in this study. The SRTR is a project of the Minneapolis Medical Research Foundation that has been supported by the Health Resources and Services Administration of the US Department of Health and Human Services under contract number 234-2005-370011C.

REFERENCES

1. Lai JC, Dodge JL, Kappus MR, et al; Multi-Center Functional Assessment in Liver Transplantation (FrAILT) Study. Changes in frailty are associated with waitlist mortality in patients with cirrhosis. *J Hepatol*. 2020;73:575–581.
2. Durand F, Levitsky J, Cauchy F, et al. Age and liver transplantation. *J Hepatol*. 2019;70:745–758.
3. Yi Z, Mayorga ME, Orman ES, et al. Trends in characteristics of patients listed for liver transplantation will lead to higher rates of waitlist removal due to clinical deterioration. *Transplantation*. 2017;101:2368–2374.
4. Burroughs AK, Sabin CA, Rolles K, et al; European Liver Transplant Association. 3-month and 12-month mortality after first liver transplant in adults in Europe: predictive models for outcome. *Lancet*. 2006;367:225–232.
5. Dutkowski P, Oberkofler CE, Bechir M, et al. The model for end-stage liver disease allocation system for liver transplantation saves lives, but increases morbidity and cost: a prospective outcome analysis. *Liver Transpl*. 2011;17:674–684.
6. Dutkowski P, Oberkofler CE, Slankamenac K, et al. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. *Ann Surg*. 2011;254:745–753.
7. Petrowsky H, Rana A, Kaldas FM, et al. Liver transplantation in highest acuity recipients: identifying factors to avoid futility. *Ann Surg*. 2014;259:1186–1194.
8. Rana A, Pallister ZS, Guiteau JJ, et al. Survival outcomes following pediatric liver transplantation (Pedi-SOFT) score: a novel predictive index. *Am J Transplant*. 2015;15:1855–1863.

9. Molinari M, Fernandez-Carrillo C, Dai D, et al. Portal vein thrombosis and renal dysfunction: a national comparative study of liver transplant recipients for NAFLD versus alcoholic cirrhosis. *Transpl Int*. 2021;34:1105–1122.
10. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001;33:464–470.
11. Coppel S, Mathur K, Ekser B, et al. Extra-hepatic comorbidity burden significantly increases 90-day mortality in patients with cirrhosis and high model for endstage liver disease. *BMC Gastroenterol*. 2020;20:302.
12. Fayek SA, Quintini C, Chavin KD, et al. The current state of liver transplantation in the United States: perspective from American Society of Transplant Surgeons (ASTS) scientific studies committee and endorsed by ASTS council. *Am J Transplant*. 2016;16:3093–3104.
13. Ravaioli M, Grande G, Di Gioia P, et al. Risk avoidance and liver transplantation: a single-center experience in a national network. *Ann Surg*. 2016;264:778–786.
14. 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;6:783–790.
15. Halldorson JB, Bakthavatsalam R, Fix O, et al. D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching. *Am J Transplant*. 2009;9:318–326.
16. Molinari M, Ayloo S, Tsung A, et al. Prediction of perioperative mortality of cadaveric liver transplant recipients during their evaluations. *Transplantation*. 2019;103:e297–e307.
17. Aloia TA, Knight R, Gaber AO, et al. Analysis of liver transplant outcomes for United Network for Organ Sharing recipients 60 years old or older identifies multiple model for end-stage liver disease-independent prognostic factors. *Liver Transpl*. 2010;16:950–959.
18. Rana A, Jie T, Porubsky M, et al. The survival outcomes following liver transplantation (SOFT) score: validation with contemporaneous data and stratification of high-risk cohorts. *Clin Transplant*. 2013;27:627–632.
19. 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 Transplant*. 2008;8:2537–2546.
20. Molinari M, Jorgensen D, Ayloo S, et al. Preoperative stratification of liver transplant recipients validation of the LTRS. *Transplantation*. 2020;104:e332–e341.
21. Ashwat E, Kaltenmeier C, Liu H, et al. Validation of the liver transplant risk score in Europe. *Br J Surg*. 2023;110:302–305.
22. World Medical Association. Ethics Unit. Declaration of Helsinki. 2007. www.wma.net/e/ethicsunit/helsinki.htm.
23. von Elm E, Altman DG, Egger M, et al; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370:1453–1457.
24. Malinchoc M, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology*. 2000;31:864–871.
25. World Health Organization. Obesity and Overweight. Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed November 8, 2024.
26. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.
27. Harrell FE Jr, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–2546.
28. Ivanics T, So D, Claasen M, et al. Machine learning-based mortality prediction models using national liver transplantation registries are feasible but have limited utility across countries. *Am J Transplant*. 2023;23:64–71.