Received: 18 June 2018 | Revised: 4 October 2018 | Accepted: 22 October 2018

DOI: 10.1111/ajt.15172

ORIGINAL ARTICLE

AJT

Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation

Dimitris Bertsimas[1] | Jerry Kung[1] | Nikolaos Trichakis[1] | Yuchen Wang[1] | Ryutaro Hirose[2] | Parsia A. Vagefi[3]

[1]Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts

[2]Department of Surgery, University of California, San Francisco, California

[3]Department of Surgery, University of Texas Southwestern Medical Center, Dallas, Texas

Correspondence
Dimitris Bertsimas
Email: dbertsim@mit.edu

Since 2002, the Model for End-Stage Liver Disease (MELD) has been used to rank liver transplant candidates. However, despite numerous revisions, MELD allocation still does not allow for equitable access to all waitlisted candidates. An optimized prediction of mortality (OPOM) was developed (http://www.opom.online) utilizing machine-learning optimal classification tree models trained to predict a candidate's 3-month waitlist mortality or removal utilizing the Standard Transplant Analysis and Research (STAR) dataset. The Liver Simulated Allocation Model (LSAM) was then used to compare OPOM to MELD-based allocation. Out-of-sample area under the curve (AUC) was also calculated for candidate groups of increasing disease severity. OPOM allocation, when compared to MELD, reduced mortality on average by 417.96 (406.8-428.4) deaths every year in LSAM analysis. Improved survival was noted across all candidate demographics, diagnoses, and geographic regions. OPOM delivered a substantially higher AUC across all disease severity groups. OPOM more accurately and objectively prioritizes candidates for liver transplantation based on disease severity, allowing for more equitable allocation of livers with a resultant significant number of additional lives saved every year. These data demonstrate the potential of machine learning technology to help guide clinical practice, and potentially guide national policy.

KEYWORDS

ethics and public policy, liver transplantation/hepatology, liver transplantation: auxiliary, simulation, statistics

1 | INTRODUCTION

The successful clinical application of liver transplantation has generated a discrepancy between supply and demand, thereby generating a persistent insufficient organ supply that results in thousands of candidate deaths every year while candidates await liver transplantation. Given the scarcity of this resource, one of the most crucial challenges in liver transplantation involves accurately prioritizing a waitlisted candidate's likelihood of death within the near future, so that the limited supply of donated livers can be allocated to maximize the benefit from transplantation.

Since 2002, liver allocation has depended on the Model for End-Stage Liver Disease (MELD) score to rank disease severity and, consequently, priority for receiving a liver transplant.[1]

Abbreviations: AUC, area under the curve; CART, classification and regression tree; DABD, donation after brain death; DSA, donor service area; HCC, hepatocellular carcinoma; HRSA, Health Resources and Services Administration; INR, international normalized ratio; LSAM, Liver Simulated Allocation Model; MELD, model for end-stage liver disease; MMRF, Minneapolis Medical Research Foundation; OCT, optimal classification tree; OPOM, optimized prediction of mortality; OPTN, Organ Procurement and Transplantation Network; SRTR, Scientific Registry of Transplant Recipients; STAR, Standard Transplant Analysis and Research; UNOS, United Network for Organ Sharing.

Certain patient populations, however, are at risk of death or of becoming too sick and unsuitable for transplantation based on disease progression that is not captured in their lab-based MELD score calculation. To allow them to contend for liver offers, these candidate populations have been granted "artificial" points (MELD exception points). Although overall the MELD score has allowed for a more objective ranking of candidates awaiting liver transplantation, compared to the pre-MELD era, the process of MELD exception point granting has emerged as a significant weakness in the allocation process, leading to inequitable and undesirable outcomes.[2] In particular, the arbitrary MELD score exception points policy has overly prioritized the subpopulation of liver transplant candidates with hepatocellular carcinoma (HCC).[3] Indeed, since the adoption of the MELD score, there have been multiple policy revisions to reduce the amount of exception points for HCC candidates to more accurately reflect this population's risk of waitlist removal from death or tumor progression. Notwithstanding these revisions, there remains a higher risk of waitlist death/removal for candidates without exception points, when compared to those candidates with exception points.

We sought to utilize a state-of-the-art machine-learning method—termed optimal classification trees (OCTs)—to generate a more accurate prediction of a liver candidate's 3-month waitlist mortality or removal, that would in-return allow for a more appropriate prioritization of candidates awaiting liver transplantation. The following prediction problem was posed: *What is the probability that a patient will either die or become unsuitable for liver transplantation within 3 months, given his or her individual characteristics?*

2 | MATERIALS AND METHODS

2.1 | Data

Waitlist, deceased-donor, transplant, and follow-up information was obtained for the period January 1, 2002, to September 5, 2016, from the Organ Procurement and Transplantation Network (OPTN) Standard Transplant Analysis and Research (STAR) dataset.

2.2 | Prediction methods

The prediction problem was addressed using data analytics models that were trained on historical data. Specifically, a model was calibrated based on optimal classification trees (or OCTs), which represented a state-of-the-art machine-learning prediction method that afforded interpretability and high prediction accuracy.[4] The end result was a classification tree that predicted the probability of a patient dying or becoming unsuitable for transplant within 3 months (the dependent variable), given observations of certain patient characteristics (the independent variables).

Classification trees are hierarchically organized structures of nodes that make predictions by sequentially "splitting the data" based on values of independent variables until a "leaf node" is reached. Given a certain tree, its predictive power is assessed by evaluating the accuracy of its predictions on historical observations. In theory, an infinite number of trees could be constructed, by varying the number of nodes, the independent variables used as splitting variables at the nodes, the associated splitting thresholds, and the predictions at the leaf nodes. OCTs, which were used to train the model, leverage mixed-integer optimization to methodically sweep through all such candidate trees. In this process, OCTs assess the predictive power of each tree, and in the end select the most favorable one, as detailed in model calibration below. Once trained, the model predicted as output the dependent variables, given (potentially previously unseen) observations of the independent variables. Henceforth the model is referred to as optimized prediction of mortality (OPOM; http://www.opom.online/).

To exemplify, Figure 1 depicts a sample classification tree, in which the data are first split at the Root Node based on the patient's MELD score. Proceeding in this fashion, a prediction for the dependent variable is made once one of the leaf nodes—Nodes 3-6 in this example—is reached. The dependent variable (dying or becoming unsuitable for transplant within 3 months) for a patient with MELD of 28 and bilirubin of 6.2, for example, is predicted to be 49% by this tree. By splitting the data merely twice—based on MELD and bilirubin—to make a prediction, this example tree had limited predictive power' the tree found to achieve the highest predictive power performed up to 10 splits to make a prediction, based on additional independent variables.

2.3 | Observations; dependent and independent variables

An observation corresponded to a patient at the time of a check-in visit, so that observed characteristics were all up-to-date. All such available observations for patients older than 12 years of age, dated after the implementation of MELD, were retrieved and totaled 1 618 966 observations. For each observation, the dependent variable was set to 1 if the patient died or was removed from the waitlist as unsuitable for transplant within the 3-month follow-up period from the observation date, and to 0 otherwise. A total of 28 independent variables were recorded for each observation, detailed in the supplementary materials (Table S1). Of note, all variables were readily retrieved from UNOSNet; of the 28 variables examined, 20 are variables associated with the traditional MELD, but in this instance applied with use of trajectories of these lab values (eg, change in international normalized ratio [INR] since previous check-in).

There were 374 666 observations that were missing their dependent variables due to the candidate receiving a liver transplant during the follow-up period. Two methods were used in the management of the transplanted cohort: (1) the missing dependent variables were imputed using a machine-learning approach,[5] which has demonstrated the ability to outperform other related extant methods; or (2) the transplanted cohort observations were excluded from the dataset. Both methods for dealing with observations missing

their dependent variables yielded statistically similar results. For brevity, only the results obtained by excluding these observations were reported in the Results.

2.4 | Model calibration

OPOM comprised 2 models: one for non-HCC candidates (independent variables 1-25), and one for HCC candidates (independent variables 1-28).

The observations of each patient were all randomly assigned to either the training, the validation, or the testing set, with probabilities 50%, 20%, and 30%, respectively. OPOM models were fit on the training set and then the out-of-sample accuracy value for the validation set was computed. Models with different tree depths (1 to 10) and different numbers of minimum observations in the leaves (1, 5, or 10) for OPOM were computed, and models that yielded the highest accuracy for the validation set were selected. The top 3 layers of the selected models can be found in the supplementary materials (Figure S1). Assessment of the independent variables that contributed the most predictive power is also demonstrated in the supplementary materials (Figure S2).

2.5 | Allocation outcomes

The latest version of Liver Simulated Allocation Model (or LSAM) was used (https://www.srtr.org/media/1203/lsam.pdf). LSAM is a program developed by the Scientific Registry of Transplant Recipients (SRTR) that uses historical real-world data from 2007 to 2011 to simulate the allocation of livers to candidates during that period. LSAM simulates allocations based on Match MELD, that is, the MELD score with consideration of exception points as per the 2014 national allocation policy. To measure the impact of OPOM, the simulation was run after substituting all patients' Match MELD scores with their corresponding OPOM scores, which, for consistency, were re-scaled to range between 6 and 40, instead of 0%-100%, and also to match the original MELD-score distribution. Through this substitution, all LSAM features, including its organ acceptance model, were retained. Of note, Status 1A candidates were listed using the same criteria

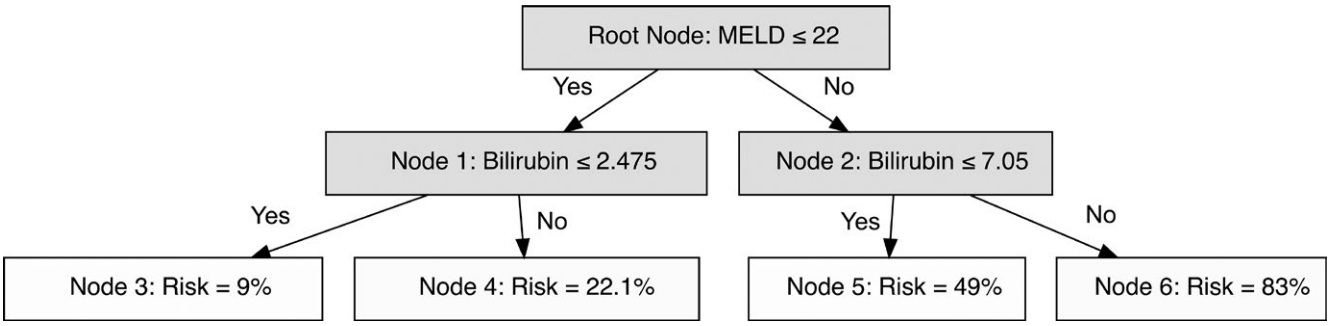

FIGURE 1 Example of a classification tree that predicts risk of dying or becoming unsuitable for transplant within 3 months. Although this sample tree splits the data twice—based on MELD and bilirubin—to make a prediction, OPOM performed up to 10 splits to make a prediction, based on additional independent variables

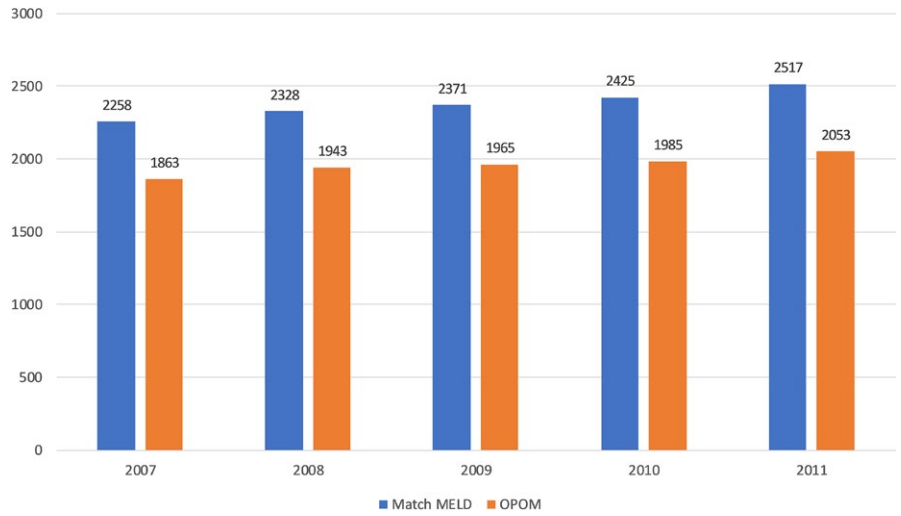

FIGURE 2 Simulated deaths by year for sample LSAM run: match MELD vs OPOM [Color figure can be viewed at wileyonlinelibrary.com]

as is done currently, were not assigned an OPOM score, and were ranked above non–Status-1A patients.

2.6 | Out-of-sample AUC

Performance was also evaluated by measuring out-of-sample area under the curve (or AUC) on the testing set. A model's AUC corresponded to the probability that a randomly drawn observation whose dependent value was 1 (ie, patient died or was removed from the list) had a higher score under that model than a randomly drawn observation whose dependent values were 0.[6] Therefore, OPOM's and MELD's AUC values measured their ability to identify patients who would die or become unsuitable for transplant within 3 months from patients who would not.

AUC was measured considering different patient populations based on exception status, and for both Match MELD and for MELD-Na, that is, the MELD score based on lab values, with no consideration of exception points, but with inclusion of the serum sodium level.

AUC was also measured for subpopulations of patients with increasing disease severity. For a fair comparison, when calculating OPOM's AUC, MELD was used to determine disease severity when stratifying patients and vice versa.

2.7 | Disclaimer

This study used data from the SRTR. The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services, provides oversight to the activities of the OPTN and SRTR contractors.

3 | RESULTS

3.1 | Simulation results

Figure 2 depicts the simulated average number of patient deaths each year by Match MELD and OPOM. Allocation of livers based

TABLE 1 LSAM simulated average annual deaths and transplants by candidate demographics under Match MELD and OPOM allocation

	Match MELD transplants	OPOM transplants	Match MELD deaths	OPOM deaths
Sex				
Male	3995.8 (3998-4001.4)	3798.36 (3790.2-3808.8)	1486.8 (1473.2-1494.8)	1213.48 (1200.4-1220)
Female	2181.76 (2170.2-2192.6)	2340.56 (2332.4-2358.6)	892.88 (886-898.2)	748.24 (741.8-753.2)
Race				
White	4247.64 (4233.4-4260)	4301.72 (4294.4-4317.2)	1674.64 (1665.2-1688.4)	1387.36 (1376.8-1397.2)
Black	699.28 (686.4-709.6)	640.28 (634.2-648.2)	219.32 (214.2-224.2)	175.12 (171.2-178.6)
Hispanic	862.44 (856.6-868.8)	893.2 (886.6-901.4)	373.04 (371.8-375.2)	304.48 (302.4-306.4)
Asian	294.32 (290-298.4)	231.32 (225.8-238.6)	84.72 (80.8-88.4)	72.52 (69.4-73.6)
Other	73.88 (72.2-76.2)	72.4 (70.2-75.2)	27.96 (26-29.4)	22.24 (20.8-23.8)
Blood type				
O	2782.4 (2771.8-2790.8)	2801.8 (2773.4-2810.6)	1153.12 (1142.4-1162.4)	945.6 (940-956.4)
A	2177.16 (2171-2182)	2185 (2175-2194.4)	909.36 (892.2-920.4)	757 (747.8-764.2)
B	852.68 (847-861.2)	823.2 (814.6-831.6)	253.9 (245.8-264)	206.64 (198.8-213.6)
AB	365.32 (362.4-366.8)	328.92 (324-331.8)	63.28 (61.6-64)	52.48 (49-57)
Cause of liver disease				
Acute hepatic necrosis	387 (382.4-392.4)	380.2 (371.8-394.6)	103.64 (101.4-109.2)	94.6 (92-97)
Cholestatic liver disease	457.2 (449.2-464.2)	473.4 (466.6-476.2)	160.48 (157-165.4)	131.36 (128-135)
Malignant neoplasms	646.96 (642-658)	376.44 (369.2-381.4)	155.96 (149.8-161)	112.52 (107.2-116.8)
Noncholestatic cirrhosis	3801.68 (3788.2-3818.6)	4201.68 (4195.6-4204.8)	1700.04 (1687.6-1707.6)	1407.52 (1399.6-1417.2)
Other	884.72 (880-889.4)	707.2 (701.8-717.4)	259.56 (255-262.8)	215.72 (207-224.4)
Candidate demographics				
Average age	50.3 (50.2-50.3)	52.81 (52.75-52.89)	54.4 (54.4-54.5)	54.24 (54.18-54.3)
Average cumulative waiting time (d)	152.5 (151.9-154.2)	222.04 (219.55-224.97)	313.1 (312.1-313.9)	331.26 (329.15-332.74)
Average BMI	27.8 (27.8-27.8)	28.12 (28.08-28.16)	28.3 (28.3-28.3)	28.31 (28.27-28.36)

on OPOM scores, rather than Match MELD, resulted in 417.96 (17.6%) fewer deaths each year. The demographic profiles of candidates transplanted through OPOM allocation vs Match MELD, are demonstrated in Table 1. Notably, a higher number of female candidates received transplants when OPOM allocation was utilized. Further analysis demonstrated that OPOM reduced the number of deaths across all United Network for Organ Sharing (UNOS) regions when compared to Match MELD (range of reduction 11.4%-23%; Table 2, Figure 3). In addition, OPOM allocation demonstrated a decrease in waitlist deaths/removals across every disease severity bracket when compared to MELD allocation (Table 3), with the largest reduction in mortality being in those candidates with a MELD score of 16 to 20 (30% decrease).

The simulated average annual number of deaths (waitlist deaths, removed patients' deaths, and posttransplant deaths) by patient status for both models is demonstrated in Table 4. Compared to Match MELD, OPOM decreased deaths of waitlisted candidates by 23.3%, decreased deaths of candidates removed from the waitlist by 21.5%, and decreased posttransplant deaths by 1.8%. OPOM allocated more livers to non-HCC patients, and fewer to HCC patients, when compared to Match MELD. However, OPOM, when compared to Match MELD, decreased the number of waitlist deaths and removals for both HCC patients and non-HCC patients. The overall number of transplantations performed was simulated to remain stable when OPOM allocation was compared to Match MELD (6138.92 vs 6177.56).

3.2 | AUC

OPOM considerably outperformed both MELD variants when predicting the 3-month probability of dying or becoming unsuitable for transplant for all patients (0.859 vs 0.841 for MELD-Na, and 0.823 for Match MELD) and across all exception statuses (Table S2). In addition, analysis of out-of-sample AUC for OPOM, Match MELD, and MELD-Na, for subpopulations of patients with increasing disease severity, revealed a notable decline in predictive power for Match MELD and MELD-Na as disease severity increased, whereas OPOM's predictive power was maintained (Figure 4). The largest divergence in predictive power between OPOM and MELD was at the higher disease severity brackets, with OPOM outperforming Match MELD by up to 16%.

TABLE 2 LSAM simulated average annual deaths and average cumulative waiting time by UNOS region under Match MELD and OPOM allocation

	Match MELD		OPOM		
Region	Average cumulative waiting time to death (d)	Average number of deaths	Average cumulative waiting time to death (d)	Average number of deaths	Percentage reduction in deaths by OPOM (%)
1	336.73 (323.17-343.86)	129.44 (126-132.4)	355.57 (348.89-364.98)	105.32 (102.4-109.4)	18.63
2	335.63 (326.42-340.84)	327.32 (325-328.4)	356.42 (347.45-372.05)	273.84 (268.8-283.6)	16.34
3	161.33 (157.301-166.88)	201.8 (199.4-210.2)	167.93 (164.76-169.43)	155.28 (150-160.8)	23.05
4	306.49 (301.43-310.14)	273.72 (271.2-276.8)	309.03 (300.52-320.89)	234.76 (228.2-238.4)	14.23
5	385.15 (378.21-388.59)	490.92 (484.2-497)	403.42 (390.53-409.23)	404.84 (398.6-410)	17.53
6	260.21 (247.43-272.68)	53.16 (51-54.6)	264.79 (244.83-275.33)	47.12 (44.2-50.2)	11.36
7	316.54 (311.6-321.22)	194.44 (192.8-196.6)	342.41 (334.91-347.66)	160.12 (159-161)	17.65
8	294.59 (279.66-312.52)	144.12 (140.8-147.8)	309.41 (299.36-328.2)	117.52 (114.8-119)	18.46
9	376.05 (372.89-379.59)	234.2 (232.6-235)	404.14 (397.15-410.95)	194.12 (191.4-196.2)	17.11
10	207.78 (202.93-212.49)	159.32 (156.2-161.4)	210.11 (201.88-217.54)	127.36 (123.8-131.6)	20.06
11	280.47 (277.43-285.28)	171.24 (169.6-175.2)	313.19 (303.25-320.11)	141.44 (132.4-148)	17.40
Nationwide	313.1 (312.1-313.9)	2379.68 (2369.8-2393)	331.26 (329.15-332.74)	1961.72 (1950.6-1970)	17.56

Regions with higher waitlist mortality rate, as measured by inverse of average cumulative waiting time, notably Region 3, tended to exhibit higher percentage reduction in deaths (last column).

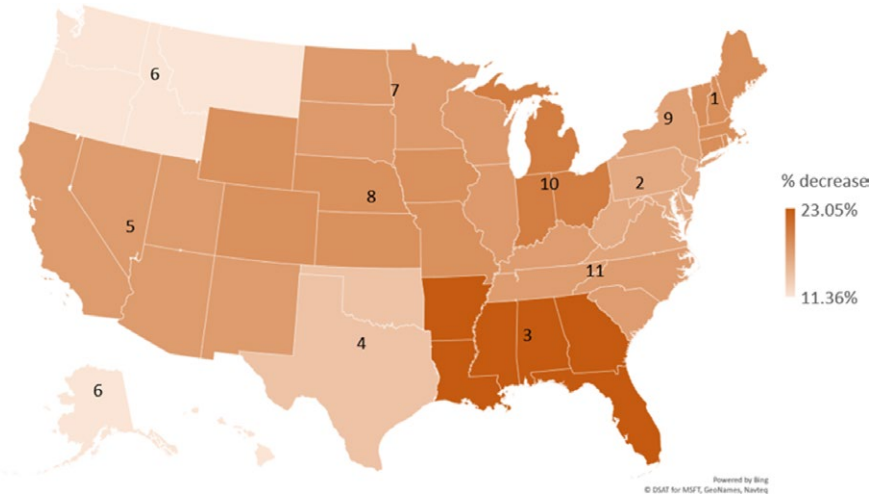

FIGURE 3 LSAM simulated annual percent decrease in deaths by UNOS region using OPOM allocation (as compared to Match MELD) [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 LSAM simulated average annual waitlist deaths reduction with OPOM allocation (as compared to Match MELD allocation)

Reduction in waitlist deaths as categorized by last-known match MELD	
6-10	16.18%
11-15	24.98%
16-20	29.95%
21-25	28.22%
26-30	28.34%
31-35	21.96%
36-40	10.80%

3.3 | Sample match run: OPOM vs MELD

Table 5 depicts LSAM-generated sample match runs for an organ procurement organization in Region 3 for O blood-type candidates simulated to be offered a 66-year-old donation after brain death (DABD) donor. OPOM, compared to MELD, replaced 12 of the top 20 ranked candidates with individuals predicted to have a higher probability of waitlist death/removal. Indeed, of the 12 candidates introduced by OPOM, 8 were simulated to experience waitlist death/removal. Conversely, of the 12 candidates removed by OPOM, 2 were simulated to experience waitlist death/removal.

4 | DISCUSSION

For almost 2 decades now, MELD has served as the scoring system used to rank liver transplant candidates on the waitlist. Although it is the case that the MELD score and its components (bilirubin, INR, and creatinine) are effective predictors of 3-month mortality, they are not the *only* relevant predictors. Indeed, although a simple method to stratify candidates awaiting liver transplantation, the MELD score is a linear regression method that does not accurately predict mortality for all candidates who can benefit from liver transplantation. The latter is demonstrated by our results demonstrating a significant deterioration in MELD predictive capabilities with increasing disease severity when compared to OPOM. It is important to note that it is the candidates with the highest disease severity that warrant the most accurate mortality prediction, to in return allow for the most accurate prioritization on the liver transplant waitlist. Differentiation within the latter cohort of the highest disease acuity represents the greatest challenge of this prediction problem. In contrast to MELD, which demonstrated decreasing AUC values as sicker patient strata are considered, OPOM maintained significantly higher AUCs, especially within the sickest candidate population, thus allowing for a more accurate prediction of waitlist mortality.

The use of MELD exception points within the current scoring system has represented an arbitrary, yet advantageous, solution for certain subpopulations of candidates, most notably those candidates with HCC. Indeed, Berry and Ioannou, through a competing risks analysis, demonstrated a near-complete lack of survival benefit among patients undergoing liver transplantation on the basis of MELD exception points, and thus calling into question the need for a system that artificially raised MELD scores.[7] The latter "HCC advantage" has been addressed through first serial downgrades in the amount of MELD exception points granted, and subsequently, more recently, with both a delayed initiation of MELD exception points (6-month delay), as well as a cap on the extent of points an individual can achieve (MELD 34 cap).[8] These modifications have been implemented with the hopes of decreasing waitlist mortality and increasing transplant rates in the non-HCC population; however, they have thus far represented insufficient and inexact changes in adequately equalizing access to liver transplants for the non-HCC population. Although well intentioned, the quest to equalize priority between the HCC and non-HCC candidates has been fundamentally inadequate, as they have utilized the assignment of exception points based on an imprecise mortality prediction.

Herein, we introduce OPOM, a novel system based on a state-of-the-art machine-learning method that has allowed for a more

TABLE 4 LSAM simulated average annual deaths and transplants by candidate status and exception status under Match MELD and OPOM allocation

	Match MELD	OPOM
Deaths by patient status		
Waitlist deaths	1201.76 (1198.4-1204.2)	922.12 (916-926.8)
Removed patient deaths	593.84 (590.2-599.8)	466.28 (463.4-470)
Postgraft deaths	584.08 (572.8-539.2)	573.32 (565.4-581)
All	2379.68 (2369.8-2393)	1961.72 (1950.6-1970)
Deaths by patient exception status		
HCC patients	293.8 (289-298)	212.04 (208.2-215.8)
Non-HCC patients	2085.88 (2075-2097.4)	1749.68 (1735.2-1761.8)
All patients	2379.68 (2369.8-2393)	1961.72 (1950.6-1970)
Removals by patient exception status		
HCC patients	255.32 (252.4-258.4)	237.16 (234.8-240.4)
Non-HCC patients	2153.8 (2404.4-2416.4)	1961.24 (1953.2-1969.4)
All patients	2409.12 (2404.4-2416.4)	2198.4 (2191.8-2209.8)
Transplants by patient exception status		
HCC patients	1178.72 (1171.4-1190.2)	690.96 (681.6-701.6)
Non-HCC patients	4998.84 (4994-5002.6)	5447.96 (5438-5457)
All patients	6177.56 (6166-6184.2)	6138.92 (6127-6148.8)

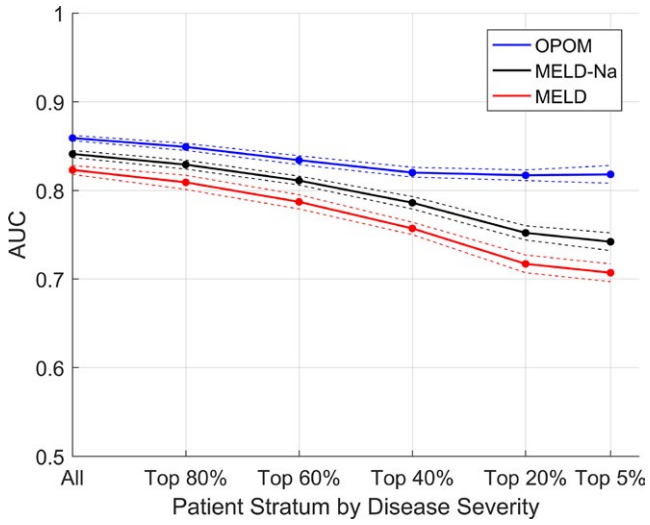

FIGURE 4 Out-of-sample AUC for OPOM by disease severity (as measured by Match MELD), and out-of-sample AUC for MELD-Na, and Match MELD, by disease severity (as measured by OPOM) [Color figure can be viewed at wileyonlinelibrary.com]

accurate prediction of 3-month mortality rate for *all* patients on the liver transplant waitlist. OPOM allocation outperformed the currently used MELD-based prediction method. In simulations, OPOM averted significantly more waitlist deaths/removals for both HCC and non-HCC candidates, and yet maintained overall transplant rates, thereby allowing for more equitable and efficient allocation of liver grafts for candidates awaiting transplantation across all levels of disease severity. As demonstrated using LSAM, the use of OPOM in place of current Match MELD scores, would save on average at least 418 more lives each year, with every UNOS region benefiting from this effect. It is notable that the overall number of transplants remains stable with OPOM allocation, albeit with an acceptable, and expected, decrease in HCC transplants to accommodate the increase in transplants of nonexception point candidates. Unlike MELD allocation, which relies on the cumbersome and inexact approach of exception point assignment, OPOM allows for accurate prioritization of all candidates based on individual characteristics, thus negating MELD's varying levels of success in predicting mortality for different patient populations. For candidates with hepatocellular cancer, OPOM's predictive ability is strengthened by the incorporation within the model of α-fetoprotein (AFP) levels, as well as tumor size and number.

The accurate prediction of an individual candidate's risk of waitlist mortality/removal is paramount to ensure equitable access to liver transplantation. Whereas on the one hand MELD-based allocation with inclusion of exception points has overprioritized exception point candidates at the expense of those candidates listed with lab MELD scores, on the other hand, utilizing only a lab-based MELD score for waitlist prioritization would shift the pendulum in the opposite direction, resulting in an allocation process that greatly underserves those in need of a liver transplant but with a lab MELD score that does not reflect their severity of disease. OPOM achieves an evidence-based, unbiased, and objective middle ground for all waitlisted candidates by utilizing multiple variables with associated trajectories. Notably, there is a higher number of transplants in the female population with OPOM allocation, perhaps overcoming the systematic bias noted in MELD-based allocation for female candidates.[9,10] The latter has been attributed to the inability of MELD to accurately capture the female candidate's degree of renal insufficiency based on serum creatinine levels, resulting in lower MELD scores and thus lower transplantation rates. OPOM has provided

TABLE 5 LSAM simulated match runs in a Region 3 organ procurement organization for blood type O under Match MELD (top table) and OPOM (bottom table) allocation

Match run under match MELD allocation					
Rank order	**Candidate ID**	**Waiting time (d)**	**Age (y)**	**MELD score**	**Waitlist death/removal**
1	A	7.89	0.8	Status 1A	
2	B	0.27	62.6	25	
3	C	20.33	56.16	24	
4	D	6.24	52.1	23	
5	E	50.94	61.99	22	
6	F	13.59	62.03	22	
7	G	72.4	59.06	22	Yes
8	H	476.98	62.98	19	
9	I	22.52	62.85	19	
10	J	227.13	61.16	18	Yes
11	K	1240.82	55.67	18	Yes
12	L	156.98	54.23	18	
13	M	778.88	50.99	17	
14	N	294.31	70.35	16	Yes
15	O	178.6	56.07	16	
16	P	430.45	60.61	16	
17	Q	27.52	63.06	16	Yes
18	R	574.66	67.05	15	
19	S	287.19	57.46	15	
20	T	246.28	48.2	15	
…	…	…	…	…	…
Match run under OPOM allocation					
Rank order	**Candidate ID**	**Waiting time (d)**	**Age (y)**	**OPOM score**	**Waitlist death/removal**
1	A	7.89	0.80	Status 1A	
2		50.76	73.58	40	Yes
3	B	0.27	62.60	25	
4		1074.37	69.22	23	Yes
5	D	6.24	52.10	23	
6	N	294.31	70.35	22	Yes
7	H	476.98	62.98	22	
8		310.90	64.46	22	Yes
9		771.13	69.38	22	Yes
10		937.42	59.64	21	Yes
11		36.56	24.57	20	
12	Q	27.52	63.06	19	Yes
13		57.03	66.70	19	
14		560.52	62.99	19	Yes
15		2666.90	53.94	18	Yes
16	O	178.60	56.07	18	
17		146.37	67.29	18	
18		170.14	58.80	18	Yes
19		148.94	66.92	18	
20	K	1240.82	55.67	17	Yes
…	…	…	…	…	…

OPOM, compared to Match MELD, replaced 12 of the top 20 candidates with individuals with a higher predicted probability of waitlist death or removal. The remaining 8 candidates on the original Match MELD rank order list, with the exception of the Status 1A candidate (Rank Order 1), experienced a change in their rank order under OPOM allocation. The last column reports whether the patient was simulated to die or be removed from the waitlist by becoming unsuitable for transplant.

a more complete picture of the individual candidate's true waitlist mortality that in return has allowed for a more accurate prediction of need for liver transplantation. Although there is a decrease noted in transplants for Black and Asian candidates with OPOM allocation, with an increase in White and Hispanic patients transplanted, it should be noted that there is also a decrease in waitlist deaths for all of these candidate populations.

The 418 waitlist deaths averted with OPOM utilization is significantly more than the number predicted with implementation of MELD-Na. Indeed, MELD-Na, which was approved by UNOS in June 2014 and implemented in January 2016, was predicted through similar LSAM analyses to decrease waitlist deaths by only 52 patients a year.[3] Similarly, the application of a 6-month delay in awarding exception points for HCC candidates was simulated in LSAM to achieve a higher rate of transplants for non-HCC candidates, at the expense of a lower transplant rate for HCC candidates. The downstream effect on waitlist mortality with this proposed change, compared with the current policy, was a net reduction of only 30 deaths in the non-HCC population. The latter policy was adopted in October 2015, and much like the acceptance of MELD–Na, although well intentioned, represented nominal changes in waitlist mortality in simulations when compared to liver allocation through OPOM. Although the actual number of waitlist deaths averted with LSAM under OPOM allocation may represent an overestimation, it is important to note the ability of LSAM to predict the overall directionality of change.[11]

Machine learning holds the potential to become an indispensable tool for clinicians, with optimized predictions based on large amounts of data.[12,13] OCTs are a state-of-the-art machine-learning method.[4] OCTs are decision trees similar to the classification and regression trees (CARTs), but are solved to global optimality with a novel method using mixed-integer optimization that outperforms the classical CART algorithms.[14] We utilized OCTs to develop an analytical tool that takes all available patient information to predict whether the waitlisted candidate will undergo the adverse events of either death or becoming unsuitable for transplantation within 3 months. In contrast to the piecemeal way in which current policy has been constructed, our tool is trained on historical outcomes in a unified fashion, utilizing millions of data points. Instead of adding in exceptions and cutoffs ex post to decrease mortality on the waitlist, machine-learning analytical tools tackle the problem directly by building these different criteria into the model itself. The out-of-sample AUC and accuracy illustrated that OPOM performs well not only on patients without exceptions, but also on patients with HCC exceptions. Furthermore, the OPOM advantage over MELD is most notable among sicker patient populations, with OPOM outperforming both match MELD and MELD-Na in AUC analysis, thus allowing OPOM to achieve a greater "sickest-first" allocation policy.

The use of readily available, reproducible, and objective data that accurately predict liver-related mortality is essential. Although OPOM utilizes a larger number of variables than MELD does, it is important to note that many of these additional variables are linked to MELD, and it is the trajectories of change in these lab values that power OPOM's accuracy. The latter concept is in line with studies examining the utility of changes in MELD scores for both waitlist and posttransplant mortality prediction, as well as liver transplant allocation.[15,16] At first glance OPOM's complexity, in comparison to MELD, can be overwhelming. However, it is notable that no additional data collection would be required by the transplant practitioner, as OPOM was generated based on available data within the STAR files, data that are routinely collected on all waitlisted candidates. Furthermore, OCTs are versatile tools that can allow for additional variables to be included/excluded with ease should additional priorities in liver allocation require that OPOM be modified. This could be crucial for appropriate allocation to the group of candidates with non-HCC standardized exceptions (eg, those candidates with hepatopulmonary syndrome, portopulmonary hypertension, and so on), and those candidates with non-HCC, non-standardized exceptions.[17] Although for the purposes of the initial creation and application of OPOM these non-HCC exceptions populations were grouped in the non-HCC patients population, they nonetheless would benefit from an optimized prediction method based on incorporation of consensus variables that accurately gauge their risk of mortality. To this point, granularity in the varying types of MELD exceptions within LSAM would also allow for a more accurate assessment of the differing classes of exception-point candidates, instead of a simple HCC vs non-HCC candidate comparison. It should be noted that LSAM analysis is also limited in that it only allows for an accurate assessment of waitlist deaths, as waitlist removals include not only candidates with deterioration in their condition, but also those removed due to improvement in their condition. Although additional analysis with consideration of a shorter or longer interval of waitlist risk could be considered, the risk of waitlist mortality at the 3-month interval was assessed to allow for accurate comparisons to MELD score calculations. Despite these limitations, and the fact that LSAM cannot account for center or practitioner changes in listing or acceptance behavior, LSAM remains the current simulation model employed to assess and implement national policy changes in liver allocation and distribution.

It should be noted that OPOM allocation does not address the issues in liver distribution, and the resultant geographic disparity that exists between UNOS regions and donor service areas (DSAs). However, it is worth noting that the application of this machine-learning tool is capable of saving an additional 418 lives every year—of the same magnitude as that achieved with LSAM models of wide broader sharing. Thus, the implementation of OPOM represents an avenue to achieve more equitable liver allocation within any defined geographic unit.

The application of an OPOM-based allocation system would more accurately adhere to the "sickest-first" principle. Indeed, the decrease in waitlist mortality/removal achieved through utilization of OPOM would not only represent the potential for more equitable allocation, but also would represent an important facet toward alleviating the discrepancy between supply and demand.

ACKNOWLEDGMENTS

The authors would like to thank Jack Dunn for his help with developing http://www.opom.online and Drs. Jorge Marrero and Richard Freeman for their helpful comments regarding this work. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the SRTR. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

AUTHOR CONTRIBUTIONS

Problem definition: DB, JK, NT, YW, RH, PV; development of approach: DB, JK, NT, YW; implementation: JK, NT, YW; writing: DB, JK, NT, YW, PV; editing: DB, JK, NT, YW, RH, PV.

DISCLOSURE

The authors of this manuscript have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

ORCID

Dimitris Bertsimas https://orcid.org/0000-0002-1985-1003
Nikolaos Trichakis https://orcid.org/0000-0002-8324-9148
Yuchen Wang https://orcid.org/0000-0003-2139-9279
Ryutaro Hirose https://orcid.org/0000-0002-2307-0170
Parsia A. Vagefi https://orcid.org/0000-0002-0461-5438

REFERENCES

1. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124:91-96.
2. Massie AB, Caffo B, Gentry SE, et al. MELD exceptions and rates of waiting list outcomes. *Am J Transplant*. 2011;11:2362-2371.
3. Elwir S, Lake J. Current status of liver allocation in the United States. *Gastroenterol Hepatol*. 2016;12:166.
4. Bertsimas D, Dunn J. Optimal classification trees. *Mach Learn*. 2017;106:1039-1082.
5. Bertsimas D, Pawlowski C, Zhuo Y. From predictive methods to missing data imputation: an optimization approach. *J Mach Learn Res*. 2018;18(196):1-39.
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
7. Berry K, Ioannou GN. Comparison of liver transplant–related survival benefit in patients with versus without hepatocellular carcinoma in the United States. *Gastroenterology*. 2015;149:669-680.
8. Heimbach JK, Hirose R, Stock PG, et al. Delayed hepatocellular carcinoma model for end-stage liver disease exception score improves disparity in access to liver transplant in the United States. *Hepatology*. 2015;61:1643-1650.
9. Cholongitas E, Marelli L, Kerry A, et al. Female liver transplant recipients with the same GFR as male recipients have lower MELD scores—a systematic bias. *Am J Transplant*. 2007;7:685-692.
10. Moylan CA, Brady CW, Johnson JL, et al. Disparities in liver transplantation before and after introduction of the MELD score. *JAMA*. 2008;300:2371-2378.
11. Goel A, Kim WR, Pyke J, et al. Liver simulated allocation modeling: were the predictions accurate for share 35? *Transplantation*. 2018;102(5):769-774.
12. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216.
13. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *JAMA*. 2016;315:551-552.
14. Breiman L, Friedman J, Olshen RA, Stone CJ. *Classification and Regression Trees*. Boca Raton, FL: CRC Press; 1984.
15. Freeman RB. Mathematical models and behavior: assessing delta MELD for liver allocation. *Am J Transplant*. 2004 Nov;4(11):1735-1736.
16. D'Amico G. Developing concepts on MELD: delta and cutoffs. *J Hepatol*. 2005 Jun;42(6):790-792.
17. Goldberg DS, Olthoff KM. Standardizing MELD exceptions: current challenges and future directions. *Curr Transpl Rep*. 2014;1(4):232-237.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bertsimas D, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant*. 2019;19:1109–1118. https://doi.org/10.1111/ajt.15172