

# Journal Pre-proof

MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era

W. Ray Kim, MD, Ajitha Mannalithara, PhD, Julie K. Heimbach, MD, Patrick S. Kamath, MD, Sumeet K. Asrani, MD, Scott W. Biggins, MD, Nicholas L. Wood, PhD, Sommer E. Gentry, PhD, Allison J. Kwong, MD



PII: S0016-5085(21)03469-7  
DOI: <https://doi.org/10.1053/j.gastro.2021.08.050>  
Reference: YGAST 64558

To appear in: *Gastroenterology*  
Accepted Date: 18 August 2021

Please cite this article as: Kim WR, Mannalithara A, Heimbach JK, Kamath PS, Asrani SK, Biggins SW, Wood NL, Gentry SE, Kwong AJ, MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era, *Gastroenterology* (2021), doi: <https://doi.org/10.1053/j.gastro.2021.08.050>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 by the AGA Institute

**MELD 3.0: The Model for End-stage Liver Disease Updated for the Modern Era**

W. Ray Kim, MD<sup>1</sup>, Ajitha Mannalithara, PhD<sup>1</sup>, Julie K. Heimbach, MD<sup>2</sup>, Patrick S. Kamath, MD<sup>2</sup>, Sumeet K. Asrani, MD<sup>3</sup>, Scott W. Biggins, MD<sup>4</sup>, Nicholas L. Wood, PhD<sup>5</sup>, Sommer E. Gentry, PhD<sup>5</sup>, Allison J. Kwong, MD<sup>1</sup>

<sup>1</sup>Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Stanford, CA

<sup>2</sup>Transplant Center, Mayo Clinic, Rochester, MN

<sup>3</sup>Baylor University Medical Center, Dallas, TX

<sup>4</sup>University of Washington, Seattle, WA

<sup>5</sup>Department of Mathematics, United States Naval Academy, Annapolis, MD

**Grant Support:** This work was funded by the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK-034238). Dr. Kwong is supported by the National Institute of Allergy and Infectious Diseases (R25 AI-147369), the National Institute on Alcohol Abuse and Alcoholism (K23 AA-029197), the AASLD Foundation Clinical, Translational, and Outcomes Research Award. The funding organizations played no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

**Abbreviations:** Alcohol-associated liver disease (ALD), Child-Turcotte-Pugh (CTP), international normalized ratio (INR), Model for End-Stage Liver Disease (MELD), United States (US)

**Correspondence:**

W. Ray Kim

**Address:** 430 Broadway Street, Floor 3, Redwood City, CA 94063-3132

**Telephone:** 650-723-5135

**Fax:** 650-723-5488

**E-mail:** [wrkim@stanford.edu](mailto:wrkim@stanford.edu)

**Disclosures:** None declared for all authors.

**Author Contributions**

W. Ray Kim: Study concept and design; analysis and interpretation of data; drafting of the manuscript; obtained funding; study supervision

Ajitha Mannalithara: Acquisition of data, statistical analysis

Julie K. Heimbach: Critical revision of the manuscript for important intellectual content

Patrick S. Kamath: Critical revision of the manuscript for important intellectual content

Sumeet K. Asrani: Critical revision of the manuscript for important intellectual content

Scott W. Biggins: Critical revision of the manuscript for important intellectual content

Nicholas Wood: Statistical analysis, critical revision of the manuscript for important intellectual content

Sommer Gentry: Statistical analysis

Allison J. Kwong: Analysis and interpretation of data, drafting of the manuscript, study supervision

**ABSTRACT**

**Background:** The model for end-stage liver disease (MELD) has been established as a reliable indicator of short-term survival in patients with end-stage liver disease. The current version (MELDNa), consisting of INR and serum bilirubin, creatinine, and sodium, has been used to determine organ allocation priorities for liver transplantation in the United States (US). The objective was to optimize MELD further by taking into account additional variables and updating coefficients with contemporary data.

**Methods:** All candidates registered on the liver transplant waitlist in the US national registry from Jan 2016 – Dec 2018 were included. Uni- and multivariable Cox models were developed to predict survival up to 90 days after waitlist registration. Model fit was tested using the concordance statistic and reclassification, and the liver simulated allocation model (LSAM) was used to estimate the impact of replacing MELDNa with the new model.

**Results:** The final multivariable model was characterized by (1) additional variables of female sex and serum albumin, (2) interactions between bilirubin and sodium and between albumin and creatinine, and (3) an upper bound for creatinine at 3.0mg/dL. The final model (MELD 3.0, henceforth), had better discrimination than MELDNa (concordance statistic 0.869 versus 0.862,  $p < 0.01$ ). Importantly, MELD 3.0 correctly reclassified a net of 8.8% of decedents to a higher MELD tier, affording them a meaningfully higher chance of transplant, particularly in women. In the LSAM analysis, MELD 3.0 resulted in fewer waitlist deaths compared to MELDNa (7,788 versus 7,850,  $p = 0.02$ ).

**Conclusion:** MELD 3.0 affords more accurate mortality prediction in general than MELDNa and addresses determinants of waitlist outcomes including the sex disparity.

**Keywords:** End-stage liver disease, waitlist mortality, outcome prediction

Journal Pre-proof

## Background

Since its original description, the Model for End-Stage Liver Disease (MELD) has proven to be a reliable predictor of short-term survival in patients with end-stage liver disease.<sup>1</sup> The current version of the MELD score, commonly referred to as MELDNa, incorporates serum concentrations of total bilirubin, creatinine and sodium, and the international normalized ratio (INR) of prothrombin time. MELDNa has been utilized to determine priorities for allocation of livers for transplant in the US since 2016.<sup>2</sup>

More recently, questions have been raised whether the accuracy of prediction of mortality by MELD may have decreased.<sup>3</sup> There may be a number of potential reasons for the concern, ranging from changes in liver disease epidemiology and development of therapies that alter disease prognosis to changes in the distribution of MELD scores and increasing age and comorbidity in patients awaiting transplant. In addition, there has been a growing concern that women are disadvantaged in the current system for a number of reasons, including serum creatinine overestimating renal function in women and thus underestimating their risk of mortality.<sup>4</sup>

Even before these observations were reported, many attempts have been made to improve MELD. A common approach is to incorporate additional variables. An important historical perspective is that a large part of the acceptance of MELD was the lack of variables that could be subjectively interpreted. Thus, while the Child-Turcotte-Pugh (CTP) score has proven to be a highly useful clinical tool to assess severity of hepatic decompensation, the advantage of MELD is that the data elements are verifiable and auditable for policy implementation.<sup>5</sup> A relevant, recent example may be sarcopenia and frailty, which have been consistently associated with poor prognosis in patients with many chronic illnesses including end-stage liver disease.<sup>6</sup> Similar to ascites and encephalopathy, however, these variables are not as objectively verifiable as laboratory data, making them difficult to be included for the purpose of allocation.

In this work, we set out to investigate whether the fit of the MELD score could be further optimized by considering alternate coefficients or by including additional variables in predicting short term mortality in the modern era. Some of the principles that guided our work included (1) consideration of biomedical insight in addition to statistical significance in determining model parameters, and (2) incorporation of objective, generalizable and easily verifiable variables. Under the current urgency-based liver allocation policy in the United States, the role of the MELD score is to inform the organ allocation system of the biological predictors of mortality, independent of the transplant policy and practices that could also affect waitlist outcome, such as donor-recipient size matching, geography, or healthcare access to transplant.

## **Methods**

### *Patients and Data Elements*

The main portion of this analysis was performed on the OPTN Standard Transplant Analysis and Research (STAR) files with data curated as of 03/15/2019. For the purpose of the analysis, the OPTN data, consisting of liver transplant candidates waitlisted in the United States, represent the population to which the results are directly applicable. Out of the data set, we created a cohort of liver transplant candidates newly waitlisted between 01/15/2016 and 12/31/2018. The primary inclusion criteria for the analysis were (1) adults aged 18 years or older (2) registered for primary liver transplant with (3) end-stage liver disease. Patients listed for (1) multi-organ transplant, other than simultaneous liver-kidney transplant, those with (2) history of previous liver transplant, and those with (3) exception points at the time of registration, were excluded. These inclusion and exclusion criteria were consistent with prior iterations of the MELD score.

The cohort was then randomly divided in a 70:30 ratio into model development and validation data sets. A wide array of variables was extracted as potential predictors of waitlist survival, including

demographic information, components of MELD and CTP scores, and additional laboratory parameters. In the selection of the variables, the same principle was used as in the original MELD score that the variables must be measurable in an objective fashion and generalizable. As such, ascites and encephalopathy were excluded from the model development. Age, sex, race, serum sodium, creatinine, INR, bilirubin, albumin, and height were considered for inclusion in the model. Estimated GFR (eGFR), as a better measure of renal function, was not considered, since the most common estimating equations, MDRD-4 and CKD-EPI, include race in addition to sex and serum creatinine. While the latter two variables are already in the mix, the inclusion of race could be problematic — given the same serum creatinine, the estimated GFR in a black patient would be calculated to be higher, potentially underestimating the risk of death and magnifying racial inequity in access to liver transplantation.<sup>7,8</sup> A non-race based measure such as cystatin C would be preferable but is not widely available.

Given the prior literature on the potential impact of height on the probability of transplant and waitlist mortality, we conducted an exploratory analysis considering height and sex as potentially confounding variables. The overall result was that sex and height were collinear, which makes a model containing both terms suboptimal and the coefficients unreliable. Among women < 175 cm, there was a higher risk of waitlist mortality that decreased linearly with increasing height — whereas among men, height had no effect (Supplementary Figure 1). In considering multivariable models with separate terms for height in both men and women, we determined that the effect of sex was larger and more consistent than that of height. With or without height in the model, the other variables remained remarkably consistent in terms of coefficients and statistical significance. Thus, sex was selected over height for inclusion in the final model (Supplementary Table 1).

For calculation of MELD and MELDNa, the following standard formulas were used as previously described:

**MELD** =  $9.57 \times \log_e(\text{creatinine}) + 3.78 \times \log_e(\text{bilirubin}) + 11.20 \times \log_e(\text{INR}) + 6.43$ ,

where creatinine (mg/dL), bilirubin (mg/dL) and INR values below 1.0 were set to 1.0 and creatinine values to 4.0 mg/dL if serum creatinine was  $\geq 4$  mg/dL or the patient received 2 or more dialysis treatments within the prior week.<sup>1,9</sup> The resulting score was rounded to the nearest whole number to yield the MELD score.

**MELDNa** =  $\text{MELD} + 1.32 \times (137 - \text{Na}) - [0.033 \times \text{MELD} \times (137 - \text{Na})]$ , where the serum sodium concentration (Na) is bound between 125 and 137 mmol/L, as defined by the OPTN.<sup>9</sup> The resulting score was rounded to the nearest whole number to yield the MELDNa score. For the purpose of organ allocation in the US, MELDNa is applied only if MELD is greater than 11.

#### *Data Analysis*

The main outcome variable in our time-to-event analysis was survival up to 90 days from the time of waitlist registration, a time frame used in prior work to develop and validate MELD and MELDNa. Waitlist mortality was defined as removal from the waitlist for death or being too sick. Surviving patients were censored at (1) 90 days from waitlist registration, (2) waitlist removal for transplant or another reason than death or being too sick to transplant, (3) receipt of exception points for any reason, or (4) December 31, 2018, whichever occurred first. This setup was similar to prior iterations of the MELD score. Survival was estimated using Kaplan-Meier methods. The Kaplan-Meier estimate and Cox proportional hazards model consider survival probability without transplant, which is appropriate in the context of developing a score to rank patients based on their mortality risk. By contrast, a competing risk analysis, which treats liver transplantation as a competing event, would be relevant to analyze the survival probability in the presence of a transplant system, e.g. to investigate waitlist disparities or effects of organ allocation policy.

The initial approach was to evaluate individual variables that are associated with 90-day mortality by the univariate proportional hazard (Cox) regression analysis. For laboratory variables found to be predictive of survival, a generalized additive model form of the Cox model was applied, which describes the relation between each variable and risk of death in a flexible shape via a smoothing spline.<sup>10</sup> Goodness of fit for each variable with and without logarithmic transformation was compared using partial likelihood ratio tests. The resulting fit was assessed both visually and with formal tests for linearity and/or significance. These models were executed in a multivariable fashion — in determining the effect of one variable, all other variables in the model were considered simultaneously. Thus, the relationship between the first variable and mortality can be identified as independent of the effects of the other variables.

Using the smoothing splines, we examined the extent to which the relationship between each variable and the risk of death is linear and whether setting lower and upper bounds — the limits beyond which linearity of the relationship breaks down — would improve the fit.<sup>10</sup> This was accomplished first by visual inspection, followed by formal testing for the non-linearity at the putative lower or upper bounds. The presence/absence of the bounds and the cutoff values for each variable were examined in an iterative fashion until the optimal bounds were found. The final determination of the upper or lower bounds was not only based on statistical significance, but also on clinical interpretation of the data.

Once individual variables with multivariable significance were identified, we considered possible two-way interactions between the variables. The final multivariable Cox regression model consisted of independently significant variables and interaction terms. A risk score was created as sum of the products between the coefficients and variables (and relevant interactions). The score then was rescaled to have a similar distribution compared to that of MELDNa. We elected to set the 80th percentile on both MELDNa and the new score to coincide. This was achieved by identifying the 80th percentile of MELDNa in the model development data set, subtracting the constant (i.e., 6), calculating

the multiplier needed to rescale the new model to equate its 80th percentile score with that of MELDNa (minus 6), and then finally adding back the constant.

Once these models were constructed, we assessed their performance against MELD and MELDNa in the validation data set. First, discrimination by the model, namely its ability to rank patients according to the risk of death within 90 days, was evaluated using the concordance (c) statistic. Of the several methods to calculate concordance, methods by Harrell and Uno were used.<sup>12,13</sup> Second, reclassification by MELD 3.0 vis-à-vis MELDNa was described for the number of patients, number of deaths and the proportion of deaths. Patients were divided by the two scores in five tiers (6-9, 10-19, 20-29, 30-39, 40+) and a 5x5 table was created for each of the metric above. The proportions of decedents correctly (MELD 3.0 tier > MELDNa tier) and incorrectly (MELDNa tier < MELD 3.0 tier) reclassified were calculated.

We conducted a sensitivity analysis in which the entire modeling procedure was repeated removing albumin as a candidate variable. Since the inception of MELD, serum albumin has been considered as a potential variable in MELD. While hypoalbuminemia is a well-known physiological consequence of liver dysfunction, there has been a concern that the serum albumin concentration may be temporarily raised by external administration and thus, incorporating hypoalbuminemia in liver allocation might discourage albumin infusion even when it is clinically indicated.<sup>14</sup> A second temporal validation analysis was performed for liver transplant candidates listed in 2019, to test robustness of the model using more recent data. From the STAR file, waitlist registrants between 1/1/2019 and 12/31/2019, not overlapping with the main analysis set, were selected using the eligibility criteria listed above. Although some data were available, listings in 2020 and afterward were not considered due to the unpredictable impact of the COVID-19 pandemic. We also considered etiology of liver disease, given the recent rise in patients undergoing liver transplant for alcohol-related liver disease (ALD).

Finally, MELD 3.0 with and without albumin was compared to MELDNa using the liver simulated

allocation model (LSAM) provided by SRTR, a discrete event simulator that uses historical data to model the US liver allocation system and can predict the effects of changes to liver allocation policy on waitlist outcomes. We ran 10 replications of liver allocation for the time period from July 1, 2013 to June 30, 2016 under 3 allocations schemes: MELDNa, MELD 3.0, and MELD 3.0 without albumin. Results for the number of waitlist deaths were averaged across the 10 LSAM iterations over the 3-year study period and compared to MELDNa via matched pair t-tests.

For all analyses, a p-value of  $< 0.05$  was considered significant, and all tests were 2-tailed. In descriptive analyses, variables were compared among groups using the t-test, the chi-square test, the 1-way analysis of variance, and the Wilcoxon rank-sum test, as appropriate. Statistical analyses were performed using SAS 9.4 (Cary, NC) and R 3.6.2 (Vienna, Austria). The study, consisting of analysis of deidentified data, was deemed exempt from the Institutional Review Board at Stanford University.

## Results

During the study period (2016-2018), there were 29,410 eligible patients listed for liver transplant. Supplementary Figure 2 describes formation of the cohort, which was then divided into a development set ( $n=20,587$ , 70%) and a validation set ( $n=8,823$ , 30%). As expected, the two sets were similar to each other with no significant difference in age, sex, race, or liver disease severity (Table 1). The median age of the development set was 58 years (interquartile range [IQR] 51-64) and 37% were women. Ascites and hepatic encephalopathy were present in 73% and 60%, respectively. The median MELD was 16 (IQR 11-23), median MELDNa 18 (IQR 11-25) and median CTP score 9 (IQR 7-11).

In the development set, the 90-day Kaplan-Meier survival was 91.3%. Supplementary Table 2 represents results of the univariate Cox model analyzing survival up to 90 days. All of the variables considered were significantly associated with death within 90 days, including female sex, MELDNa and all of its components, and serum albumin.

Figure 1 illustrates smoothing splines for the five laboratory variables considered, namely total bilirubin, creatinine, INR, sodium and albumin. Logarithmically transformed variables produced better fit for total bilirubin, creatinine and INR, whereas the natural scale was appropriate for sodium and albumin. With total bilirubin and INR, the risk of death rose continuously with no apparent lower or upper limit. Serum creatinine was linear up to a point, beyond which the risk did not increase further. Based on the p-spline and clinical insights, serum creatinine of 3.0 mg/dl was selected as the inflection point. Consistent with prior versions of MELD, bilirubin, INR and creatinine values below 1.0 were set to 1.0. Both serum sodium and albumin displayed a U-shaped relation. However, as hyponatremia and hypoalbuminemia are the main physiological consequences of worsening end-stage liver disease, we only modeled the lower aspects of the curves. The lower and upper bounds of the current MELDNa for serum sodium, namely 125 mEq/L and 137 mEq/L, respectively, were still appropriate, whereas for serum albumin, lower and upper bounds of 1.5 g/dL and 3.5 g/dL, respectively, were selected.

Taking into account these details of each predictor variable, we constructed a multivariable Cox model predicting mortality up to 90 days. Considered in the model were not only the individual variables but also possible interactions between them. Supplementary Table 3a summarizes the final model, which includes female sex, total bilirubin, INR, creatinine, sodium and albumin. In addition, significant interactions were found between bilirubin and sodium and between creatinine and albumin. The resulting risk estimating equation, noted in the table, was then rescaled such that the lowest score would be 6 and the 80<sup>th</sup> percentile score 28, arriving at the following formula:

$$\begin{aligned} \text{MELD 3.0} = & 1.33 \text{ (if female)} + 4.56 * \log_e(\text{bilirubin}) + 0.82 * (137 - \text{Na}) - 0.24 * (137 - \text{Na}) * \log_e(\text{bilirubin}) \\ & + 9.09 * \log_e(\text{INR}) + 11.14 * \log_e(\text{creatinine}) + 1.85 * (3.5 - \text{albumin}) - 1.83 * (3.5 - \text{albumin}) * \log_e \\ & (\text{creatinine}) + 6, \text{ which is rounded to the nearest integer.} \end{aligned}$$

Supplementary Table 3b represents the survival function for the mortality prediction model: for a patient with the average risk score in the development set (MELD 3.0 = 20), predicted mortality was 1.9% at 30 days and 5.4% at 90 days. Examples in Supplementary Table 3c illustrate MELD 3.0 scores for men and women with laboratory variables at the 50<sup>th</sup> and 75<sup>th</sup> percentiles and their predicted survival.

In the validation set, the median MELD 3.0 score was 19 (IQR 13-26) with 3.4% of subjects having scores of > 40. Of 8,823 candidates in the set, 318 died within 30 days and 514 within 90 days. The concordance statistic for 90 day mortality of MELD 3.0 was 0.8693 and that of MELDNa 0.8622 (Harrell's method, Table 3). Although the numerical difference appeared modest, the difference was statistically significant ( $p < 0.01$ ).

Table 2 demonstrates reclassification between MELDNa and MELD 3.0. The distribution of both MELDNa and MELD 3.0 scores was skewed to the right with 51% of the patients having both MELDNa and MELD 3.0 < 20. There were more patients up-categorized ( $n=890$ , 10.1%) in general than down-categorized ( $n=306$ , 3.5%). Out of the 514 decedents, 435 (84.6%) remained in the same score categories, while 62 (12.1%) were correctly reclassified (up-categorized) and 17 (3.3%) were incorrectly reclassified (down-categorized), with a net gain of 45 (8.8%). The more meaningful shift may be in patients who were registered with MELDNa of 20-29 ( $n=195$ ) and 30-39 ( $n=168$ ) and died on the list. As 11.8% and 11.3% of those patients would have gained enough points to be up-categorized to the 30-39 and 40+ categories, respectively, they would have had a meaningfully higher chance of receiving an organ, possibly averting death. The proportion of deaths was higher for up-categorized patients and lower for down-categorized patients compared to those whose scores did not change category. Supplementary Tables 4 and 5 stratify the reclassification analysis by sex. There were more women up-categorized ( $n=543$ , 16.7%) in general than down-categorized ( $n=23$ , 0.7%). Out of the 221 female decedents, a net of 33 (14.9%) would be correctly reclassified. In men, the effect was less dramatic, yet still positive with a net gain of 12 decedents (4.1%).

The temporal validation set included 10,459 listings from 2019, of which 3,588 (34.3%) had ALD as the primary listing diagnosis. The distribution of MELD 3.0 score was similar (median 19, IQR 13-26). The concordance statistic for 90-day mortality was 0.8682 using MELD 3.0 overall, compared to 0.8641 using MELD-Na ( $p=0.02$ ). When the analysis was repeated by etiology, the concordance statistic of MELD 3.0 was overall higher in ALD patients than those with other etiologies. Among ALD patients, concordance remained higher for MELD 3.0 than MELD-Na (0.8729 v. 0.8713), although its statistical significance was lost in this smaller subset ( $p=0.58$ , 194 deaths within 90 days), whereas the difference among patients with other etiologies of liver disease was maintained (0.8665 v. 0.8618,  $p=0.03$ ).

Finally, recognizing the potential concern of including albumin in an allocation model, we conducted a sensitivity analysis to construct a model without albumin. The resulting model incorporates all of the variables of MELD 3.0 except albumin and the interaction between albumin and creatinine. Table 3 compares the MELD 3.0, MELD 3.0 without albumin, and MELDNa. Compared to MELDNa (and the original MELD), the relative weight of serum bilirubin and creatinine increased in both MELD 3.0 and MELD 3.0 without albumin, whereas that of INR and sodium decreased. Model discrimination, judged by the concordance statistic, was the best for MELD 3.0 and worst for MELDNa, with MELD 3.0 without albumin being intermediate. The difference between concordance of MELD 3.0 and MELD 3.0 without albumin was significant ( $p<0.01$  in both development and validation sets), as was that between MELD 3.0 without albumin and MELDNa ( $p<0.01$  in development set and  $p=0.03$  in validation set).

In the LSAM analysis, only MELD 3.0 resulted in fewer waitlist deaths compared to MELDNa. Across replications, the mean number of deaths with MELDNa was 7,850, compared to 7,788 using MELD 3.0 with albumin ( $p=0.02$ ) and 7,814 using MELD 3.0 without albumin ( $p=0.12$ ).

## Discussion

In this work, we present the third iteration of the MELD score, hence MELD 3.0, following the original and MELDNa versions of the score. Compared to its predecessors, the current model is derived from a recent cohort of liver transplant candidates and characterized by the following new features: (1) addition of two variables, namely female sex and serum albumin, (2) lowered ceiling for serum creatinine from 4.0 mg/dL to 3.0 mg/dL, and (3) inclusion of two interaction terms between albumin and creatinine and between bilirubin and sodium. The score was rescaled in a way to maintain “MELD intuition” that practitioners have developed over time so that a given numerical score of both models represents similar level of sickness and mortality risk. The resulting model performed significantly better than MELDNa, the current gold standard, in ranking patients according to the risk of death. We estimate that the new score would reclassify approximately 9% of patients who died while waiting and reduce at least 20 waiting list deaths per year. Applying MELD 3.0 to a contemporary validation cohort confirmed that the model was robust to potential shifts in liver transplant etiology, e.g. the increasing incidence of ALD. Finally, for the potential concern about a model containing albumin, we performed a sensitivity analysis excluding albumin. As expected, the model without albumin performed not as well as the full model, but better than MELDNa.

Table 4 illustrates several scenarios to demonstrate the impact of different variables on the scores presented in this analysis. First, in the two low-risk patients with MELD of 12, the mild hyponatremia increases MELDNa by 6 points, whereas MELD 3.0 with and without albumin would give fewer additional points especially in a male patient. Of the intermediate-risk cases who share the same bilirubin, sodium, INR and creatinine, severe hypoalbuminemia increases MELD 3.0 by 1 point, as the predicted 90-day mortality increases by 2.4 percentage points. Female sex added another point to MELD 3.0 compared to a male patient with identical laboratory values, which was associated with another 3 percentage point increase in 90-day mortality. The high-risk cases demonstrate the impact of creatinine. With serum creatinine of 1.8 mg/dL and sodium of 128 mEq/L, MELDNa was 3 points higher than MELD, to which

MELD 3.0 added another point. An increase in creatinine from 1.8 mg/dL to 2.8 mg/dL added 3 more points to MELDNa and 4 to MELD 3.0, reflecting the steeper rise in mortality in this range of creatinine. This difference was larger with the model without albumin, which may be attributed to the fact that with MELD 3.0, albumin has little impact once creatinine is elevated. Similar to other scenarios, an identical female patient would receive one more MELD 3.0 point.

These cases illustrate the strengths of the new score. First, the current data point to the need to lower the ceiling for serum creatinine from the previous 4.0 mg/dL to 3.0 mg/dL. Serum creatinine is intended to represent renal function, which can be underestimated in patients with malnutrition and sarcopenia.<sup>15</sup> With MELDNa, in which serum creatinine is capped at 4.0 mg/dL, the maximum component score attributable to creatinine would be 13 points, whereas with MELD 3.0, the maximum creatinine of 3.0 mg/dL limits this to 12 points. Critics of MELD-based allocation have argued that the weight given to creatinine in MELDNa is excessive, creating an unfair advantage, including access to simultaneous liver kidney transplantation, to patients with high serum creatinine.<sup>16</sup> The lower impact of creatinine in MELD 3.0 is also relevant to the changing demographics of chronic liver disease, as abnormal creatinine in the increasing number of patients with NAFLD today with diabetic and/or hypertensive nephropathy may reflect chronic kidney disease rather than acute kidney injury that the creatinine term in the original MELD was purported to address.<sup>17</sup>

It has been consistently reported that women are significantly less likely to receive a transplant compared to men with the same MELD score, which may be related to a number of factors.<sup>4</sup> First, the predominant biological effect is that serum creatinine overestimates GFR and thus underestimates the risk of death in women compared to men with the same creatinine.<sup>18</sup> It was estimated that women receive 1 to 2.4 fewer creatinine-derived MELD points than men with similar renal dysfunction. Second, women tend to have a smaller abdominal cavity, which limits their ability to receive larger organs.<sup>19</sup> Third, certain conditions may affect men and women differently (e.g., hepatocellular carcinoma), which

may confound priorities in liver allocation. In our data, female sex was associated with a significantly higher risk of death and MELD 3.0 credits an extra 1.3 points to women, which will help mitigate the gender disparity in access to transplantation. It is also important to note that the score does not simply add extra points for women; it improves prediction for the population overall. Third, with regard to albumin, the model incorporates the interaction term between creatinine and albumin in such a way that as the creatinine increases, albumin becomes less important. In fact, when serum creatinine is 2.7 mg/dL or higher, hypoalbuminemia starts to lower the score, albeit by a small increment. Given the fact that MELD 3.0 was superior in discrimination, we propose that albumin is a meaningful variable to be included in the model.

These strengths notwithstanding, there may be concerns about the new score. First, the improvement between MELDNa and MELD 3.0 may appear small. The concordance statistics (by Harrell) were 0.862 and 0.869, respectively. However, this difference is statistically significant, and similar to that observed between the original MELD and MELDNa (0.868 versus 0.877).<sup>9</sup> Similarly, in our LSAM analysis the number of waitlist deaths would decrease by approximately 20 per year, which is approximately half of what was predicted for MELDNa compared to the original MELD. Thus, in our view, MELD 3.0 represents a meaningful improvement, especially when we consider that the new score adds dimensions that are biologically and clinically relevant and addresses the inherent gender disparity created by the use of MELD or MELDNa for liver allocation. Second, as was pointed out earlier, a potential concern may be raised that adding albumin may discourage clinicians from infusing albumin, when doing so, as recommended by guidelines, would be beneficial to the patient. In most of these circumstances, however, serum creatinine is likely to be elevated and would diminish, if not negate, the impact of albumin. Nonetheless, in case there is consensus that albumin should not be included, we provide a version without. Third, some predictors were not included in the final model. Differences in waitlist mortality based on race were observed, but the reasons why minorities, particularly black

patients, experience worse outcomes are often not genetic or biological, but rather due to external, largely socioeconomic factors rooted in structural racism. Thus, inclusion in a risk prediction score without fully understanding the underlying reasons for the racial disparity may have unintended consequences. Additionally, while the effect of sex was dominant, height also influences waitlist outcome and transplant probability. MELD 3.0 addresses individual urgency and the risk of waitlist mortality without transplant but does not account for potential size mismatch and access to size-appropriate organs among shorter men and women. The national allocation policy-making process by the OPTN is designed to address these issues. Finally, in this analysis, we did not set a maximum score of 40 for MELD 3.0 (and other scores). The cap was put in place when MELD was first implemented nearly two decades ago. The proportion of patients with high MELD scores awaiting transplantation has increased over time, and it has been observed that patients with MELD >40 experience greater waitlist mortality compared those with MELD 40, leading some to advocate removing the cap.<sup>20</sup> MELD 3.0 was scaled in such a way that the distribution of the score is similar to prior scores without a presumption that the score would be capped at 40.

A recent study by Godfrey et al. suggested that the predictive accuracy of the MELD score has declined over time, attributed to the changing demographics of liver disease, with a c-statistic of 0.80 in 2003 to 0.70 in 2015.<sup>21</sup> Applying a time-dependent c statistic, which appropriately accounts for censoring and was used in the development and validation of MELD and MELDNa, to the same dataset, we reported a c-statistic of 0.839 for MELDNa in 2015, which is consistent with the findings of the present study.<sup>22</sup> The demographics of liver disease have indeed changed since development of the original MELD score, and as shown in our study, recalibration using contemporary data and consideration of additional variables can further improve upon the prediction of waitlist mortality.

In conclusion, based on recent data consisting of liver transplant candidates in the US, we identify additional variables that are meaningfully associated with short term mortality including female sex and

serum albumin. We also found evidence to support lowering the serum creatinine ceiling to 3 mg/dL. Based on these data, we created an updated version of the MELD score, which improves mortality prediction compared to the current MELDNa model, including the recognition of female sex as a risk factor for death. We believe that the new model represents an opportunity to lower waitlist mortality in the US and propose it to be considered to replace the current version of MELD in determining allocation priorities in liver transplantation.

Journal Pre-proof

## References

1. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology (Baltimore, Md)*. 2001;33(2):464-470.
2. Nagai S, Chau LC, Schilke RE, et al. Effects of Allocating Livers for Transplantation Based on Model for End-Stage Liver Disease-Sodium Scores on Patient Outcomes. *Gastroenterology*. 2018;155(5):1451-1462.e1453.
3. Asrani SK, Jennings LW, Kim WR, et al. MELD-GRAIL-Na: Glomerular Filtration Rate and Mortality on Liver-Transplant Waiting List. *Hepatology (Baltimore, Md)*. 2020;71(5):1766-1774.
4. Locke JE, Shelton BA, Olthoff KM, et al. Quantifying Sex-Based Disparities in Liver Allocation. *JAMA Surgery*. 2020;155(7):e201129-e201129.
5. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91-96.
6. Lai JC, Covinsky KE, Dodge JL, et al. Development of a novel frailty index to predict mortality in patients with end-stage liver disease. *Hepatology (Baltimore, Md)*. 2017;66(2):564-574.
7. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020 Aug 27;383(9):874-82.
8. Eneanya ND, Yang W, Reese PP. Reconsidering the Consequences of Using Race to Estimate Kidney Function. *JAMA*. 2019 Jul 9;322(2):113-4.
9. Peng Y, Khaled T, Liu J, et al. Data Request from the MELD Enhancement Subcommittee of the Liver and Intestinal Organ Transplantation Committee. Data Request ID: HR2011\_02. June 7, 2011.
10. Hastie T, Tibshirani R. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*. 1990;46(4):1005-1016.

11. Leise MD, Kim WR, Kremers WK, Larson JJ, Benson JT, Therneau TM. A revised model for end-stage liver disease optimizes prediction of mortality among patients awaiting liver transplantation. *Gastroenterology*. 2011;140(7):1952-1960.
12. Harrell FE, Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the Yield of Medical Tests. *JAMA*. 1982;247(18):2543-2546.
13. Uno H, Cai T, Tian L, Wei LJ. Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models. *Journal of the American Statistical Association*. 2007;102(478):527-537.
14. Bajaj JS, Tandon P, O'Leary JG, et al. The Impact of Albumin Use on Resolution of Hyponatremia in Hospitalized Patients With Cirrhosis. *The American journal of gastroenterology*. 2018;113(9):1339.
15. Asrani SK, Jennings LW, Trotter JF, et al. A Model for Glomerular Filtration Rate Assessment in Liver Disease (GRAIL) in the Presence of Renal Dysfunction: Hepatology. *Hepatology (Baltimore, Md)*. 2019;69(3):1219-1230.
16. Merola J, Formica RN, Mulligan DC. Changes in united network for organ sharing policy for simultaneous liver-kidney allocation. *Clinical Liver Disease*. 2017;9(1):21-24.
17. Musso G, Gambino R, Tabibian JH, et al. Association of Non-alcoholic Fatty Liver Disease with Chronic Kidney Disease: A Systematic Review and Meta-analysis. *PLOS Medicine*. 2014;11(7):e1001680.
18. Allen AM, Heimbach JK, Larson JJ, et al. Reduced Access to Liver Transplantation in Women: Role of Height, MELD Exception Scores, and Renal Function Underestimation. *Transplantation*. 2018;102(10):1710-1716.
19. Lai JC, Terrault NA, Vittinghoff E, Biggins SW. Height contributes to the gender difference in wait-list mortality under the MELD-based liver allocation system. *Am J Transplant*. 2010;10(12):2658-2664.

20. Nadim MK, DiNorcia J, Ji L, et al. Inequity in organ allocation for patients awaiting liver transplantation: Rationale for uncapping the model for end-stage liver disease. *Journal of hepatology*. 2017;67(3):517-525.
21. Godfrey EL, Malik TH, Lai JC, Mindikoglu AL, Galván NTN, Cotton RT, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. *Am J Transplant*. 2019 Sep 4;ajt.15559.
22. Kwong AJ, Mannalithara A, Kim WR. Reply to: "The decreasing predictive power of MELD in an era of changing etiology of liver disease." *Am J Transplant*. 2020 Mar;20(3):901–2.

**Figure Legend**

**Figure 1.** Multivariable smoothing splines relating predictor variables with relative risk of death within 90 days. (A) Bilirubin (B) INR (C) Creatinine (D) Sodium (E) Albumin.

Journal Pre-proof

Figure 1. Multivariable smoothing splines relating predictor variables with relative risk of death in 90 days.

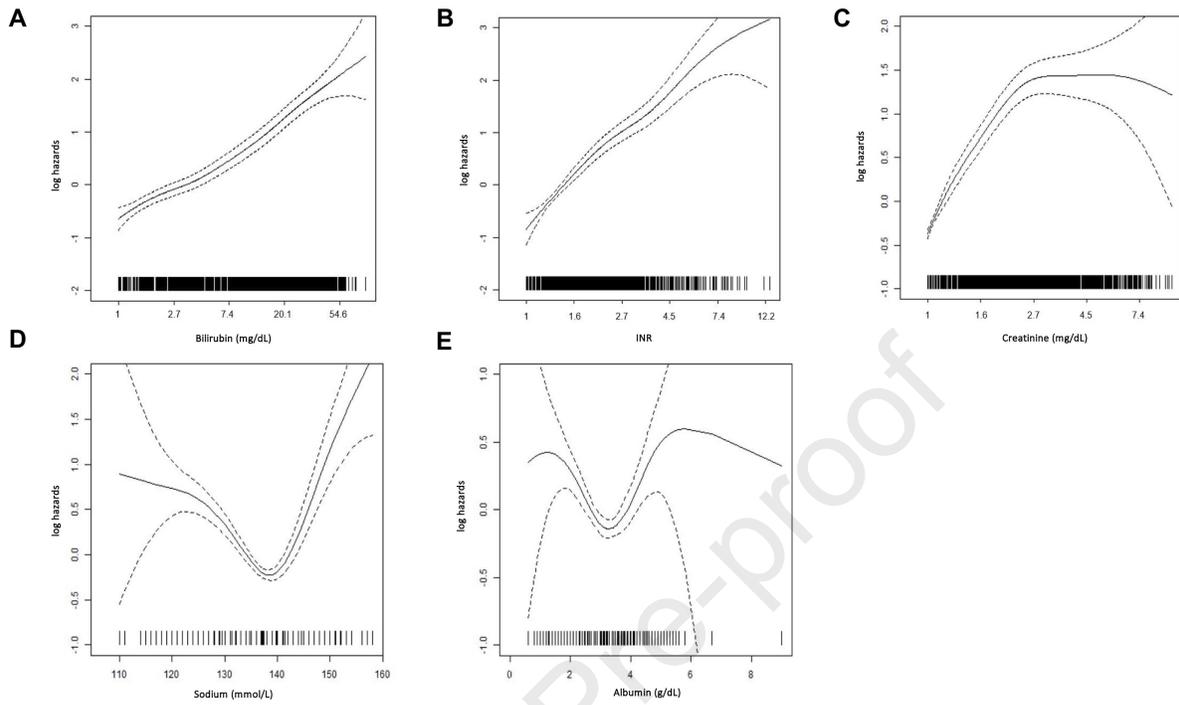


Table 1

Baseline characteristics of liver transplant waitlist registrants

|                       | Overall<br>(n=29,410) | Development Set<br>(n=20,587) | Validation Set<br>(n=8,823) |
|-----------------------|-----------------------|-------------------------------|-----------------------------|
| Age (yr)              | 58.0 (51.0-64.0)      | 58.0 (51.0-64.0)              | 58.0 (51.0-64.0)            |
| Women, n (%)          | 10,835 (36.8)         | 7,592 (36.9)                  | 3,243 (36.8)                |
| Race, n (%)           |                       |                               |                             |
| White                 | 20,661 (70.3)         | 14,484 (70.4)                 | 6,177 (70.0)                |
| Hispanic              | 4,835 (16.4)          | 3,424 (16.6)                  | 1,411 (16.0)                |
| Black                 | 2,185 (7.4)           | 1,490 (7.2)                   | 695 (7.9)                   |
| Asian                 | 1,214 (4.1)           | 832 (4.0)                     | 382 (4.3)                   |
| Other                 | 515 (1.8)             | 357 (1.7)                     | 158 (1.8)                   |
| Diabetes, n (%)       | 8,863 (30.2)          | 6,252 (30.5)                  | 2,611 (29.7)                |
| Ascites, n (%)        |                       |                               |                             |
| Absent                | 7,870 (26.8)          | 5,537 (26.9)                  | 2,333 (26.4)                |
| Slight                | 13,502 (45.9)         | 9,450 (45.9)                  | 4,052 (45.9)                |
| Moderate              | 8,038 (27.3)          | 5,600 (27.2)                  | 2,438 (27.6)                |
| Encephalopathy, n (%) |                       |                               |                             |
| None                  | 11,843 (40.3)         | 8,328 (40.5)                  | 3,515 (39.8)                |
| 1-2                   | 15,368 (52.3)         | 10,739 (52.2)                 | 4,629 (52.5)                |
| 3-4                   | 2,199 (7.5)           | 1,520 (7.4)                   | 679 (7.7)                   |
| Sodium                | 137.0 (133.0-139.0)   | 137.0 (133.0-139.0)           | 137.0 (133.0-139.0)         |
| Creatinine            | 1.0 (0.8-1.5)         | 1.0 (0.8-1.4)                 | 1.0 (0.8-1.5)               |
| INR                   | 1.4 (1.2-1.8)         | 1.4 (1.2-1.8)                 | 1.4 (1.2-1.8)               |
| Bilirubin             | 2.5 (1.2-5.7)         | 2.5 (1.3-5.7)                 | 2.5 (1.2-5.7)               |
| Albumin               | 3.2 (2.7-3.6)         | 3.2 (2.7-3.6)                 | 3.2 (2.7-3.6)               |
| MELD                  | 16.0 (11.0-23.0)      | 16.0 (11.0-23.0)              | 16.0 (12.0-23.0)            |
| MELDNa                | 18.0 (11.0-25.0)      | 18.0 (11.0-25.0)              | 18.0 (12.0-25.0)            |
| CTP Score             | 9.0 (7.0-11.0)        | 9.0 (7.0-11.0)                | 9.0 (7.0-11.0)              |

Table 2

Reclassification of liver transplant candidates between MELDNa and MELD3.0 in the validation set. (A) the number of patients, (B) the number of deaths and (C) the proportion of death (B divided by A). Red-demarcated areas indicate up-scoring (MELD3.0 category higher than MELDNa) and blue-demarcated areas the opposite.

| A. Patients (n) |       | MELD 3.0 |       |       |       |       |
|-----------------|-------|----------|-------|-------|-------|-------|
|                 |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+   |
| MELDNa          | 6-9   | 1047     | 334   | -     | -     | -     |
|                 | 10-19 | 66       | 3093  | 341   | -     | -     |
|                 | 20-29 | -        | 150   | 2182  | 140   | -     |
|                 | 30-39 | -        | -     | 64    | 1007  | 75    |
|                 | 40+   | -        | -     | -     | 26    | 298   |
| B. Deaths (n)   |       | MELD 3.0 |       |       |       |       |
|                 |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+   |
| MELDNa          | 6-9   | 6        | 4     | -     | -     | -     |
|                 | 10-19 | -        | 45    | 16    | -     | -     |
|                 | 20-29 | -        | 6     | 166   | 23    | -     |
|                 | 30-39 | -        | -     | 8     | 141   | 19    |
|                 | 40+   | -        | -     | -     | 3     | 77    |
| C. Death (%)    |       | MELD 3.0 |       |       |       |       |
|                 |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+   |
| MELDNa          | 6-9   | 0.6%     | 1.2%  | -     | -     | -     |
|                 | 10-19 | -        | 1.5%  | 4.7%  | -     | -     |
|                 | 20-29 | -        | 4.0%  | 7.6%  | 16.4% | -     |
|                 | 30-39 | -        | -     | 12.5% | 14.0% | 25.3% |
|                 | 40+   | -        | -     | -     | 11.5% | 25.8% |

Table 3

Comparison of coefficients and concordance for MELD 3.0, MELDNa and MELD 3.0 with no albumin. The concordance data are from the validation set.

|  |                       | MELD 3.0 | MELDNa | MELD 3.0 No Albumin |
|--|-----------------------|----------|--------|---------------------|
| Coefficient  |                       |          |        |                     |
|  | $\log_e$ (Bilirubin)  | 4.56     | 3.78   | 4.85                |
|  | $\log_e$ (INR)        | 9.09     | 11.20  | 9.66                |
|  | $\log_e$ (Creatinine) | 11.14    | 9.57   | 10.47               |
|  | Na                    | 0.82     | 1.32   | 0.88                |
|  | Albumin               | 1.85     | NA     | NA                  |
|  | Female                | 1.33     | NA     | 1.4                 |
| Concordance  |                       |          |        |                     |
|  | by Harrell            | 0.8693   | 0.8622 | 0.8665              |
|  | by Uno                | 0.8378   | 0.8294 | 0.8342              |
| MELD 3.0 No Albumin = 1.40 (if female) + 4.85 * $\log_e$ (bilirubin) + 0.88 * (137-Na) – 0.25 (137-Na) * $\log_e$ (bilirubin) + 9.66 * $\log_e$ (INR) + 10.47 * $\log_e$ (creatinine) + 6. |                       |          |        |                     |

Table 4. Illustrative low-, intermediate- and high-risk cases with MELD, MELDNa and MELD 3.0 scores

|                     | Risk Level |      |              |            |            |            |            |            |
|---------------------|------------|------|--------------|------------|------------|------------|------------|------------|
|                     | Low        |      | Intermediate |            |            | High       |            |            |
| Data                |            |      |              |            |            |            |            |            |
| Sex                 | M          | F    | M            | M          | F          | M          | M          | F          |
| Bilirubin           | 2.5        | 2.5  | 6.0          | 6.0        | 6.0        | 12.0       | 12.0       | 12.0       |
| Na                  | 131        | 131  | 131          | 131        | 131        | 128        | 128        | 128        |
| INR                 | 1.0        | 1.0  | 1.5          | 1.5        | 1.5        | 2.2        | 2.2        | 2.2        |
| Creatinine          | 1.2        | 1.2  | 1.5          | 1.5        | 1.5        | <b>1.8</b> | <b>2.8</b> | <b>2.8</b> |
| Albumin             | 3.8        | 3.8  | <b>3.5</b>   | <b>2.2</b> | <b>2.2</b> | 2.0        | 2.0        | 2.0        |
| Scores              |            |      |              |            |            |            |            |            |
| MELD                | 12         | 12   | 22           | 22         | 22         | 30         | 35         | 35         |
| MELDNa              | 18         | 18   | 26           | 26         | 26         | 33         | 36         | 36         |
| MELD 3.0            | 16         | 17   | 25           | 26         | 27         | 34         | 38         | 39         |
| Delta*              | -2         | -1   | -1           | 0          | 1          | 1          | 2          | 3          |
| MELD 3.0 no albumin | 16         | 18   | 25           | 25         | 27         | 34         | 39         | 40         |
| Predicted Mortality |            |      |              |            |            |            |            |            |
| 30 day              | 0.9%       | 1.1% | 4.3%         | 5.3%       | 6.4%       | 19.8%      | 35.8%      | 41.8%      |
| 90 day              | 2.6%       | 3.1% | 12.1%        | 14.5%      | 17.5%      | 47.1%      | 72.3%      | 79.1%      |

\* Difference between MELDNa and MELD 3.0 (Delta score = MELD 3.0 - MELDNa)

## Supplementary Materials

**Supplementary Table 1.** Comparison between MELD 3.0 (prior to rescaling) and a model with separate terms for height in men and women, adjusted for all other MELD 3.0 components.

**Supplementary Table 2.** Univariate cox proportional hazards regression for 90-day mortality since waitlist registration (Development data set)

**Supplementary Table 3.** The final multivariable Cox proportional hazards regression for 90-day survival. (A) Variables and coefficients; (B) Survival function for MELD 3.0; (C) Examples

**Supplementary Table 4.** Reclassification of female liver transplant candidates between MELDNa and MELD3.0 in the validation set. (A) the number of patients, (B) the number of deaths and (C) the proportion of death (B divided by A). Red-demarcated areas indicate up-scoring (MELD3.0 category higher than MELDNa) and blue-demarcated areas the opposite.

**Supplementary Table 5.** Reclassification of male liver transplant candidates between MELDNa and MELD3.0 in the validation set. (A) the number of patients, (B) the number of deaths and (C) the proportion of death (B divided by A). Red-demarcated areas indicate up-scoring (MELD3.0 category higher than MELDNa) and blue-demarcated areas the opposite.

**Supplementary Figure 1.** Multivariable smoothing spline describing relative hazard of 90-day mortality based on height, stratified by sex.

**Supplementary Figure 2.** Study subjects.

**Supplementary Table 1.** Comparison between MELD 3.0 (prior to rescaling) and a model with separate terms for height in men and women, adjusted for all other MELD 3.0 components.

| Model without height (MELD 3.0) |                    |         | Model incorporating height        |                    |         |
|---------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| Parameter                       | Parameter Estimate | P_Value | Parameter                         | Parameter Estimate | P_Value |
| Female sex                      | <b>0.236</b>       | <0.001  | pspline (height_women), linear    | -0.01593           | 0.009   |
|                                 |                    |         | pspline (height_women), nonlinear |                    | 0.127   |
|                                 |                    |         | pspline (height_men), linear      | -0.0012            | 0.789   |
|                                 |                    |         | pspline (height_men), nonlinear   |                    | 0.261   |
| ln(bilirubin)                   | 0.809              | <0.001  | ln(bilirubin)                     | 0.812              | <0.001  |
| 137-Na                          | 0.145              | <0.001  | 137-Na                            | 0.145              | <0.001  |
| ln(bilirubin)*(137-Na)          | -0.042             | <0.001  | ln(bilirubin)*(137-Na)            | -0.042             | <0.001  |
| ln(INR)                         | 1.612              | <0.001  | ln(INR)                           | 1.608              | <0.001  |
| ln(creatinine)                  | 1.976              | <0.001  | ln(creatinine)                    | 1.962              | <0.001  |
| albumin                         | 0.328              | <0.001  | albumin                           | 0.318              | <0.001  |
| ln(creatinine)*albumin          | -0.324             | 0.010   | ln(creatinine)*albumin            | -0.298             | 0.017   |

**Supplementary Table 2.** Univariate cox proportional hazards regression for 90-day survival since waitlist registration (Development data set).

| <b>Variable</b> | <b>Hazard Ratio (95% CI)</b> | <b>P value</b> |
|-----------------|------------------------------|----------------|
| Women           | 1.370 (1.219-1.541)          | <0.001         |
| MELD            | 1.172 (1.165-1.179)          | <0.001         |
| Sodium          | 0.940 (0.929-0.951)          | <0.001         |
| MELDNa          | 1.186 (1.178-1.195)          | <0.001         |
| Albumin         | 0.839 (0.766-0.920)          | <0.001         |
| Creatinine      | 1.130 (1.112-1.148)          | <0.001         |
| INR             | 1.768 (1.723-1.813)          | <0.001         |
| Bilirubin       | 1.076 (1.073-1.080)          | <0.001         |

**Supplementary Table 3.** The final multivariable Cox proportional hazards regression for 90-day survival.

(A) Variables and coefficients

| Parameter                                   | Parameter estimate | Standard error | p      |
|---|--------------------|----------------|--------|
| Female                                      | 0.236              | 0.0598         | <0.001 |
| Log <sub>e</sub> Bilirubin                  | 0.809              | 0.040          | <0.001 |
| (137-Na)                                    | 0.145              | 0.0163         | <0.001 |
| Log <sub>e</sub> INR                        | 1.612              | 0.0924         | <0.001 |
| Log <sub>e</sub> Creatinine                 | 1.976              | 0.0982         | <0.001 |
| (3.5-Albumin)                               | 0.328              | 0.0839         | <0.001 |
| Log <sub>e</sub> Bilirubin * (137-Na)       | -0.042             | 0.0074         | <0.001 |
| Log <sub>e</sub> Creatinine * (3.5-Albumin) | -0.324             | 0.1249         | 0.010  |

(B) Survival function for MELD 3.0

| t (days) | 15    | 30    | 45    | 60    | 75    | 90    |
|----------|-------|-------|-------|-------|-------|-------|
| So (t)   | 0.991 | 0.981 | 0.971 | 0.963 | 0.955 | 0.946 |

$$S_i(t) = S_0(t) \wedge \exp(0.17698 * \text{MELD}3.0 - 3.56)$$

(C) Examples

| Lab values                   | Sex | MELD 3.0 | Predicted survival |       |       |       |       |       |
|------------------------------|-----|----------|--------------------|-------|-------|-------|-------|-------|
|                              |     |          | 15                 | 30    | 45    | 60    | 75    | 90    |
| 50 <sup>th</sup> percentile† | M   | 14       | 0.997              | 0.993 | 0.990 | 0.987 | 0.984 | 0.981 |
|                              | F   | 15       | 0.996              | 0.992 | 0.988 | 0.984 | 0.981 | 0.977 |
| 75 <sup>th</sup> percentile‡ | M   | 26       | 0.975              | 0.947 | 0.920 | 0.899 | 0.878 | 0.855 |
|                              | F   | 28       | 0.963              | 0.922 | 0.883 | 0.853 | 0.823 | 0.791 |

†Laboratory values at the 50<sup>th</sup> percentile: Bilirubin 2.5 mg/dl, Na 137 mEq/L, INR 1.4, Creatinine 1.0 mg/dL, Albumin 3.2 g/dL

‡Laboratory values at the 75<sup>th</sup> percentile: Bilirubin 5.7 mg/dL, Na 133 mEq/L, INR 1.8, Creatinine 1.5 mg/dL, Albumin 2.7 g/dL

**Supplementary Table 4.** Reclassification of female liver transplant candidates between MELD-Na and MELD3.0 in the validation set. (A) the number of patients, (B) the number of deaths and (C) the proportion of death (B divided by A). Red-demarkated areas indicate up-scoring (MELD 3.0 category higher than MELDNa) and blue-demarkated areas the opposite.

| A<br>Patients (n) |       | MELD 3.0 |       |       |       |     |
|-------------------|-------|----------|-------|-------|-------|-----|
|                   |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+ |
| MELDNa            | 6-9   | 274      | 187   | -     | -     | -   |
|                   | 10-19 | -        | 1034  | 213   | -     | -   |
|                   | 20-29 | -        | 11    | 838   | 98    | -   |
|                   | 30-39 | -        | -     | 7     | 417   | 45  |
|                   | 40+   | -        | -     | -     | 5     | 114 |

| B<br>Deaths (n) |       | MELD 3.0 |       |       |       |     |
|-----------------|-------|----------|-------|-------|-------|-----|
|                 |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+ |
| MELDNa          | 6-9   | -        | 1     | -     | -     | -   |
|                 | 10-19 | -        | 17    | 10    | -     | -   |
|                 | 20-29 | -        | 1     | 71    | 19    | -   |
|                 | 30-39 | -        | -     | 1     | 65    | 6   |
|                 | 40+   | -        | -     | -     | 1     | 29  |

| C<br>Death (%) |       | MELD 3.0 |       |       |       |       |
|----------------|-------|----------|-------|-------|-------|-------|
|                |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+   |
| MELDNa         | 6-9   | -        | 0.5%  | -     | -     | -     |
|                | 10-19 | -        | 1.6%  | 4.7%  | -     | -     |
|                | 20-29 | -        | 9.1%  | 8.5%  | 19.4% | -     |
|                | 30-39 | -        | -     | 14.3% | 15.6% | 13.3% |
|                | 40+   | -        | -     | -     | 20.0% | 25.4% |

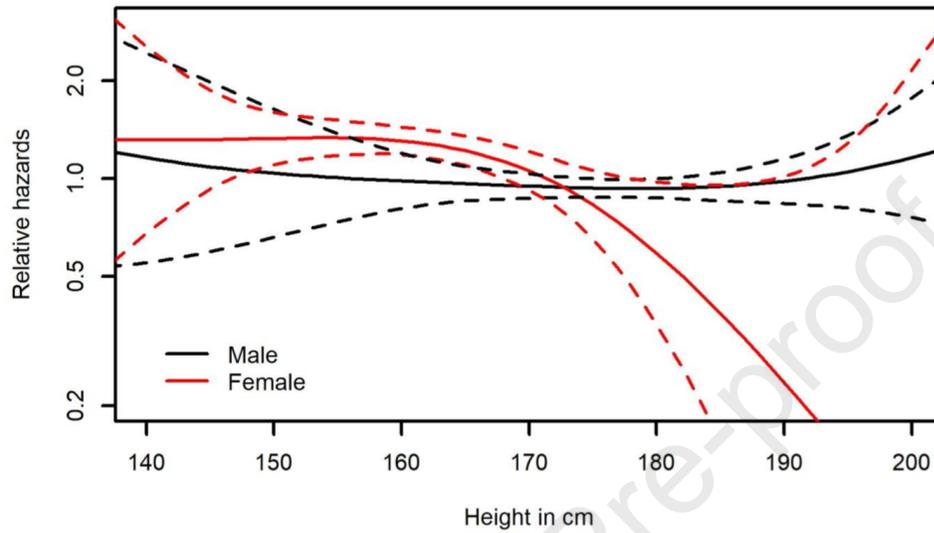
**Supplementary Table 5.** Reclassification of male liver transplant candidates between MELDNa and MELD3.0 in the validation set. (A) the number of patients, (B) the number of deaths and (C) the proportion of death (B divided by A). Red-demarkated areas indicate up-scoring (MELD3.0 category higher than MELDNa) and blue-demarkated areas the opposite.

| A<br>Patients (n) |       | MELD 3.0 |       |       |       |     |
|-------------------|-------|----------|-------|-------|-------|-----|
|                   |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+ |
| MELDNa            | 6-9   | 773      | 147   | -     | -     | -   |
|                   | 10-19 | 66       | 2059  | 128   | -     | -   |
|                   | 20-29 | -        | 139   | 1344  | 42    | -   |
|                   | 30-39 | -        | -     | 57    | 590   | 30  |
|                   | 40+   | -        | -     | -     | 21    | 184 |

| B<br>Deaths (n) |       | MELD 3.0 |       |       |       |     |
|-----------------|-------|----------|-------|-------|-------|-----|
|                 |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+ |
| MELDNa          | 6-9   | 6        | 3     | -     | -     | -   |
|                 | 10-19 | -        | 28    | 6     | -     | -   |
|                 | 20-29 | -        | 5     | 95    | 4     | -   |
|                 | 30-39 | -        | -     | 7     | 76    | 13  |
|                 | 40+   | -        | -     | -     | 2     | 48  |

| C<br>Death (%) |       | MELD 3.0 |       |       |       |       |
|----------------|-------|----------|-------|-------|-------|-------|
|                |       | 6-9      | 10-19 | 20-29 | 30-39 | 40+   |
| MELDNa         | 6-9   | 0.8%     | 2.0%  | -     | -     | -     |
|                | 10-19 | -        | 1.4%  | 4.7%  | -     | -     |
|                | 20-29 | -        | 3.6%  | 7.1%  | 9.5%  | -     |
|                | 30-39 | -        | -     | 12.3% | 12.9% | 43.3% |
|                | 40+   | -        | -     | -     | 9.5%  | 26.1% |

**Supplementary Figure 1.** Multivariable smoothing spline describing relative hazard of 90-day mortality based on height, stratified by sex.



Supplementary Figure 2. Study subjects.

