JAMA Surgery | Original Investigation | PACIFIC COAST SURGICAL ASSOCIATION

Evaluation of Early Allograft Function Using the Liver Graft Assessment Following Transplantation Risk Score Model

Vatche G. Agopian, MD; Michael P. Harlander-Locke, MPH; Daniela Markovic, MS; Wethit Dumronggittigule, MD; Victor Xia, MD; Fady M. Kaldas, MD; Ali Zarrinpar, MD, PhD; Hasan Yersiz, MD; Douglas G. Farmer, MD; Jonathan R. Hiatt, MD; Ronald W. Busuttil, MD, PhD

IMPORTANCE Early allograft dysfunction (EAD) following a liver transplant (LT) unequivocally portends adverse graft and patient outcomes, but a widely accepted classification or grading system is lacking.

OBJECTIVE To develop a model for individualized risk estimation of graft failure after LT and then compare the model's prognostic performance with the existing binary EAD definition (bilirubin level of ≥10 mg/dL on postoperative day 7, international normalized ratio of ≥1.6 on postoperative day 7, or aspartate aminotransferase or alanine aminotransferase level of >2000 U/L within the first 7 days) and the Model for Early Allograft Function (MEAF) score.

DESIGN, SETTING, AND PARTICIPANTS This retrospective single-center analysis used a transplant database to identify all adult patients who underwent a primary LT and had data on 10 days of post-LT laboratory variables at the Dumont-UCLA Transplant Center of the David Geffen School of Medicine at UCLA between February 1, 2002, and June 30, 2015. Data collection took place from January 4, 2016, to June 30, 2016. Data analysis was conducted from July 1, 2016, to August 30, 2017.

MAIN OUTCOMES AND MEASURES Three-month graft failure–free survival.

RESULTS Of 2021 patients who underwent primary LT over the study period, 2008 (99.4%) had available perioperative data and were included in the analysis. The median (interquartile range [IQR]) age of recipients was 56 (49-62) years, and 1294 recipients (64.4%) were men. Overall survival and graft-failure-free survival rates were 83% and 81% at year 1, 74% and 71% at year 3, and 69% and 65% at year 5, with an 11.1% (222 recipients) incidence of 3-month graft failure or death. Multivariate factors associated with 3-month graft failure–free survival included post-LT aspartate aminotransferase level, international normalized ratio, bilirubin level, and platelet count, measures of which were used to calculate the Liver Graft Assessment Following Transplantation (L-GrAFT) risk score. The L-GrAFT model had an excellent C statistic of 0.85, with a significantly superior discrimination of 3-month graft failure–free survival compared with the existing EAD definition (C statistic, 0.68; $P < .001$) and the MEAF score (C statistic, 0.70; $P < .001$). Compared with patients with lower L-GrAFT risk, LT recipients in the highest 10th percentile of L-GrAFT scores had higher Model for End-Stage Liver Disease scores (median [IQR], 34 [26-40] vs 31 [25-38]; $P = .005$); greater need for pretransplant hospitalization (56.8% vs 44.8%; $P = .003$), renal replacement therapy (42.9% vs 30.5%; $P < .001$), mechanical ventilation (35.8% vs 18.1%; $P < .001$), and vasopressors (22.9% vs 11.0%; $P < .001$); longer cold ischemia times (median [IQR], 436 [311-539] vs 401 [302-506] minutes; $P = .04$); greater intraoperative blood transfusions (median [IQR], 17 [10-26] vs 10 [6-17] units of packed red blood cells; $P < .001$); and older donors (median [IQR] age, 47 [28-56] vs 41 [25-52] years; $P < .001$).

CONCLUSIONS AND RELEVANCE The L-GrAFT risk score allows a highly accurate, individualized risk estimation of 3-month graft failure following LT that is more accurate than existing EAD and MEAF scores. Multicenter validation may allow for the adoption of the L-GrAFT as a tool for evaluating the need for a retransplant, for establishing standardized grading of early allograft function across transplant centers, and as a highly accurate clinical end point in translational studies aiming to mitigate ischemia or reperfusion injury by modulating donor quality and recipient factors.

JAMA Surg. 2018;153(5):436-444. doi:10.1001/jamasurg.2017.5040
Published online December 20, 2017. Corrected on May 16, 2018.

Invited Commentary
page 444

Supplemental content

Author Affiliations: Dumont-UCLA (University of California, Los Angeles) Transplant and Liver Cancer Centers, Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles (Agopian, Harlander-Locke, Dumronggittigule, Kaldas, Zarrinpar, Yersiz, Farmer, Hiatt, Busuttil); Department of Biomathematics, UCLA, Los Angeles (Markovic); Department of Surgery, Faculty of Medicine, Siriraj Hospital, Mahidol University, Salaya, Thailand (Dumronggittigule); Department of Anesthesia, David Geffen School of Medicine at UCLA, Los Angeles (Xia).

Corresponding Author: Vatche G. Agopian, MD, Dumont-UCLA Transplant and Liver Cancer Centers, Department of Surgery, David Geffen School of Medicine at UCLA, 757 Westwood Plaza, Ste 8501-B, Los Angeles, CA 90095 (vagopian@mednet.ucla.edu).

The success and durability of a liver transplant (LT), unequivocally the gold-standard treatment for all patients with irreversible liver failure,[1,2] critically depend on the function of the liver allograft. *Early allograft dysfunction* (EAD), a term used to describe initial poor function of the transplanted liver, represents the clinical phenotype of severe ischemia-reperfusion injury due to a variety of recipient, donor, and perioperative factors.[3-5] Not only has EAD been shown to result in inferior graft and patient survival following both cadaveric[6-8] and living donor LT,[9] but it has also been associated with short-term and long-term renal impairment and increased use of resources.[10,11]

The prognostic importance of EAD is undeniable, but a universally accepted definition remains elusive. The earliest characterizations of EAD relied primarily on peak serum aminotransferase levels, which reflected graft injury,[12-14] and a subsequent functional definition proposed by Deschênes and colleagues[15] that incorporated posttransplant bilirubin level and prothrombin time to reflect the metabolic and synthetic functions of the allograft. In the Model for End-Stage Liver Disease (MELD) era,[16] Olthoff et al[6] have validated what is now the most widely used definition of EAD: the presence of 1 or more of 3 variables, including (1) a peak serum aspartate aminotransferase (AST) or alanine aminotransferase (ALT) level higher than 2000 U/L (to convert to microkatals per liter, multiply by 0.167) within the first 7 postoperative days (PODs), (2) a serum bilirubin level of 10 mg/dL or higher (to convert to micromole per liter, multiply by 17.104) on POD 7, and (3) an international normalized ratio (INR) of 1.6 or higher on POD 7.

Under this current definition, the binary categorization of all patients as either having or not having EAD fails to capture the continuum along which graft dysfunction manifests and results in an inability to discriminate patients with EAD who may experience early graft loss from most patients with EAD who may enjoy long-term graft failure–free survival. Recently, Pareja et al[17] reported on the Model for Early Allograft Function (MEAF) score (score range: 0 to 10, with higher scores indicating increasing hazard ratio [HR] for patient mortality at 3 months), which allows for the continuous grading of early allograft dysfunction and more accurately assesses graft function compared with the categorical classifications of EAD, but did not assess the model's accuracy in estimating graft failure.

With this study, we sought to develop a model—called the Liver Graft Assessment Following Transplantation (L-GrAFT)—based on posttransplant laboratory variables that allows for the individualized calculation of graft failure risk following LT and to compare the L-GrAFT's prognostic performance with the existing binary EAD definition and MEAF score.

Key Points

Question What are the important posttransplant variables associated with graft outcomes after primary liver transplant?

Findings In this single-center analysis of 2008 adult recipients of a primary liver transplant, measures of posttransplant aspartate aminotransferase and bilirubin levels, international normalized ratios, and platelet counts were highly associated with 3-month graft-failure–free survival, which led to the creation of a risk scoring model called the Liver Graft Assessment Following Transplantation (L-GrAFT), designed for the continuous grading of early allograft dysfunction following a liver transplant.

Meaning The L-GrAFT risk score may serve as a tool for the standardized grading of early allograft function in transplant centers and as a highly accurate clinical end point in translational studies that aim to improve donor allograft quality and liver transplant outcomes.

Methods

We retrospectively reviewed a prospectively maintained transplant database to identify all adult patients (≥18 years of age) who underwent a primary LT and had available data on posttransplant laboratory studies, including serum AST, ALT, and total bilirubin levels, INR, and platelet counts. All LTs were performed at the Dumont-UCLA Transplant Center of the David Geffen School of Medicine at UCLA from February 1, 2002, to June 30, 2015. The UCLA institutional review board approved the study and waived the patient consent requirement because of the retrospective, deidentified nature of the study. Data collection took place from January 4, 2016, to June 30, 2016. Data analysis was conducted from July 1, 2016, to August 30, 2017.

The primary objective of this study was to develop a model allowing for individualized estimation of graft failure (ie, need for a retransplant or death) following an LT and then to compare its prognostic performance to the existing binary EAD definition[6] (bilirubin level of ≥10 mg/dL on POD 7, INR of ≥1.6 on POD 7, or AST or ALT level of >2000 U/L within the first 7 PODs) and to the MEAF score ($ALT_{max.3POD}$, $INR_{max.3POD}$, $Bilirubin_{3POD}$).[17] Variables for analysis included LT recipient demographics (age; sex; primary end-stage liver disease diagnosis; and diabetes, hypertension, and coronary artery disease status), laboratory results (physiological MELD score and pretransplant and posttransplant serum AST, ALT, total bilirubin levels, INR, and platelet count), pretransplant acuity (requirement for hospitalization, renal replacement therapy, mechanical ventilation, and vasopressors), and donor and operative characteristics (age, sex, cold ischemia time [CIT], warm ischemia time [WIT], and operative transfusion of units of packed red blood cells [uPRBCs]).

Statistical Analysis

Continuous variables were summarized as median values and interquartile ranges (IQRs), and categorical variables were summarized as percentages. Graft failure–free survival and overall patient survival curves were computed using Kaplan-Meier methods and were compared using log-rank tests.

Laboratory studies in the first 10 post-LT days were summarized to identify the variables most associated with graft failure, including measures of overall graft injury (AST and ALT levels), serum bilirubin level, INR, and platelet count. These candidate variables for AST and ALT levels included the overall 10-day mean (measured as the area under the curve [AUC]), maximum post-LT value, and rate of normalization

(slope = unit change/d); for INR, the maximum post-LT INR and 10-day AUC; and for bilirubin level and platelet count, the total 10-day AUC, rate of normalization, change from pre-LT to last post-LT value, and post-LT days 7 and 10 values. All candidate laboratory measures were determined to be skewed on the original scale and, thus, were transformed to the logarithmic scale where they displayed a normal distribution.

Hazard rate ratios for time to graft failure were computed under a Cox regression model, allowing for interactions of each specified candidate variable with interval of follow-up time. The proportional hazards assumption for each variable was evaluated by plotting the HR for graft failure stratified by time interval of event (0-3 months, 3-6 months, 6-12 months, and >1 year). A model—the L-GrAFT—for calculating 3-month graft failure was developed using a logistic regression analysis based on the 18 post-LT candidate variables (eTable 1 in the Supplement). The initial model allowed for nonlinear effects using restricted cubic splines allowing for 3 knots (2 regression terms per variable), whereas the final multivariate model was selected using the backward stepwise procedure for variable selection and a liberal $P < .15$ as the retention criterion. The nonlinear effects observed on the log odds scale for the slope log AST, AUC log AST, and AUC log bilirubin were approximated by a quadratic curve and, hence, modeled by including quadratic terms to the logistic model.

Model accuracy was evaluated using receiver operating characteristic analysis under the logistic model, and the AUC was compared with the accuracy of the EAD definition and MEAF score, each of which was calculated for the UCLA cohort. The AUCs were compared between each pair of models nonparametrically using the method by DeLong et al.[18] Goodness of fit was assessed using the Hosmer-Lemeshow method. The C statistic for the L-GrAFT model was internally validated on 200 bootstrapped samples taken with replacement from the original data, with each bootstrapped sample containing the same number of observations as the original sample (n = 2008). A risk score was calculated for each patient as the weighted sum of the covariates, with weights equal to the parameter estimates (log odds ratio) under the final multivariate model. Five risk groups for 3-month graft failure were defined on the basis of risk score distribution, including (1) very low risk: 50th percentile or less with a risk score lower than −3.23; (2) low risk: greater than 50th percentile with a risk score of −3.23 or higher to the 90th percentile or less with a risk score lower than −1.18; (3) moderate risk: greater than 90th percentile with a risk score of −1.18 or higher to the 93.3 percentile or less with a risk score lower than −0.57; (4) moderate to high risk: greater than 93.3 percentile with a risk score of −0.57 or greater to the 96.6 percentile or less with a risk score lower than 1.3; and (5) high risk: greater than 96.6 percentile with a risk score greater than 1.3. Mean profiles for each laboratory measure were evaluated using the mixed-effects regression model for the overall cohort and by risk groups. Recipient, donor, and operative characteristics were compared among low-risk groups (1 and 2) and moderate-to-high risk groups (3-5) using the Wilcoxon rank sum test (continuous variables) or the Fisher or χ^2 test (categorical variables). Analyses were performed using SAS, version 9.4 (SAS Institute Inc), and R, version 3.0.2 (R Foundation for Statistical Computing).

Results

Of the 2021 patients who underwent a primary LT over the study period, 2008 (99.4%) had available perioperative data and were included in the analysis. Recipient, donor, and operative characteristics are shown in eTable 2 in the Supplement. Among the recipients, the median (IQR) age was 56 (49-62) years, and 1294 (64.4%) were men. (Note that data for some variables were missing for some patients. Thus, the numbers and percentages reported here are based on only patients with available information.)

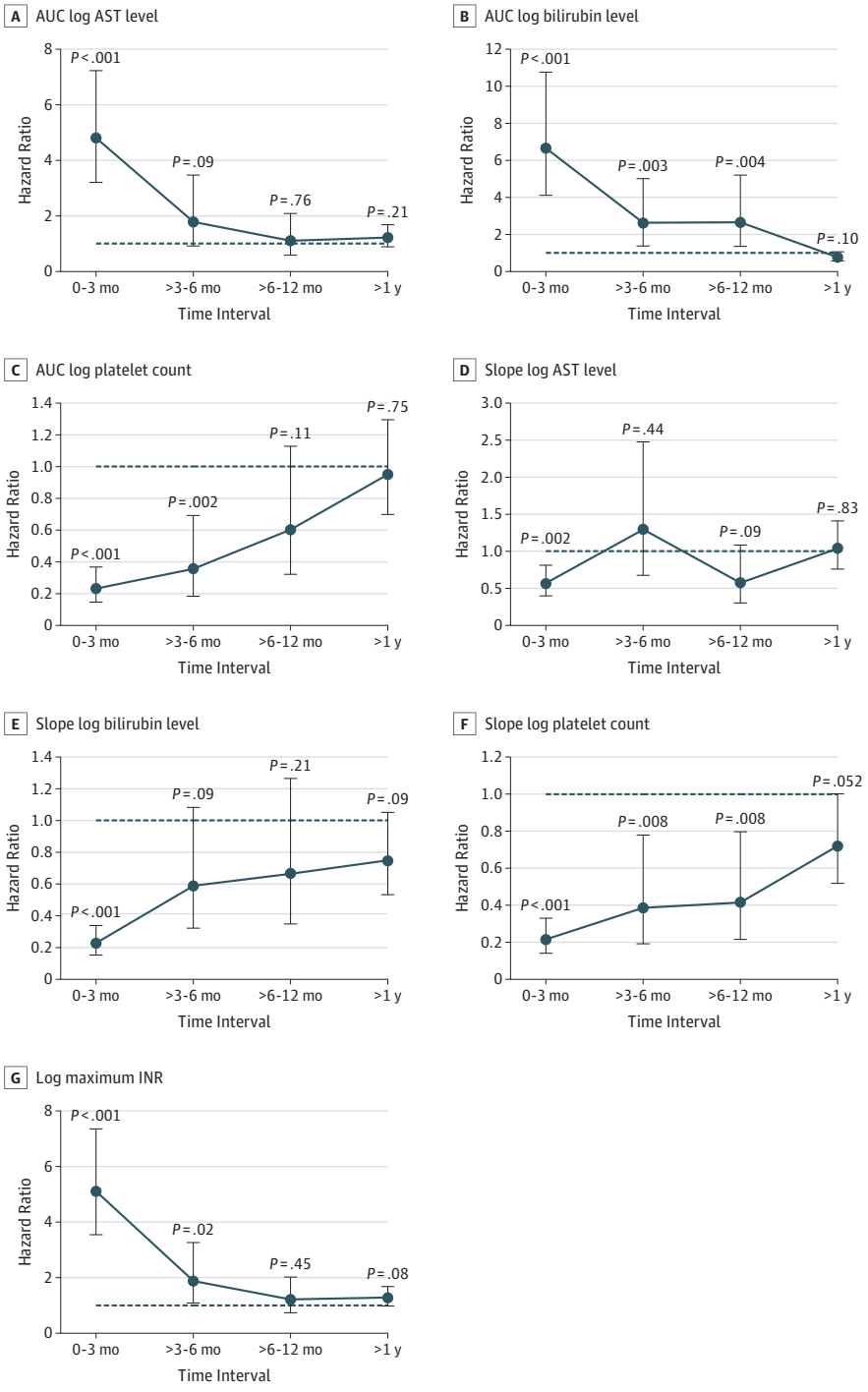
The most frequent underlying cause of liver disease was hepatitis C (38.5%), with diabetes, hypertension, and coronary artery disease in 25.9%, 29.6%, and 7.0% of patients, respectively. The median (IQR) recipient laboratory MELD score was 31 (25-38), with 46.1% of patients requiring pretransplant hospitalization, 31.8% requiring renal replacement therapy, 19.8% requiring mechanical ventilation, and 12.2% requiring vasopressors. Among the donors, the median (IQR) age was 41 (25-53) years, and 1294 (64.4%) were male. For this group, an LT was performed with a median (IQR) CIT of 404 (304-511) minutes, a median (IQR) WIT of 40 (35-47) minutes, and a median (IQR) transfusion of 11 (7-18) uPRBCs. At a median (IQR) follow-up of 36.8 (13.2-78.3) months, the overall patient and graft failure–free survival rates were 83% and 81% at year 1, 74% and 71% at year 3, and 69% and 65% at year 5, with an 11.1% (n = 222) incidence of 3-month graft failure (eFigure 1 in the Supplement). Of the 222 recipients, 180 (81.4%) died and 42 (18.9%) required a retransplant.

L-GrAFT Model Derivation

The ability of post-LT laboratory measures to prognosticate graft failure was first assessed using Cox regression analysis. The proportional hazards assumptions that each variable would have a constant association with graft failure over time was evaluated by plotting the HRs by interval of time for all 18 measures (eTable 1 in the Supplement). As shown in **Figure 1** for 7 of these variables included in the L-GrAFT model, the association of each laboratory variable with graft failure was most significant at 3 months but diminished significantly thereafter, consistently approaching nonsignificant HRs of 1 at 1 year. Specifically, the cumulative 10-day post-LT exposure and the rates of normalization of AST (AST log AUC [Figure 1A] and slope log AST [Figure 1D]) and bilirubin (bilirubin log AUC [Figure 1B] and slope log bilirubin [Figure 1E]) levels and platelet count (platelets log AUC [Figure 1C] and slope log platelets [Figure 1F]) and the maximum post-LT INR (log max INR [Figure 1G]) significantly affected graft failure at 3 months, but each approached nonsignificant HRs at 1 year.

Based on the results of the Cox regression, the final multivariate L-GrAFT model was constructed as a logistic regression evaluating the association of post-LT variables with the development of 3-month graft failure. Multivariate factors in 3-month graft failure are shown in **Table 1** and include the total 10-day AST

Figure 1. Association of Each Laboratory Variable With the Hazard Ratios (HRs) for Graft Failure Stratified by Time Interval



Reported HRs correspond with changes from the fourth (highest) quartile to the first (lowest) quartile for all measures except rate of normalization (slope) of AST (D) and bilirubin (E) levels, where HRs correspond with changes from the lowest quartile (more negative slope indicates faster rate of normalization) to the highest quartile (less negative slope indicates slower rate of normalization). For all measures, HRs were significant and most pronounced for graft failures occurring within 3 months and universally declined and approached nonsignificant levels for graft failures occurring after 12 months. AST indicates aspartate aminotransferase; AUC, area under the curve; INR, international normalized ratio.

exposure (AUC_{10} log AST: odds ratio [OR], 1.30 per SD increase; 95% CI, 1.05-1.60; P = .02) and early rate of normalization over 7 days ($Slope_7$ log AST: OR, 0.52 per SD decrease; 95% CI, 0.36-0.75; P = .02), maximum INR within the 10 posttransplant days (log maximum INR_{10}: OR, 1.22 per SD increase; 95% CI, 1.02-1.45; P = .001), total 10-day bilirubin (AUC_{10} log bilirubin: OR, 1.90 per SD increase; 95% CI, 1.54-2.34; P < .001) and rate of decrease ($Slope_{10}$ log bilirubin: OR, 0.62 per SD decrease; 95% CI, 0.51-0.75; P < .001), and total 10-day platelet count (AUC_{10} log platelets: OR, 0.80 per SD increase; 95% CI, 0.65-0.9; P = .040) and rate of increase ($Slope_{10}$ log platelets: OR, 0.73 per SD increase; 95% CI, 0.60-0.88; P = .001). A modified L-GrAFT model based on 7-day post-LT variables ($L\text{-}GrAFT_7$) is shown in eTable 3 in the Supplement.

Table 1. Multivariate Factors in the 3-Month Post-LT Graft Failure Risk Assessment Used in the L-GrAFT Model[a]

Variable	OR (95% CI)	*P* Value
AUC_{10} log AST level[b]	1.30 (1.05-1.60)	.02
$Slope_7$ log AST level decrease, rate of change/d[c]	0.52 (0.36-0.75)	.001
Log maximum INR_{10} level	1.22 (1.02-1.45)	.03
AUC_{10} log bilirubin level[b]	1.90 (1.54-2.34)	<.001
$Slope_{10}$ log bilirubin level decrease, rate of change/d	0.62 (0.51-0.75)	<.001
AUC_{10} log platelets count	0.80 (0.65-0.99)	.04
$Slope_{10}$ log platelets count increase, rate of change/d[b,c]	0.73 (0.60-0.88)	.001

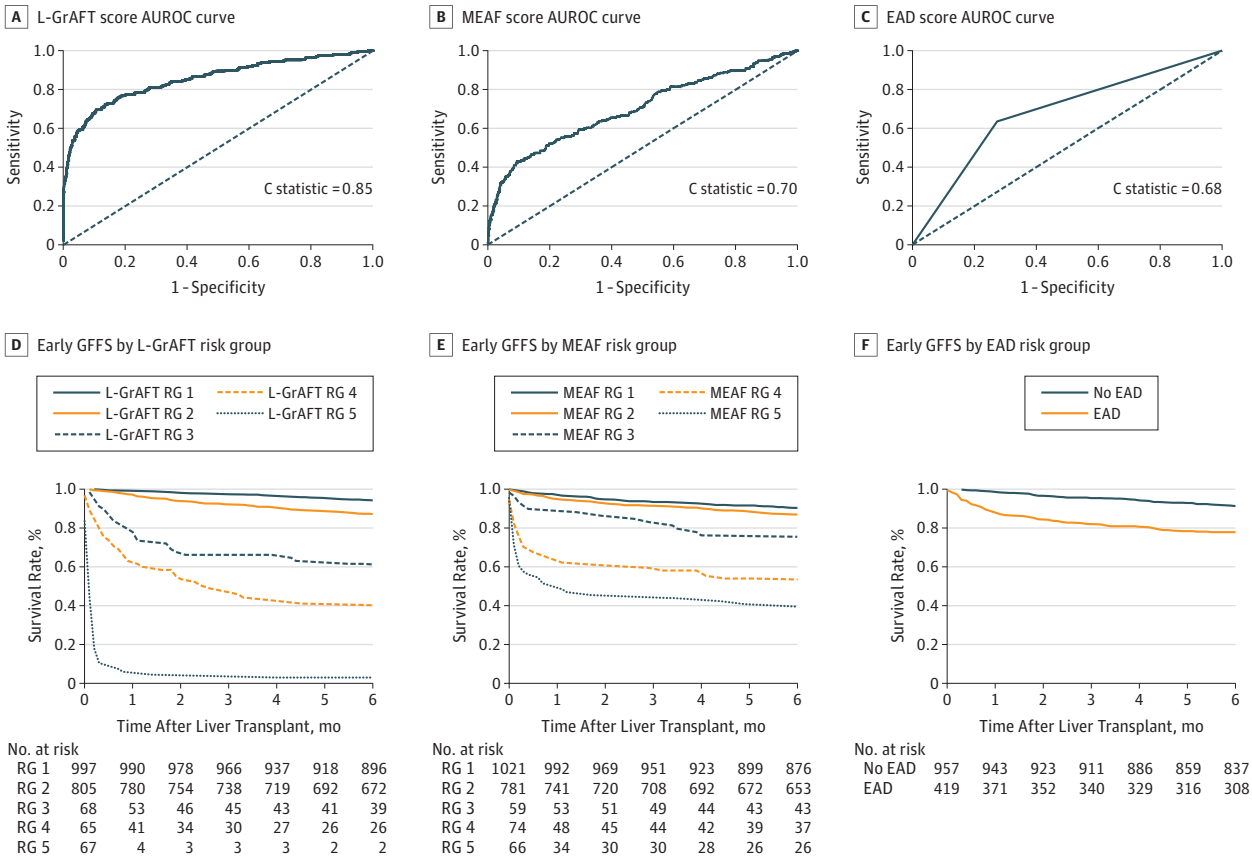
Abbreviations: AST, aspartate aminotransferase; AUC, area under the curve (over 10 post-LT days); INR, international normalized ratio; L-GrAFT, Liver Graft Assessment Following Transplantation; LT, liver transplant; OR, odds ratio.

[a] C statistic, 0.85 (200 bootstrapped C statistic, 0.84).

[b] For nonlinearly associated measures, reported odds ratios correspond to 1-SD change relative to the median.

[c] $Slope_7$ indicates early rate of change over 7 days. $Slope_{10}$ indicates rate of change over a 10-day period.

Figure 2. Comparison of Model Accuracies as Measured by the Area Under the Receiver Operating Characteristic (AUROC) Curve Among 3 Models of Early Allograft Dysfunction



No. at risk

RG 1	997	990	978	966	937	918	896
RG 2	805	780	754	738	719	692	672
RG 3	68	53	46	45	43	41	39
RG 4	65	41	34	30	27	26	26
RG 5	67	4	3	3	3	2	2

No. at risk

RG 1	1021	992	969	951	923	899	876
RG 2	781	741	720	708	692	672	653
RG 3	59	53	51	49	44	43	43
RG 4	74	48	45	44	42	39	37
RG 5	66	34	30	30	28	26	26

No. at risk

No EAD	957	943	923	911	886	859	837
EAD	419	371	352	340	329	316	308

The Liver Graft Assessment Following Transplantation (L-GrAFT) score (A; C statistic, 0.85) had the highest AUROC statistically significantly superior to both the Model for Early Allograft Function (MEAF) score (B; C statistic, 0.70; *P* < .001) and the early allograft dysfunction (EAD) score (C; C statistic, 0.68; *P* < .001). The L-GrAFT model allowed for greater discrimination of 3-month graft-failure risk (D) compared with the MEAF score (E) and the binary EAD definition (F). Five risk groups (RGs) for 3-month graft failure for both L-GrAFT and MEAF were defined on the basis of risk score distribution, including (1) very low risk (≤50th percentile), (2) low risk (>50th to ≤90th percentile), (3) moderate risk (>90th to ≤93.3 percentile), (4) moderate-to-high risk (>93.3 to ≤96.6 percentile), and (5) high risk (>96.6 percentile). GFFS indicates graft failure–free survival.

Comparison of L-GrAFT Model With Existing Models of Early Allograft Dysfunction

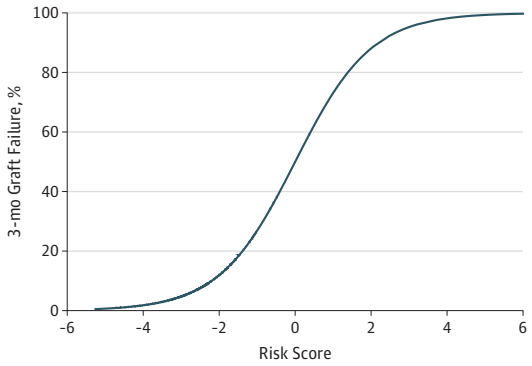
The L-GrAFT model had good calibration (goodness of fit *P* = .41; eFigure 2 in the Supplement) with an excellent C statistic of 0.85 (internally validated C statistic of 0.84), which is significantly superior to the calculated MEAF score (C statistic, 0.70; *P* < .001) and EAD (C statistic, 0.68; *P* < .001) in determining 3-month graft failure in the entire cohort (**Figure 2**A-C), and excellent discrimination of risk when stratified by risk score groups (Figure 2D-F). The L-GrAFT risk score

calculator was developed for individualized estimation of 3-month graft-failure risk; the higher the risk score, the greater the risk for 3-month graft failure (**Figure 3**). The formula for risk-score calculation is as follows: risk score = 11.27 − 0.429 × (AUC log AST) + 0.005 × (AUC log AST^2) + 4.607 × (early slope log AST) + 4.413 × (early slope log AST^2) + 0.890 × (log max INR − 0.049 × (AUC log TBIL) + 0.004 × (AUC log TBIL^2) + 5.336 × (slope log TBIL) − 0.046 × (AUC log PLT) − 5.249 × (slope log PLT) + 13.086 × (slope log PLT^2), where TBIL stands for total bilirubin and PLT stands for platelets.

Post-LT Laboratory Profiles and Perioperative Characteristics Among L-GrAFT Risk Groups

The mean post-LT laboratory profile plots for the 5 L-GrAFT risk groups are shown in eFigure 3 in the Supplement. Compared with patients in lower-risk score groups (L-Graft 1 and 2), patients with higher L-Graft scores (L-Graft 3, 4, and 5) had increased overall AST (eFigure 3A in the Supplement) and ALT (eFigure 3B in the Supplement) levels, higher bilirubin levels with slower normalization (eFigure 3C in the Supplement), higher maximum INRs (eFigure 3D in the Supplement), and lower overall platelet counts with a slower rate of increase (eFigure 3E in the Supplement). Evaluation of recipient, donor, and operative characteristics among L-GrAFT risk score groups are shown in **Table 2**. Compared with patients with lower L-GrAFT scores (L-Graft 1 and 2), LT recipients with higher L-GrAFT scores (L-Graft 3, 4, and 5) had significantly greater median MELD scores (median [IQR] score, 34 [26-40] vs 31 [25-38]; $P = .005$); higher pretransplant serum bilirubin levels (median [IQR] level, 14.8 vs 6.4 mg/dL; $P < .001$); a greater need for pretransplant hospitalization (56.8% vs 44.8% patients; $P = .003$), mechanical ventilation (35.8% vs 18.1% patients; $P < .001$), renal replacement therapy (42.9% vs 30.5% patients; $P < .001$), and vasopressors (22.9% vs 11.0%; $P < .001$); longer CITs (median [IQR] time, 436 [311-539] vs 401 [302-506] minutes; $P = .04$); more operative blood transfusions (median [IQR], 17 [10-26] vs 10 [6-17] uPRBCs; $P < .001$); and a greater frequency of older donors (median [IQR] age, 47 [28-56] vs 41 [25-52] years; $P < .001$).

Figure 3. Association of Estimated Risk for 3-Month Graft Failure With Liver Graft Assessment Following Transplantation (L-GrAFT) Risk Score



The mathematical formula for calculating L-GrAFT risk score can be found in the Results section. The L-GrAFT scores corresponding to the 5 risk groups are as follows: (1) very low risk (<−3.23), (2) low risk (≥−3.23 to<−1.18), (3) moderate risk (≥−1.18 to<−0.57), (4) moderate-to-high risk (≥−0.57 to <1.3), and (5) high risk (>1.3).

Discussion

The durability of an LT in treating end-stage liver disease has created a critical shortage of donor organs, with far more patients in need of organs than are currently available. Use of extended criteria donors as a strategy to expand the donor pool may be effective in mitigating waitlist mortality,[19] but use of marginal grafts for an ever-increasing population of higher-acuity recipients risks the development of EAD,[2,20] which negatively affects post-LT outcomes.[21,22] With an increasing number of translational studies aiming to ameliorate the ischemia-reperfusion injury responsible for EAD,[23-27] it is critical to have a reliable measure of EAD that allows for comparison of meaningful graft-related outcomes across all transplant centers. Based on our analysis of more than 2000 patients who underwent primary LT, we developed the L-GrAFT risk score that allows for a highly accurate, individualized calculation of 3-month graft failure.

Our L-GrAFT model incorporates 10 post-LT days of serum AST levels, a marker of graft injury, and the INR and bilirubin levels as measures of graft synthetic and metabolic functions. Most previous definitions of EAD have incorporated these variables (eg, maximum value within 3-7 days for the serum transaminase level and the INR, as well as the maximum value at day 3 or 7 for the bilirubin level)[6,12-15,17] as a static measure within a specified post-LT time frame. However, to our knowledge, the L-GrAFT model is the first comprehensive measure of EAD to also take into account the rate of change (slope) and trend in these variables. We show that faster normalization of both serum AST and bilirubin levels affect superior graft survival, independent of the absolute arbitrary values that have been used in previous models. For example, the existing binary definition would categorize all patients with a bilirubin level of 10 mg/dL or higher on POD 7 as having EAD; however, there is clearly a difference in graft function between a patient with a pretransplant bilirubin level of 40 mg/dL whose bilirubin is normalizing and a patient with liver cancer with a physiological MELD 7 who has a normal pretransplant bilirubin level and has developed significant cholestasis 1 week following LT. In addition, the total exposure (AUC) of AST or bilirubin over 10 days was more of a determinant of graft failure than any specific maximum or ultimate value, presumably because it reflects an average measure of graft assessment over time. Finally, to our knowledge, our L-GrAFT risk score is the first to incorporate postoperative platelet counts into a measure of EAD, with lower post-LT platelet counts and slower rates of increase highly associated with greater post-LT graft failure. Our findings are consistent with those in recent reports demonstrating an association with low postoperative platelet counts and delayed liver function following partial hepatectomy,[28] as well as increased post-LT complications and decreased 90-day graft and patient survival following LT.[29,30]

Table 2. Comparison of Characteristics Among Low- and High-Risk Groups in the L-GrAFT Model

Demographic	L-GrAFT Risk Group		*P* Value
	Low (n = 1808)	High (n = 200)	
Recipient			
Age, median (IQR), y	56 (49-62)	55 (48-60)	.09
Male sex, %	64.7	62.5	.55
Diagnosis, %			
Hepatitis C	38.8	35.5	.89
Cryptogenic cirrhosis	13.9	13.5	
Alcohol abuse	13.3	17.0	
Cholestatic/autoimmune disease	7.7	7.0	
Nonalcoholic steatohepatitis	6.7	8.0	
Hepatitis B	6.9	6.0	
Acute liver failure	6.3	7.0	
Other	6.4	6.0	
Hypertension, %	30.2	24.6	.11
Diabetes, %	25.9	25.8	.97
Coronary artery disease, %	7.1	6.1	.60
Pretransplant acuity			
Laboratory MELD score, median (IQR)	31 (25-38)	34 (26-40)	.005
Serum bilirubin level, median (IQR), mg/dL	6.4 (2.0-23.2)	14.8 (2.9-36.6)	<.001
INR, median (IQR)	1.6 (1.3-2.1)	1.7 (1.3-2.1)	.46
Hospitalization, %	44.8	56.8	.003
Renal replacement therapy, %	30.5	42.9	<.001
Mechanical ventilation, %	18.1	35.8	<.001
Vasopressors, %	11.0	22.9	<.001
Donor and operative			
Donor age, median (IQR), y	41 (25-52)	47 (28-56)	<.001
Donor male sex, %	64.7	62.5	.55
Nonstandard donor (DCD or split graft), %	9.8	9.5	.88
Cold ischemia time, median (IQR), min	401 (302-506)	436 (311-539)	.04
Warm ischemia time, median (IQR), min	41 (35-48)	40 (35-47)	.35
Transfusion of uPRBCs, median (IQR)	10 (6-17)	17 (10-26)	<.001

Abbreviations: DCD, donation after cardiac death; INR, international normalized ratio; IQR, interquartile range; L-GrAFT, Liver Graft Assessment Following Transplantation model; MELD, Model for End-Stage Liver Disease; uPRBCs, units of packed red blood cells.

Note: Data for some variables were missing for some patients. Thus, the numbers and percentages reported here are based on only patients with available information.

SI conversion factor: To convert serum bilirubin level to micromole per liter, multiply by 17.104.

The L-GrAFT risk score, as a dynamic assessment of the graft over 10 days, demonstrated far superior discrimination of post-LT graft failure compared with both the EAD definition and the MEAF score. Based on the binary definition, 30.4% patients in our cohort were characterized as having EAD, with a modest discrimination of 3-month graft failure (AUC, 0.68) when comparing patients with EAD (17.5%) with patients without EAD (4.6%). Conversely, the individualized calculation of 3-month graft failure along a continuum of L-GrAFT risk scores (AUC, 0.85) allowed for significantly superior discrimination of risk compared with both the EAD definition (AUC, 0.68) and the MEAF score (AUC, 0.70), which itself was not superior to the binary EAD categorization. In fact, despite the MEAF score allowing for continuous grading of EAD severity and its external validation in an independent cohort of LT recipients, the authors of that study did not report the MEAF model's accuracy, limiting any objective assessment of added utility when it is compared with the EAD definition.

Another notable finding was that the magnitude of the association between post-LT laboratory variables and graft failure did not remain constant over time, with the largest association seen early (at 3 months) and subsequently diminishing with HRs universally approaching 1 after 12 months. This empirical rejection of the proportional hazards assumption was the primary reason we developed the L-GrAFT risk score as a logistic regression of 3-month graft failure. Not surprisingly, an examination of factors associated with higher L-GrAFT risk scores identified recipient acuity measures (ie, higher MELD scores, higher number of patients who underwent pretransplant renal replacement therapy, and greater need for intubation and vasopressors), as well as donor characteristics (increased age) and operative characteristics (higher CITs and uPRBC transfusions), that have strongly been associated with inferior graft outcomes. The well-established Survival Outcomes Following Liver Transplantation[31] and Balanced Assessment of Risk scores,[32] which model 3-month patient survival, have identified similar recipient, donor, and operative characteristics, but both scores demonstrate modest model accuracy (C statistic, 0.70) compared with the L-GrAFT risk score (C statistic, 0.85). This finding perhaps is not unexpected given that potentially unknowable pretransplant variables not included in these models ultimately affect the early graft func-

tion or transplant phenotype, which is more accurately characterized by post-LT laboratory studies.

Limitations

This study has several limitations. The requirement for 10 days of post-LT variables may be more cumbersome compared with the requirement for other measures of EAD that rely on only 3 or 7 days of laboratory values; however, the significantly superior accuracy of the L-GrAFT in determining 3-month graft failure may justify the inclusion of more data points. In fact, the evaluation of the L-GrAFT score using only 7 days of post-LT variables nonetheless yielded a C statistic of 0.83, which is far superior to that achieved by existing models. The actual L-GrAFT score is mathematically complicated, but an online L-GrAFT calculator allows transplant providers to simply enter the post-LT laboratory values to obtain an estimated risk of 3-month graft failure without the need for complex calculations. Finally, although our L-GrAFT model showed excellent accuracy in determining 3-month graft failure in a cohort of high-acuity recipients in a transplant center in the western United States, its results may not necessarily be generalizable to transplant recipients in different practice environments. To this end, external validation of the L-GrAFT risk score is necessary to establish its role as an important tool for grading early allograft function across all transplant centers and practice environments.

Conclusions

The L-GrAFT risk score model allows for highly accurate, individualized risk estimation of 3-month graft failure following LT and is superior to the existing binary EAD classification and MEAF score. External multicenter validation may lead to the adoption of the L-GrAFT as a tool for determining the need for a retransplant and for establishing a standardized grading system of early allograft function across transplant centers. Finally, with innumerable translational studies aiming to improve the quality of marginal grafts to expand the donor pool, or to modify recipient factors to mitigate ischemia-reperfusion injury, the L-GrAFT score, as a highly accurate measure of early graft failure, may serve as an excellent clinical end point by which to measure and compare the efficacy of interventions.

ARTICLE INFORMATION

Accepted for Publication: August 25, 2017.

Correction: This article was corrected on May 16, 2018, to fix an error in the presentation of data in the Results section.

Published Online: December 20, 2017. doi:10.1001/jamasurg.2017.5040

Author Contributions: Dr Agopian had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Agopian, Harlander-Locke, Xia, Zarrinpar, Farmer, Busuttil.
Acquisition, analysis, or interpretation of data: Agopian, Harlander-Locke, Markovic, Dumronggittigule, Xia, Kaldas, Zarrinpar, Yersiz, Hiatt, Busuttil.
Drafting of the manuscript: Agopian, Harlander-Locke, Hiatt, Busuttil.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Agopian, Markovic, Dumronggittigule.
Administrative, technical, or material support: Agopian, Harlander-Locke, Kaldas, Yersiz, Busuttil.
Study supervision: Agopian, Farmer, Busuttil.

Conflict of Interest Disclosures: None reported.

Meeting Presentation: This study was presented in part at the 88th Annual Meeting of the Pacific Coast Surgical Association; February 19, 2017; Indian Wells, California.

REFERENCES

1. Starzl TE, Marchioro TL, Vonkaulla KN, Hermann G, Brittain RS, Waddell WR. Homotransplantation of the liver in humans. *Surg Gynecol Obstet*. 1963;117:659-676.

2. Agopian VG, Petrowsky H, Kaldas FM, et al. The evolution of liver transplantation during 3 decades: analysis of 5347 consecutive liver transplants at a single center. *Ann Surg*. 2013;258(3):409-421.

3. Ali JM, Davies SE, Brais RJ, et al. Analysis of ischemia/reperfusion injury in time-zero biopsies predicts liver allograft outcomes. *Liver Transpl*. 2015;21(4):487-499.

4. Zhai Y, Petrowsky H, Hong JC, Busuttil RW, Kupiec-Weglinski JW. Ischaemia-reperfusion injury in liver transplantation—from bench to bedside. *Nat Rev Gastroenterol Hepatol*. 2013;10(2):79-89.

5. Deschenes M. Early allograft dysfunction: causes, recognition, and management. *Liver Transpl*. 2013;19(suppl 2):S6-S8.

6. Olthoff KM, Kulik L, Samstein B, et al. Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl*. 2010;16(8):943-949.

7. Lee DD, Singh A, Burns JM, Perry DK, Nguyen JH, Taner CB. Early allograft dysfunction in liver transplantation with donation after cardiac death donors results in inferior survival. *Liver Transpl*. 2014;20(12):1447-1453.

8. Lee DD, Croome KP, Shalev JA, et al. Early allograft dysfunction after liver transplantation: an intermediate outcome measure for targeted improvements. *Ann Hepatol*. 2016;15(1):53-60.

9. Pomposelli JJ, Goodrich NP, Emond JC, et al. Patterns of early allograft dysfunction in adult live donor liver transplantation: the A2ALL experience. *Transplantation*. 2016;100(7):1490-1499.

10. Croome KP, Hernandez-Alejandro R, Chandok N. Early allograft dysfunction is associated with excess resource utilization after liver transplantation. *Transplant Proc*. 2013;45(1):259-264.

11. Wadei HM, Lee DD, Croome KP, et al. Early allograft dysfunction after liver transplantation is associated with short- and long-term kidney function impairment. *Am J Transplant*. 2016;16(3):850-859.

12. Ardite E, Ramos C, Rimola A, Grande L, Fernández-Checa JC. Hepatocellular oxidative stress and initial graft injury in human liver transplantation. *J Hepatol*. 1999;31(5):921-927.

13. González FX, Rimola A, Grande L, et al. Predictive factors of early postoperative graft function in human liver transplantation. *Hepatology*. 1994;20(3):565-573.

14. Ploeg RJ, D'Alessandro AM, Knechtle SJ, et al. Risk factors for primary dysfunction after liver transplantation—a multivariate analysis. *Transplantation*. 1993;55(4):807-813.

15. Deschênes M, Belle SH, Krom RA, Zetterman RK, Lake JR; National Institute of Diabetes and Digestive and Kidney Diseases Liver Transplantation Database. Early allograft dysfunction after liver transplantation: a definition and predictors of outcome. *Transplantation*. 1998;66(3):302-310.

16. Wiesner R, Edwards E, Freeman R, et al; United Network for Organ Sharing Liver Disease Severity Score Committee. Model for End-Stage Liver Disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91-96.

17. Pareja E, Cortes M, Hervás D, et al. A score model for the continuous grading of early allograft dysfunction severity. *Liver Transpl*. 2015;21(1):38-46.

18. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.

19. Barshes NR, Horwitz IB, Franzini L, Vierling JM, Goss JA. Waitlist mortality decreases with increased use of extended criteria donor liver grafts at adult liver transplant centers. *Am J Transplant*. 2007;7(5):1265-1270.

20. Lunsford KE, Bodzin AS, Markovic D, et al. Avoiding futility in simultaneous liver-kidney transplantation: analysis of 331 consecutive patients listed for dual organ replacement. *Ann Surg*. 2017;265(5):1016-1024.

21. Briceño J, Ciria R, de la Mata M, Rufián S, López-Cillero P. Prediction of graft dysfunction based on extended criteria donors in the model for end-stage liver disease score era. *Transplantation*. 2010;90(5):530-539.

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.

23. Bärthel E, Rauchfuss F, Hoyer H, Breternitz M, Jandt K, Settmacher U. The PRAISE study: a prospective, multi-center, randomized, double blinded, placebo-controlled study for the evaluation of iloprost in the early postoperative period after liver transplantation (ISRCTN12622749). *BMC Surg*. 2013;13:1.

24. Bharathan VK, Chandran B, Gopalakrishnan U, et al. Perioperative prostaglandin e1 infusion in living donor liver transplantation: a double-blind, placebo-controlled randomized trial. *Liver Transpl*. 2016;22(8):1067-1074.

25. de Rougemont O, Breitenstein S, Leskosek B, et al. One Hour Hypothermic Oxygenated Perfusion (HOPE) protects nonviable liver allografts donated after cardiac death. *Ann Surg*. 2009;250(5):674-683.

26. Guarrera JV, Henry SD, Samstein B, et al. Hypothermic machine preservation facilitates successful transplantation of "orphan" extended criteria donor livers. *Am J Transplant*. 2015;15(1):161-169.

27. Henry SD, Nachber E, Tulipan J, et al. Hypothermic machine preservation reduces molecular markers of ischemia/reperfusion injury in human liver transplantation. *Am J Transplant*. 2012;12(9):2477-2486.

28. Alkozai EM, Nijsten MW, de Jong KP, et al. Immediate postoperative low platelet count is associated with delayed liver function recovery after partial liver resection. *Ann Surg*. 2010;251(2):300-306.

29. Lesurtel M, Raptis DA, Melloul E, et al. Low platelet counts after liver transplantation predict early posttransplant survival: the 60-5 criterion. *Liver Transpl*. 2014;20(2):147-155.

30. Takahashi K, Nagai S, Putchakayala KG, et al. Prognostic impact of postoperative low platelet count after liver transplantation. *Clin Transplant*. 2017;31(3).

31. 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(12):2537-2546.

32. 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(5):745-753.

Invited Commentary

Determinants of Failure in Hepatic Transplant

Jorge D. Reyes, MD

One of the most enduring truisms in transplantation of all organs was written by Starzl almost 60 years ago: "The provision of a viable and minimally damaged homograft is undoubtedly the most important single factor in the determinant of success."[1] Presently, there are far more patients in need of transplant than there are organs available, and the use of expanded criteria donors has become an important source of organs for an ever-increasing acuity of patients. Consequently, the specter of graft failure and the ability to assess and predict outcomes to better intervene with management strategies has become the frontier in transplantation.

Related article page 436

In this issue of *JAMA Surgery*, Agopian et al[2] present a study using the Liver Graft Assessment Following Transplantation (L-GrAFT) model, which proposes to individualize prediction of graft failure after liver transplant (ie, need for retransplantation or death) and is based on posttransplant laboratory variables most predictive of graft failure, which included serum aspartate aminotransferase level, alanine aminotransferase level, total bilirubin level, international normalized ratio, and platelet counts, as well as donor and recipient variables. The hazard ratios for time to graft failure for each predictor (allowing for interactions) were plotted by time interval of event, and model accuracy was evaluated using receiver operating characteristic analysis and area under the curve (AUC) analysis, comparing them with the accuracy of the early allograft dysfunction (EAD) and Model for Early Allograft Function scores. Using Cox regression analysis, the L-GrAFT model was derived by plotting the hazard rate ratios by interval for 18 variables found to be most significant at 3 months; the final multivariate model evaluated the effectiveness of these predictors on the development of graft failure within 3 months. A risk score was calculated as a weighted sum of covariates, and then 5 risk groups for graft failure within 3 months were defined based on the risk score distribution from very low risk to high risk. The C statistic for the L-GrAFT model was internally validated on 200 bootstrapped samples taken with replacement from the original data. The authors report on 2008 liver transplant recipients over a study period from February 2002 to June 2015; at a median follow-up of 36.8 months, there was an 11% incidence (210 recipients) of 3-month graft failure, of whom 170 (81.0%) died and 40 (19.0%) required retransplant.[2] The predictive performance of the L-GrAFT model was compared with reports of EAD and several validation studies of EAD, including the Model for Early Allograft Function score.[3-7]

The authors affirm that the strengths of this model are the analysis of the rate of change (slope) and trend as well as analysis of the total exposure (AUC) rather than maximums, thus reflecting a better graft assessment. The L-GrAFT model (AUC, 0.85) demonstrated superior discrimination than the EAD score (AUC, 0.68) and the Model for Early Allograft Function score (AUC, 0.70), which are binary assessments as opposed to individualized prediction along a continuum. The effect of recipient and donor characteristics comparing the Survival Outcomes Following Liver Transplantation score[8] and the Balanced Assessment of Risk score[9] also seemed to have modest accuracy compared with the L-GrAFT model proposed by Agopian et al.[2]

Although this is an interesting study, one major concern is using an internal validation set where 200 bootstrapped samples were used on data already seen by the model (in the validation set, some samples can be used many times); consequently, the C statistic of the model can be artificially high because it has seen the data. A better C statistic would be to