



**Charting the Path Forward for Risk Prediction in Liver  
Transplant for HCC: International Validation of HALTHCC  
amongst 4,089 patients**

Journal:	<i>Hepatology</i>
Manuscript ID	HEP-18-2422.R2
Wiley - Manuscript type:	Original
Date Submitted by the Author:	11-Jun-2019
Complete List of Authors:	<p>Firl, Daniel; Cleveland Clinic, Lerner College of Medicine  Sasaki, Kazunari; Cleveland Clinic, Lerner College of Medicine  Agopian, Vatche; The Dumont-UCLA Transplant Center, ; Pfleger Liver  Institute, Department of Surgery  Gorgen, Andre; University of Toronto Division of General Surgery, Multi-  Organ Transplant Program  Kimura, Shoko; Massachusetts General Hospital Institute for Patient Care  Dumronggittigule, Wethit; University of California Los Angeles David  Geffen School of Medicine  Mcvey, Jack; Cleveland Clinic, Lerner College of Medicine  Iesari, Samuele; Universite catholique de Louvain  Mennini, Gainluca; Universita degli Studi di Roma La Sapienza  Vitale, Alessandro; Padua University Hospital, Chirurgia Epatobiliare e  Trapianto Epatico  Finkenstedt, Armin; Medical University of Innsbruck, Department of  Medicine, Division of Gastroenterology and Hepatology  Onali, Simona; Royal Free Hospital, Sheila Sherlock Liver Centre  Hoppe-Lotichius, Maria; Johannes Gutenberg University, Department of  Transplantation and Hepatobiliopancreatic Surgery  Vennarecci, Giovanni; San Camillo Hospital, transplantation  Manzia, Tommaso; Tor Vergata University, U.O.C.Trapianti d'Organo  Nicolini, Daniele ; Universita Politecnica delle Marche  Avolio, Alfonso; Catholic University, Dept of Surgery  Agnes, Salvatore; "A.Gemelli" Catholic University of Rome, Dept of  Surgery - Transplantation Service  Vivarelli, Marco; University of Bologna, Surgery and Transplantation  Tisone, G; S. Eugenio Hospital Roma Italy, Surgical clinic  Ettorre, Giuseppe; San Camillo Hospital, General Surgery and  Transplantation Unit  Otto, Gerd; University Medical Center,  Tsochatzis, Emmanuel; Royal Free Hospital, The Royal Free Sheila  Sherlock Liver Centre and UCL Institute of Liver and Digestive Health ;  Rossi, Massimo  Viveiros, Andre; Medical University Innsbruck, Internal Medicine, II  Department  Cillo, Umberto; University of Padova, Surgery and Oncology  Markmann, James; Massachusetts General Hospital, Surgery;  Ikegami, Toru; Kyushu University, Department of Surgery and Science  Kaido, Toshimi; Kyoto University School of Medicine, Hepato-Biliary-  Pancreatic and Transplant Surgery</p>

	LAI, Quirino Sapisochin, Gonzalo; Toronto General Hospital. University Health Network, Multi-Organ Transplant. Department of Surgery Lerut, Jan; Cliniques Universitaires Saint-Luc, ; Aucejo, Federico; Cleveland Clinic, Liver Transplantation;
Keywords:	hepatocellular carcinoma, risk models, international study, liver transplantation



## **Charting the Path Forward for Risk Prediction in Liver Transplant for HCC: International Validation of HALTHCC amongst 4,089 patients**

Daniel J. Firl<sup>1,§</sup>, Kazunari Sasaki<sup>1,§</sup>, Vatche G. Agopian<sup>2</sup>, Andre Gorgen<sup>3</sup>, Shoko Kimura<sup>4</sup>, Wethit Dumronggittigule<sup>2</sup>, John C. McVey<sup>1</sup>, Samuele Iesari<sup>5</sup>, Gianluca Mennini<sup>6</sup>, Alessandro Vitale<sup>7</sup>, Armin Finkenstedt<sup>8</sup>, Simona Onali<sup>9</sup>, Maria Hoppe-Lotichius<sup>10</sup>, Giovanni Vennarecci<sup>11</sup>, Tommaso M. Manzia<sup>12</sup>, Daniele Nicolini<sup>13</sup>, Alfonso W. Avolio<sup>14</sup>, Salvatore Agnes<sup>14</sup>, Marco Vivarelli<sup>13</sup>, Giuseppe Tisone<sup>12</sup>, Giuseppe M. Ettorre<sup>11</sup>, Gerd Otto<sup>10</sup>, Emmanuel Tsochatzis<sup>9</sup>, Massimo Rossi<sup>6</sup>, Andre Viveiros<sup>8</sup>, Umberto Cillo<sup>7</sup>, James F Markmann<sup>4</sup>, Toru Ikegami<sup>15</sup>, Toshimi Kaido<sup>16</sup>, Quirino Lai<sup>5,6</sup>, Gonzalo Sapisochin<sup>3</sup>, Jan Lerut<sup>5</sup> on behalf of the European Hepatocellular Cancer Liver Transplant Study Group, and Federico N. Aucejo<sup>1</sup>.

**Affiliations:** <sup>1</sup>Department of General Surgery and Cleveland Clinic Lerner College of Medicine, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>§</sup>Current Address: danfirl@gmail.com; Department of Surgery, Duke University Hospital, Durham, NC, USA. <sup>2</sup>Dumont-UCLA Transplant and Liver Cancer Center, Department of Surgery, Ronald Reagan UCLA Medical Center and David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>3</sup>Department of Abdominal Transplant and HPB Surgical Oncology, University Health Network and University of Toronto, Toronto, Canada. <sup>4</sup>Transplant Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>5</sup>Starzl Unit of Abdominal Transplantation, St. Luc University Hospital, Université Catholique Louvain, Brussels, Belgium. <sup>6</sup>Department of General Surgery and Organ Transplantation, Umberto I Hospital, Sapienza University, Rome, Italy. <sup>7</sup>Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy. <sup>8</sup>Department of Medicine I, Medical University Innsbruck, Innsbruck, Austria. <sup>9</sup>UCL Institute for Liver and Digestive Health and Royal Free Sherlock Liver Centre, Royal Free Hospital and UCL, London, UK. <sup>10</sup>Department of Transplantation and Hepatobiliary Surgery, University of Mainz, Mainz, Germany. <sup>11</sup>Division of General Surgery and Liver Transplantation, San Camillo Hospital, Rome, Italy. <sup>12</sup>Department of Transplant Surgery, Polyclinic Tor Vergata Foundation, Tor Vergata University, Rome, Italy. <sup>13</sup>Unit of Hepatobiliary Surgery and Transplantation, Azienda Ospedaliero-Universitaria Ospedali Riuniti, Torrette Ancona, Italy. <sup>14</sup>Liver Unit, Department of Surgery, Agostino Gemelli Hospital, Catholic University, Rome, Italy. <sup>15</sup>Kyoto University Hospital, Kyoto, Japan. <sup>16</sup>Kyushu University Hospital, Kyushu, Japan.

**§Corresponding Author:** Kazunari Sasaki MD, Department of General Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, OH, USA. E: sasakik@ccf.org

**Funding Source:** KS received an American Society of Transplant Surgeons Fellowship related to this work. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

**Disclosures:** The authors declare no conflicts or competing interests as described by *Hepatology* guidelines.

**Keywords:** liver transplantation, risk models, selection criteria, international study

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Word Count:** 3000, not including title page, abstract, references, table/figure legends

**Abstract (270/275words max)**

Prognosticating outcomes in liver transplant (LT) for hepatocellular carcinoma (HCC) continues to challenge the field. While Milan Criteria (MC) generalized the practice of LT for HCC and improved outcomes, its predictive character has degraded with increasing candidate and oncological heterogeneity. We sought to validate and recalibrate a previously developed, preoperatively calculated, continuous risk score, the hazard associated with liver transplantation in HCC (HALTHCC) in an international cohort. From 2002-2014, 4,089 patients (both MC in and out [25.2%]) across 16 centers in North America, Europe, and Asia were included. A continuous risk score using pre-LT levels of alpha fetoprotein, model for end stage liver disease sodium score, and tumor burden score was recalibrated amongst a randomly selected cohort (n=1,021) and validated in the remainder (n=3,068). This study demonstrated significant heterogeneity by site and year, reflecting practice trends over the last decade. On explant pathology, both vascular invasion(VI) and poorly differentiated component(PDC) increased with increasing HALTHCC score. The lowest risk patients (HALTHCC 0-5) had lower rates of VI and PDC than the highest risk patients (HALTHCC>35) (VI:7.7%[1.2-14.2] vs 70.6%[48.3-92.9] and PDC:4.6%[0.1-9.8%] vs 47.1%[22.6-71.5];  $P<0.0001$  for both). This trend was robust to MC status. This international study was used to adjust the coefficients in the HALTHCC score. Before recalibration, HALTHCC had the greatest discriminatory ability for overall survival (C-index=0.61) compared to all previously reported scores. Following recalibration, the prognostic utility increased for both recurrence (C-index=0.71) and overall survival (C-index=0.63).

**Conclusion:** This large international trial validated and refined the role for the continuous risk metric, HALTHCC, in establishing pre-LT risk amongst candidates with HCC worldwide. Prospective trials introducing HALTHCC into clinical practice are warranted.

## Introduction (458)

Hepatocellular carcinoma (HCC) is the fifth most common cancer by global incidence.<sup>1,2</sup> For most patients, liver transplantation (LT) offers the only reasonable chance at cure since it replaces the carcinogenic background liver and leaves the diseased liver capsule and vasculature undisturbed.<sup>3,4</sup> While LT is preferred, limited donor organs and poorer outcomes from cancer recurrence limited its application until the groundbreaking Milan Criteria (MC: a single-tumor $\leq$ 5cm or up to 3 tumors, all $<$ 3cm).<sup>3-5</sup> In the subsequent two decades, MC has allowed transplant centers around the world to treat HCC patients with similar, acceptable outcomes.<sup>2,6-8</sup> However, several groups have suggested that MC is too stringent.<sup>8-10</sup> With increasing penetration of down-staging loco-regional therapy (LRT), the predictive quality of MC has decreased.<sup>9,11</sup> This can be attributed to tumor morphology alone not reflecting the biology of HCC, however, it is also worth noting LRT was uncommon when MC was developed.<sup>12</sup>

Many factors have been suggested to improve patient selection. Unfortunately many models incorporated genetic parameters/serum assays not routinely collected in clinical practice<sup>6,7,13-15</sup> or pathologic factors which cannot be used for selection.<sup>8,16,17</sup> Most studies presented dichotomous criteria; the patient is in or out. Although simple and convenient, modelling risk in binary terms discards variance, which could be associated with outcomes and results in poorer discrimination.<sup>18</sup> Such models need to be reconstructed *de novo* with population or clinical practice shifts. The hazard associated with LT for HCC (HALTHCC) model sought to remedy these strategic missteps by forming a continuous spectrum of risk upon which a patient can be judged relative to others.<sup>9</sup> Its value is calculated by common clinical variables which were determined in a single center to be relevant for post-LT survival and validated amongst the entire national experience of the United States.<sup>9</sup> It was also validated for post-LT recurrence and its longitudinal measurement was associated with dropout in a multicenter study.<sup>11</sup>

In setting potential LT candidates on a spectrum of risk, population characteristics can continuously change without the model being so stiff as to stop being useful. Further, maintaining the equation as a set of beta coefficients is a flexible framework familiar to the transplant community in the form of the model for end stage liver disease (MELD) equation and allows continual improvement in accuracy without reinvention of the concept. While HALTHCC is a valuable lens through which to observe candidate risk, its utility amongst international centers is unknown. The aim of this study was three-fold: to create a broad and heterogeneous cohort representative of world-wide LT for HCC; to test the hypothesis that HALTHCC will maintain its predictive character for post-LT outcomes in this cohort; and to test the feasibility and efficacy of recalibrating the beta coefficients for HALTHCC as has been done for MELD in the past.

**Methods (619)**

*Patient Data and Competitor Models*

This multicenter trial was conducted using data from prospectively collected transplant records of patients undergoing LT for HCC from 2002-2014. The inclusion criteria were intentionally broad to generate a heterogeneous sample: adult recipients of LT (whole organ or partial) with a primary or secondary diagnosis of HCC. Patients were included regardless of whether they had access to preoperative LRT. Across sixteen sites and eight countries, 4,089 patients were registered. Patient selection, management, and follow-up was determined by center practice and local/regional/national law. Data sharing was restricted to variables from common clinical practice to construct the investigated risk models. Competitor models were included based on literature review and available variables. Search terms, search strategy, and model inclusion are described in **Supplemental Appendix (SA) Table 1**. This study was approved by the Institutional Review Boards at participating institutes and data maintained at the Cleveland Clinic (IRB#17-772).

### Study Design and Statistical Analysis

The TRIPOD-IV framework was followed in reporting this study.<sup>19</sup> Two lines of investigation were conducted:

1. The original HALTHCC score (oHALTHCC) was examined with respect to explant pathology and then on post-LT outcomes. For pathology, focus was placed on vascular invasion (VI) and poorly differentiated component (PDC) on explant analysis since these are most strongly associated with post-LT recurrence.<sup>12,20</sup> In the same oHALTHCC analysis, post-LT outcomes were analyzed to assess discriminatory ability in the broad international cohort.
2. To test the utility of recalibration, the study population was randomized into a training set and validation set. The training set was used to find the optimal weighting of HALTHCC parameters and then this recalibrated HALTHCC score (rHALTHCC) was examined in the validation set to assess discriminatory ability of all competitor models including oHALTHCC.

HALTHCC is a continuous score based on preoperatively accessible clinical characteristics which was previously reported.<sup>9</sup> Original HALTHCC (oHALTHCC) is calculated by: **Equation 1:**  $\text{oHALTHCC} = (1.85 * \ln(\text{AFP})) + (1.27 * \text{tumor burden score (TBS)}) + (0.26 * \text{MELD} - Na)$ . Overall survival (OS) was defined as the time interval from LT to either, mortality or last follow-up (where censored). Recurrence was monitored using center specific practice and reported as location and time from LT. HCC related mortality was defined as mortality after recurrence (either intra or extra-hepatic spread). Cox proportional hazards models, ordinal logistic models, and the Kaplan-Meier (KM) method were used to assess associations with OS, recurrence, and HCC related mortality. In order to maintain statistical validity and avoid overfitting, the cohort was divided into a training and validation set for rHALTHCC. Significant differences between the original cohort and this international cohort (SA

**Table 2)** strengthen the generalizability of any findings. Block randomization was performed in a 3:1 fashion (with each center serving as a block) to maintain similar representation of every center.<sup>21</sup> Recalibration was performed on the training cohort (n=1,021) using ordinal logistic modeling of a composite endpoint (one point for either recurrence or mortality and two points for recurrence leading to mortality) in an effort better balance the competing interests in allocation policy.<sup>22-24</sup> Throughout the modeling, attention was paid to ensure an adequate ratio of events to explanatory variables (ratios maintained >10:1) and proportional hazards assumption (assessed by Schoenfeld global test [ $P=0.08$ ]). Post-estimation concordance and discrimination was applied to the validation cohort (n=3,068) using Harrell's C-index, Akaike Information Criterion (AIC), net reclassification improvement (NRI) and integrated discrimination improvement (IDI).<sup>25</sup> Calibration between the oHALTHCC and rHALTHCC was compared by estimating predicted 1-,3- and 5-year survival and overall recurrence rates for 10 randomly selected cohorts amongst the 3,068 patients. Categorical variables are displayed as counts and percentages whereas continuous variables were reported as medians and interquartile range (25<sup>th</sup>-75<sup>th</sup> percentiles). All testing performed was two-sided and compared against a 5% alpha using STATA 13 (STATA Corp, College Station, Texas, USA).

**Results (832)**

Population Composition

The population distribution and characteristics are visualized in **Figure 1** and **SA Figure 1** and summarized in **SA Table 3**. Significant differences were seen in recipient underlying liver disease and MELD-Na at listing, tumor burden and AFP at listing, waiting time, proportion of candidates receiving LRT while waiting, and tumor characteristics at LT (**Figure 1A**). In addition to heterogeneity across world region, there were trends over time (**Figure 1** and **SA Figure 1**). For example, increased waiting time in North America due to shifts in allocation policy have led to increased penetration of LRT and decreased the active tumor burden at transplantation. The



Asian cohort changed dramatically over the study period, government reimbursement and focus on outcomes have decreased the tumor burden of candidates undergoing LT. The European cohort was more stable over time with only minor trends towards increasing penetration of LRT and some, likely related decreases in tumor burden at transplantation.

### *Explant Pathology*

The relationship between risk scores and pathological features is demonstrated in **Figure 2**. Amongst MC-in patients, VI and PDC on explant pathology were present in 19.2 and 11.1% of patients; MC-out status increased the incidences to 38.3 and 18.4%, respectively ( $P<0.0001$ ). The Metroticket 2.0 paradigm has introduced three groups of candidates amongst which LT should have comparable outcomes.<sup>26</sup> There was an increase in both VI and PDC from Group 1 (AFP<200 and Up-to-7; VI: 19.0%[95%CI:17.3-20.6] and PDT: 11.0%[95%CI:9.7-12.3]) to Groups 2 (AFP 200-400 and Up-to-5; VI: 40.7%[27.5-54.0] and PDT: 22.2%[11.0-33.4]) and 3 (AFP 400-1000 and Up-to-4; VI: 32.1%[14.5-49.8] and PDT: 14.3%[10.8-27.5]). Interestingly, patients not indicated to receive LT, here termed Group 4 (AFP>1000 or sum of lesion size and number greater than allowed with a given AFP for Groups 1-3) had similar rates of VI (45.9%[41.4-50.3]) and PDC (21.3%[17.6-24.9])(**Figure 2A**). Another alternative to HALTHCC is the MORAL score. As indicated in **Figure 2B** there was poor corroboration between increasing MORAL score and explant pathology. HALTHCC had a clear and robust association between estimated pre-LT risk and explant pathology, both VI and PDC increased with increasing HALTHCC score (**Figure 2C**). The lowest risk patients (HALTHCC 0-5) had much lower rates of VI and PDC than the highest risk patients (HALTHCC>35) (VI: 7.7%[1.2-14.2] vs 70.6%[48.3-92.9] and PDC: 4.6%[0.1-9.8%] vs 47.1%[22.6-71.5];  $P<0.0001$  for both). This trend was robust to MC status indicating that HALTHCC may be a useful tool to estimate risk of poor pathologic features using only preoperatively accessible variables.

*Modeling Post-LT Outcomes*

oHALTHCC was tested for association with post-LT outcomes (**Figure 3A-B**). This analysis showed generally promising validation of the US national experience; lower risk patients had higher survival and lower recurrence. However, in this cohort, the very low risk patients (oHALTHCC<5) and the penultimate high risk group (oHALTHCC 30-35) were overlapping with nearby risk cohorts; identifying clear areas where risk estimation could be improved. oHALTHCC's discrimination was consistent with the US national experience for recurrence (C-index: 0.69), OS (0.62), and HCC related death (0.72).

*Recalibration of HALTHCC*

The ordinal logistic modeling used to recalibrate HALTHCC is detailed in **Table 2**. This confirmed the highly statistically significant nature of the original components of HALTHCC and also provided evidence that regional variation (especially Asia) was contributing to deviations from risk-outcomes relationships being observed in the US experience. Together, these estimates were used to generate **Equation 2:**
$$rHALTHCC = (2.31 * \ln (AFP)) + (1.33 * (TBS)) + (0.25 * MELD - Na) - (5.57 * Asia).$$
The validation cohort was thusly estimated and then KM estimates of survival and recurrence compared to the oHALTHCC equation (**Figures 3C-D**). This analysis provided a greatly improved range of outcomes across all levels of risk. The 5-year post-LT OS ranged from 82.4% in the lowest risk group (rHALTHCC<5, n=145/3,068) to 32.4% in the highest risk group (rHALTHCC>35, n=24/3,068). A similar trend was observed with 5-year post-LT recurrence which ranged from, 8.6% (rHALTHCC<5, n=145/3,068) to 70.0% (rHALTHCC>35, n=24/3,068). The robustness with which rHALTHCC stratified risk of post-LT events was much greater than all competitor scores (**SA Figure 2**).

*Discriminatory Ability and Accuracy of HALTHCC*

Following recalibration of HALTHCC in the training set, comparison of discrimination for post-LT mortality, recurrence, and HCC-related mortality was conducted amongst competitors. Compared to its closest competitor, rHALTHCC was a statistically superior discriminator as measured by Harrell's C (0.63 [0.61-0.65] vs Metroticket 2.0 (0.57 [0.55-0.58]);  $P < 0.0001$ ), NRI (0.330 [0.261-0.391] vs Metroticket 2.0, 0.260 [0.201-0.315]), IDI (0.043 [0.031-0.057] vs Metroticket 2.0, 0.016 [0.009-0.025]), and AIC (14332 vs Metroticket 2.0, 14401), for post-LT survival (**SA Table 4**). This difference persisted for recurrence (0.71 [0.68-0.74] vs Metroticket 2.0, 0.65 [0.62-0.67];  $P$ -value  $< 0.0001$ ; **SA Table 5**) and HCC-related mortality (0.74 [0.71-0.77] vs Metroticket 2.0, 0.66 [0.63-0.69];  $P$ -value  $< 0.0001$ ; **SA Table 6**). Lastly, estimates of 1-, 3-, 5-year post-LT survival, and overall recurrence rates derived from o-, and, rHALTHCC were compared to observed values of these post-LT outcomes to demonstrate calibration of the estimating equations underlying their discriminatory ability (**SA Figure 3**). It is clear to observe visually how the linear regressions of the randomly selected clusters improve in their estimation of the observed values moving from A-B, C-D, etc, which confirms the model's improvement.

## Discussion (1434)

In a large and heterogeneous worldwide cohort, this validation study confirmed superior discriminatory characteristics of HALTHCC for explant pathology and post-LT outcomes. Recalibration of the score improved its prognostic utility. While MC was critical to widespread success in LT for HCC, its utility has decreased over-time due to changes in clinical practice. Many models were developed to replace MC and several did possess improved prognostic discrimination,<sup>6-8,16,20,27-29</sup> however, most fell by the wayside because of poor generalizability. Those models were developed in single institutes where the patient population and treatment strategy are homogeneous. As a result, they failed to show similar excellence in other cohorts. Consequently, other groups made their own model in their cohorts and our field has been trapped in a loop. HALTHCC was developed to defeat this loop and form the basis of a new

allocation system to improve prognostic utility and fair and reasonable organ allocation between HCC and non-HCC patients. This validation study using a heterogeneous, worldwide cohort with the hypothesis: if HALTHCC is truly measuring the disease HCC, it should work in any environment. Statistical analyses showed that there were extensive differences of tumor/treatment characteristics in 4,089 worldwide patients. The original HALTHCC score's performance was consistent with past studies as a discriminator of, recurrence (C-index: 0.69), OS (0.62), and HCC related mortality (0.72). Recalibration of HALTHCC was conducted and improved statistical discrimination over oHALTHCC as measured by C-index ( $P<0.001$  for recurrence,  $P=0.051$  for survival, and  $P=0.003$  for HCC-related mortality) and also visual discrimination using Kaplan Meier and calibration curves (**Figures 3 and SA Figure 3**).

While original HALT-HCC was validated in the US national experience, it was unknown how this clinical score would perform in an entirely new population with high heterogeneity with most patients outside the US (**SA Table 2**). As shown in **Figure 1**, this study cohort consisted of populations with different treatment strategies and tumor characteristics. Although it is known that high heterogeneity decreases performance of prediction models<sup>30</sup>, the oHALTHCC score maintained a high level of discrimination, especially for recurrence. Moreover, HALTHCC showed significant association with explant pathology, such as vascular invasion and poorly differentiated component (**Figure 2**). MC demonstrated a single level increase in rates of both VI and PDC consistent with its binary nature. However, when compared against all other criteria, increased HALTHCC score was most closely and reliably related to both VI and PDC. This is without any explicit modeling of pathology done at any phase in HALTHCC's development. These results suggest this continuous risk score is measuring tumor features underlying risk of poor outcomes not statistical static.

One might entertain the critique that a continuous score's form is too complicated and that our purpose seems more about chasing statistical excellence rather than usability. However, the transplant community is already familiar with the concept of continuous scoring,

demonstrated by MELD/MELD-Na scores, those risk metrics were similarly statistically derived and optimized. Moreover, it is well known that the alternative, categorical or ordinal scoring systems discard accuracy and their cutpoints promote controversy instead of easing decision making.<sup>18</sup> After HALTHCC's development, a close competitor, the Metroticket 2.0 model was introduced. This model is derived from a statistical basis and estimates risk from a continuous formulation which is then binned into risk categories based on risk planes. While we applaud the derivation, its application to bins is overly simplistic. The benefit realized by switching from a categorical logic to a continuous risk score is significant and encourages the community to embrace statistically valid and efficient methodologies. Organ allocation is a national issue; therefore, even small improvements in predictive utility could affect and contribute to large changes in organ utilization. We believed balancing statistical accuracy with a minimally complex model would increase the number of LT for HCC without adversely affecting outcomes. In fact, there is evidence that priority allocation using HALTHCC would allow fine tuning of "acceptable outcomes" and serve as a simple metric for transplant center comparison across regions.<sup>11</sup> Further, a continuous risk score is advantageous compared to a dichotomous model since the continuous score can be recalibrated easily when necessary, as was the case with the MELD and MELD-Na scores previously.

In this study, recalibration of HALTHCC was performed as planned at the conception of HALTHCC and this statistically significantly increased its discriminatory ability.<sup>9</sup> Following randomization of the study population into similarly composed cohorts (**Table 1**), the training set underwent ordinal logistic regression using a weighted composite endpoint. The estimation amongst the randomized cohort demonstrated statistical significance for all the factors in oHALTHCC. **SA Table 7** outlines the composition of the validation cohort stratified by rHALTHCC score and MC status. If we were to re-allocate the livers in this study based on composite risk, we can provide priority to nearly 74.6% of MC-out candidates without post-LT outcomes suffering. Further, we can reallocate ~4-5% of MC-in recipients with higher composite

1  
2  
3 risk. This approach would have the net effect of greater equity and lower post-LT recurrence  
4 rates. However, care must be taken since these estimates are all derived post-hoc in a cohort  
5 influenced by MC. A consensus meeting is necessary to finalize the form and composition of  
6 predictive models our community would like to explore prior to embarking on prospective trials.  
7 For example, we previously demonstrated the value of including MELD-Na in HALTHCC to  
8 improve prediction of overall survival, however it has relatively less impact on modeling  
9 recurrence.<sup>9,11</sup> Clearly recurrence is of prime interest in LT for HCC, however while 3-year  
10 recurrence runs 7-10%, transplant related mortality at 3-years is 15-20%<sup>26</sup>; candidate selection  
11 is about more than tumor size and number and transplantation is not benign. As a community  
12 we need to decide the most critical outcomes of interest for future prospective studies and then  
13 optimize our risk measure. A single system from presentation through locoregional therapy and  
14 into the transplant period is preferred since it would simplify a complex clinical decision tree but  
15 also improve outcomes.

16  
17  
18 Moving away from waiting time based standard exception points with individualized  
19 applications to a mathematic transformation of risk and benefit has many advantages. In Figure  
20 4 six example patients are outlined to demonstrate three proposed models of allocation through  
21 which to explore competing concepts in allocation related to HCC. It is important to note these  
22 draft models are simply examples to discuss possible strategies in allocation, a final or “best  
23 model” will require advanced simulation studies to estimate the impact between non-HCC and  
24 HCC cohort dropout, transplant volumes, and post-LT outcomes. In the first draft model, post-LT  
25 survival is maximized (model alpha). This model prioritizes long term survival after liver  
26 transplantation alone, patients who have lowest HALTHCC score prior to organ allocation wait  
27 the least time, have the least disease burden and have the highest overall survival rates and  
28 lowest recurrence. However, one can argue that many of these low risk patients would go for  
29 many months after LRT without disease recurrence or a new lesion, in contrast many patients  
30 who could have benefited from timely transplantation may have progressed. Second, the model

prioritizing oncological risk in a fashion similar to the “sickest first” policy (model gamma). In this model, high HALTHCC patients have the highest priority since they presumably have the most to benefit from immediate and early transplantation. However, this model performs poorly in metrics of organ utilization and post-LT population recurrence rates. It could be argued that many moderate risk patients with lower priority may not have progressed had they been transplanted early whereas this model may inadvertently prioritize transplantation of patients with almost certain distant microscopic disease. Lastly the model balancing the middle group of patients, those with near mean risk best served by timely transplantation (model beta). These patients are clearly at risk of dropout but also have a reasonable chance at cure from timely transplant. In model beta, patients closest to the mean historical risk at transplant achieve the highest priority by simply transforming their mathematical distance from the mean using absolute values. Patients with very high risk at presentation may naturally dropout but may be able to get sufficient priority to access LT if they have an appropriate response to LRT as measured by HALTHCC.<sup>11</sup> Patients with very low risk may go many months or even years after LRT without disease recurrence, however, with the field effect of cirrhosis in place it is likely that another lesion will at some point crop up; in that case their HALTHCC score will increase and their priority for transplantation will increase. This balanced approach provides a compromise between sickest first and survival conscious or utility concepts. The beauty of all of these models compared to something like MC or other binary systems (including the complex criteria currently required to meet exceptional status under UNOS) is that there is no cutoff at which point patients with only marginally increased risk suffer from total ineligibility (eg one tumor sized 5.2cm vs one tumor sized 4.8cm or AFP level of 1005 vs 995). Lastly, these models can be easily adjusted in two future oriented ways:

1. They can have the constant term adjusted by UNOS region to help negotiate disparities in rates of transplantation by MELD/priority score.



2. They can have the multipliers modified to adjust the risk tolerance of the society if recurrence rates are too high or there is too much dropout.

These models serve as the most basic and first in a series of studies our group hopes to employ to estimate the impact advanced, HALTHCC based, allocation models can have on our society.

Despite its size, this large study has some limitations to consider. First, it is a retrospective study, therefore we cannot definitively infer any causality related to HALTHCC score. Also, due to time constraints we were not able to collect detailed longitudinal data on tumor size, serum tumor markers, exact times of multiple LRT, nor the response to LRT. Methodologically, assumptions need to be made to adjust for regional differences in outcomes. One can either allow the individual coefficients to vary for each region (ie estimate a new equation per region) or allow an omnibus adjustment at the end, assuming biological determinants like AFP, TBS, MELD-Na will have similar orders of influence by region. In scenario 1 above, we assume knowledge of the mechanism by which outcomes vary between different locations in the world. In reality, it could be population genetics, donor types, or other unspecified practice patterns which most significantly account for the difference in outcomes for a given HALTHCC score (unadjusted by region). In contrast, in scenario 2 we acknowledge that outcomes for similar patient characteristics may differ around the world. However, that doesn't mean they are not comparable, nor that the difference is not estimable. In fact, the difference is estimable, we just don't know the mechanism through which that difference acts. Our assumption in acting through scenario 2 is that AFP, TBS, and MELD-NA have a relatively similar contribution (at least in a similar order of magnitude) to risk to poorer outcome across location. However overall risk of a given outcome is not proportional strictly to those biological factors. Since the mechanism of risk modification is unknown we chose the method of omnibus adjustment of risk for a given location. In fact, this method is the most conservative from both a clinical reasoning standpoint but also statistically since we introduce far fewer additional assumptions, tests and thus reduce opportunity for bias. Further a sensitivity analysis was



performed demonstrating that the coefficients for AFP, TBS, and MELD-Na were similar between the Western and Eastern Cohorts (**SA Tables 8 and 9**). Estimating the baseline hazard function between each region separately demonstrated that proportionality was not violated but that the baseline hazard is simply lower in Asia, thus our estimates of the impact of AFP, TBS and MELD-Na are valid (**SA Figures 4A-B**). Next we calculated model utility on the Western dataset alone only which only marginally decreased the utility estimates by Harrell's C (**SA Table 10**). Additionally, calibration of HALTHCC derived overall survival and recurrence predictions compared to observations were not systematically biased by region whereas its nearest competitor model Metroticket 2.0 consistently overestimated the Asian cohorts risk (**SA Figures 4C-D**). Finally, since we did not have a complete data set of all previously reported serum markers and all the individual patients' longitudinal courses we were unable to draw comparisons on utility to several recent scoring systems including the delta-HALTHCC and TRAIN scores.<sup>7,11</sup>

In conclusion this large international trial validated and refined the role for the continuous risk metric, HALTHCC. Its utility was confirmed amongst 4,089 patients undergoing LT for HCC, composing a heterogeneous data set, which strengthens generalizability. While MC has served the transplant community well for two-decades it is time for providers in this field to have our practice "bridged" until whole exome studies provide truly personalized guidance to HCC management. We proposed various models to inform allocation algorithms based on HALTHCC. Prospective trials introducing HALTHCC into observational frameworks to remove the bias laid out by two decades of practice under MC are warranted.

References:

1. El-Serag, H. B. Hepatocellular Carcinoma. *N. Engl. J. Med.* **365**, 1118–1127 (2011).

2. El-Serag, H. B. & Kanwal, F. Epidemiology of hepatocellular carcinoma in the United States: Where are we? Where do we go?: HEPATOLOGY, Vol. XX, No. X, 2014 EL-SERAG AND KANWAL. *Hepatology* **60**, 1767–1775 (2014).

3. IWATSUKI, SHUNZABURO, STARZL, THOMAS E., SHEAHAN, DANIEL G., YOKOYAMA, ITSUO, DEMETRIS, ANTHONY J., TODO, SATURO, TZAKIS, ANDREAS G., VAN THIEL, DAVID H., CARR, BRIAN, SELBY, RICHARD & MADARIAGA, JUAN. Hepatic Resection Versus Transplantation for Hepatocellular Carcinoma. *Ann. Surg.* **214**, 221–229 (1991).

4. Llovet, J. M., Fuster, J., Bruix, J. & Barcelona Clinic Liver Cancer (BCLC) Group. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: Resection versus transplantation. *Hepatology* **30**, 1434–1440 (1999).

5. Mazzaferro, V., Regalia, E., Doci, R., Andreola, S., Pulvirenti, A., Bozzetti, F., Montalto, F., Ammatuna, M., Morabito, A. & Gennari, L. Liver Transplantation for the Treatment of Small Hepatocellular Carcinomas in Patients with Cirrhosis. *N. Engl. J. Med.* **334**, 693–700 (1996).

6. Uchiyama, H., Itoh, S., Yoshizumi, T., Ikegami, T., Harimoto, N., Soejima, Y., Harada, N., Morita, K., Toshima, T., Motomura, T. & Maehara, Y. Living donor liver transplantation for hepatocellular carcinoma: results of prospective patient selection by Kyushu University Criteria in 7 years. *HPB* **19**, 1082–1090 (2017).

7. Lai, Q., Nicolini, D., Inostroza Nunez, M., Iesari, S., Goffette, P., Agostini, A., Giovagnoni, A., Vivarelli, M. & Lerut, J. A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) Score. *Ann. Surg.* **264**, 787–796 (2016).

8. Sapisochin, G., Goldaracena, N., Laurence, J. M., Dib, M., Barbas, A., Ghanekar, A., Cleary, S. P., Lilly, L., Cattal, M. S., Marquez, M., Selzner, M., Renner, E., Selzner, N., McGilvray, I. D., Greig, P. D. & Grant, D. R. The extended Toronto criteria for liver transplantation in patients with hepatocellular carcinoma: A prospective validation study. *Hepatology* **64**, 2077–2088 (2016).
9. Sasaki, K., Firl, D. J., Hashimoto, K., Fujiki, M., Diago-Uso, T., Quintini, C., Egtesad, B., Fung, J. J., Aucejo, F. N. & Miller, C. M. Development and validation of the HALT-HCC score to predict mortality in liver transplant recipients with hepatocellular carcinoma: a retrospective cohort analysis. *Lancet Gastroenterol. Hepatol.* **2**, 595–603 (2017).
10. Yao, F. Liver transplantation for hepatocellular carcinoma: Expansion of the tumor size limits does not adversely impact survival. *Hepatology* **33**, 1394–1403 (2001).
11. Firl, D. J., Kimura, S., McVey, J., Hashimoto, K., Yeh, H., Miller, C. M., Markmann, J. F., Sasaki, K. & Aucejo, F. N. Reframing the approach to patients with hepatocellular carcinoma: Longitudinal assessment with HALTHCC improves ablate and wait strategy. *Hepatology* (2018).  
doi:10.1002/hep.29907
12. Silva, M. F. & Sherman, M. Criteria for liver transplantation for HCC: What should the limits be? *J. Hepatol.* **55**, 1137–1147 (2011).
13. Hoshida, Y., Villanueva, A., Kobayashi, M., Peix, J., Chiang, D. Y., Camargo, A., Gupta, S., Moore, J., Wrobel, M. J., Lerner, J., Reich, M., Chan, J. A., Glickman, J. N., Ikeda, K., Hashimoto, M., Watanabe, G., Daidone, M. G., Roayaie, S., Schwartz, M., Thung, S., Salvesen, H. B., Gabriel, S., Mazzaferro, V., Bruix, J., Friedman, S. L., Kumada, H., Llovet, J. M. & Golub, T. R. Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).
14. Nault, J., De Reyniès, A., Villanueva, A., Calderaro, J., Rebouissou, S., Couchy, G., Decaens, T., Franco, D., Imbeaud, S., Rousseau, F., Azoulay, D., Saric, J., Blanc, J., Balabaud, C., Bioulac-Sage, P., Laurent, A., Laurent-Puig, P., Llovet, J. M. & Zucman-Rossi, J. A Hepatocellular Carcinoma 5-Gene

Score Associated With Survival of Patients After Liver Resection. *Gastroenterology* **145**, 176–187 (2013).

15. Lee, H. W., Song, G.-W., Lee, S.-G., Kim, J. M., Joh, J.-W., Han, D. H., Kim, S. I., Kim, S. H., Kim, D.-S., Cho, J. Y. & Suh, K.-S. Patient selection by tumor markers in liver transplantation for advanced hepatocellular carcinoma. *Liver Transpl.* (2018). doi:10.1002/lt.25056

16. Marsh, J. W., Dvorchik, I., Subotin, M., Balan, V., Rakela, J., Popechitelev, E. P., Subbotin, V., Casavilla, A., Carr, B. I., Fung, J. J. & Iwatsuki, S. The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation: A pilot study. *Hepatology* **26**, 444–450 (1997).

17. Zhang, M., Yin, F., Chen, B., Li, B., Li, Y. P., Yan, L. N. & Wen, T. F. Mortality risk after liver transplantation in hepatocellular carcinoma recipients: A nonlinear predictive model. *Surgery* **151**, 889–897 (2012).

18. Royston, P., Altman, D. G. & Sauerbrei, W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat. Med.* **25**, 127–141 (2006).

19. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann. Intern. Med.* **162**, 55 (2015).

20. Agopian, V. G., Harlander-Locke, M., Zarrinpar, A., Kaldas, F. M., Farmer, D. G., Yersiz, H., Finn, R. S., Tong, M., Hiatt, J. R. & Busuttil, R. W. A Novel Prognostic Nomogram Accurately Predicts Hepatocellular Carcinoma Recurrence after Liver Transplantation: Analysis of 865 Consecutive Liver Transplant Recipients. *J. Am. Coll. Surg.* **220**, 416–427 (2015).

21. Matts, J. P. & Lachin, J. M. Properties of permuted-block randomization in clinical trials. *Control. Clin. Trials* **9**, 327–344 (1988).

22. Armstrong, P. W., Westerhout, C. M., Van de Werf, F., Califf, R. M., Welsh, R. C., Wilcox, R. G. & Bakal, J. A. Refining clinical trial composite outcomes: An application to the Assessment of the Safety and Efficacy of a New Thrombolytic—3 (ASSENT-3) trial. *Am. Heart J.* **161**, 848–854 (2011).
23. Rauch, G., Jahn-Eimermacher, A., Brannath, W. & Kieser, M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat. Med.* **33**, 1104–1120 (2014).
24. Ananth, C. Regression models for ordinal responses: a review of methods and applications. *Int. J. Epidemiol.* **26**, 1323–1333 (1997).
25. Pencina, M. J., D'Agostino, R. B., D'Agostino, R. B. & Vasan, R. S. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat. Med.* **27**, 157–172 (2008).
26. Mazzaferro, V., Sposito, C., Zhou, J., Pinna, A. D., De Carlis, L., Fan, J., Cescon, M., Di Sandro, S., Yi-Feng, H., Lauterio, A., Bongini, M. & Cucchetti, A. Metroticket 2.0 Model for Analysis of Competing Risks of Death Following Liver Transplantation for Hepatocellular Carcinoma. *Gastroenterology* (2017). doi:10.1053/j.gastro.2017.09.025
27. Lai, Q., Inostroza, M., Rico Juri, J. M., Goffette, P. & Lerut, J. Delta-slope of alpha-fetoprotein improves the ability to select liver transplant patients with hepatocellular cancer. *HPB* **17**, 1085–1095 (2015).
28. Halazun, K. J., Najjar, M., Abdelmessih, R. M., Samstein, B., Griesemer, A. D., Guarrera, J. V., Kato, T., Verna, E. C., Emond, J. C. & Brown, R. S. Recurrence After Liver Transplantation for Hepatocellular Carcinoma: A New MORAL to the Story. *Ann. Surg.* **1** (2016). doi:10.1097/SLA.0000000000001966
29. Kaido, T., Ogawa, K., Mori, A., Fujimoto, Y., Ito, T., Tomiyama, K., Takada, Y. & Uemoto, S. Usefulness of the Kyoto criteria as expanded selection criteria for liver transplantation for hepatocellular carcinoma. *Surgery* **154**, 1053–1060 (2013).

30. Ioannidis, J. P. & Lau, J. Heterogeneity of the baseline risk within patient populations of clinical trials: a proposed evaluation algorithm. *Am. J. Epidemiol.* **148**, 1117–1126 (1998).

Tables:

Table 1: Population Characteristics			
	Training Cohort (n=1,021)	Validation Cohort (n=3,068)	P-value
Transplant Year	2008 [2005, 2012]	2008 [2005, 2012]	0.920
World Region			0.740
North America	460 (45.1)	1391 (45.3)	
Europe	468 (45.8)	1399 (45.6)	
Asia	93 (9.1)	278 (9.1)	
Sex, male	689 (67.5)	2133 (69.5)	0.222
Age, years	58 [52, 63]	58 [52, 63]	0.360
Underlying Liver Disease*			0.667
HCV cirrhosis	560 (54.8)	1683 (54.9)	
HBV cirrhosis	192 (18.8)	501 (16.3)	
Laennec's cirrhosis	203 (19.9)	623 (20.3)	
NASH cirrhosis	50 (4.9)	189 (6.2)	
Other	77 (7.5)	248 (8.1)	
Listing Characteristics			
Laboratory MELD-Na, points	12.0 [9.0, 16.0]	12.0 [9.0, 16.0]	0.884
Tumor Number	1.0 [1.0, 3.0]	1.0 [1.0, 3.0]	0.655
Tumor Size, cm	2.6 [1.9, 3.6]	2.5 [1.8, 3.6]	0.477
TBS, points	3.4 [2.5, 4.8]	3.4 [2.4, 4.6]	0.445
AFP, IU/mL	10.0 [5.0, 37.9]	10.0 [4.7, 39.3]	0.553
Patients meeting MC	746 (73.1)	2173 (71.0)	0.786
HALTHCC, points	12.6 [10.0, 15.8]	12.6 [10.3, 15.9]	0.468
Waiting time, days	157 [60, 318]	149 [60, 315]	0.557
Pre-LT Characteristics			
Laboratory MELD-Na, points	13.0 [9.0, 18.0]	13.0 [9.0, 17.6]	0.914
Tumor Number	1.0 [1.0, 2.0]	1.0 [1.0, 3.0]	0.218
Tumor Size, cm	2.1 [1.1, 3.0]	2.1 [1.0, 3.0]	0.536
TBS, points	3.06 [1.4, 4.2]	3.05 [1.8, 4.2]	0.562
AFP, IU/mL	10.0 [4.5, 35.0]	9.8 [4.2, 37.4]	0.448
NLR	2.7 [1.7, 4.3]	2.7 [1.7, 4.4]	0.249
Patients meeting MC	752 (73.7)	2308 (75.2)	0.303
HALTHCC, points	12.0 [9.0, 15.6]	12.2 [9.4, 15.6]	0.170
Follow-up time, days	1369 [529, 2700]	1445 [606, 2753]	0.378
<b>Note:</b> Continuous variables: median [interquartile range]. *May not sum to unity due to overlapping etiologies.			
<b>Abbreviations:</b> AFP: preoperative alpha-fetoprotein, HALTHCC: Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma, HBV: hepatitis B virus, HCC: hepatocellular carcinoma, HCV: hepatitis C virus, MC: Milan Criteria, MELD-Na: Model of End Stage Liver Disease Sodium, NASH: non-alcoholic steatohepatitis, NLR: neutrophil-lymphocyte ratio, SRTR: Scientific Registry of Transplant Recipients, TBS: Tumor Burden Score.			

**Table 2:** Ordinal Logistic Model for Recalibrating HALTHCC using Training Set

	Coefficient	Standard error	Wald test	Odds Ratio (95% CI)	P-value*
MELD-NA, per point	0.0247	0.0088	2.82	1.025 (1.008-1.043)	0.005
TBS, pre point	0.1332	0.0233	5.71	1.143 (1.091-1.196)	<0.0001
AFP, per ln(IU/mL)	0.2308	0.0357	6.46	1.260 (1.175-1.351)	<0.0001
World Region					
North America	-Base case-	-	-	-	-
Europe	0.1817	0.141	1.29	1.199 (0.910-1.581)	0.198
Asia	-0.5571	0.261	-2.13	0.573 (0.343-0.955)	0.033

\*Based on likelihood est adjusted for the other factors in the final model.

**Abbreviations:** AFP: preoperative alpha-fetoprotein (natural log transformed), HALTHCC: Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma, MELD-Na: Model of End Stage Liver Disease Sodium, TBS: Tumor burden score

## Figure Legends

### Figure 1: Trends in Worldwide Liver Transplantation for Hepatocellular Carcinoma

A) Cartoon map of the world overlaid with the characteristics of liver transplantation (LT) for hepatocellular carcinoma (HCC) in each of the three major regions in this study. B) Proportion of LT recipients undergoing locoregional therapy prior to LT by year of LT and center region. C) MELD-Na measured just prior to LT by year of LT and center region. D) Number of viable lesions on radiographic evaluation at time of candidate listing, by year of LT and center region. E) Number of viable lesions on radiographic evaluation at last pre-LT evaluation, by year of LT and center region. F) Largest viable lesion size on radiographic evaluation at time of candidate listing, by year of LT and center region. G) Largest viable lesion size on radiographic evaluation at last pre-LT evaluation, by year of LT and center region. **NOTE:** Panel A – Boxed descriptive statistics are median [interquartile range]. Panels B-G – Bars represent means with 95% confidence intervals.

### Figure 2: Utility of Pre-Operatively Assessable Risk Metrics as Measured by Explant Pathological Features

A) Demonstrates probability of finding vascular invasion or poor tumor differentiation of explant pathology for Milan Criteria (in vs out) and Metroticket 2.0 groupings (indicated LT patients

[Groups 1-3] vs not indicated [Group 4]). B) Explant pathology rates for vascular invasion and poor tumor differentiation grading by MORAL score. C) Explant pathology rates of vascular invasion and poor tumor differentiation grading by HALTHCC score. **Abbreviations:** G: group, HALTHCC: Hazard associated with liver transplantation for hepatocellular carcinoma, MC: Milan criteria, MORAL: Model of recurrence after liver transplantation, MT2.0: Metroticket version 2.0.

**NOTE:** Bars demonstrate means with 95% confidence intervals for estimates.

**Figure 3: HALTHCC Score is Associated with Overall Survival and Recurrence and Improved after Recalibration**

A) Kaplan-Meier (KM) survival estimates for overall survival by hazard associated with liver transplantation for HCC (HALTHCC) score by to any model recalibration. B) KM estimates for recurrence by HALTHCC score prior to any model recalibration. C) KM survival estimates for overall survival following HALTHCC recalibration using the training cohort. D) KM estimates for recurrence following HALTHCC recalibration using the training cohort. **NOTE:** Only validation cohort patients are included in the analysis of Panels A-D, the training cohort for recalibration was excluded to avoid overfitting.

**Figure 4: Example Patients to Demonstrate Utility of HALTHCC Based Allocation**

Six example patients illustrate the ease and utility of longitudinal assessment with HALTHCC of HCC patients undergoing evaluation for LT. The proposed models vary in emphasis from purely post-operative survival to a sickest first and finally balanced approach. Further a simple adjustment for regional competitiveness or to balance risk of dropout can be conducted through modification of the constant terms.



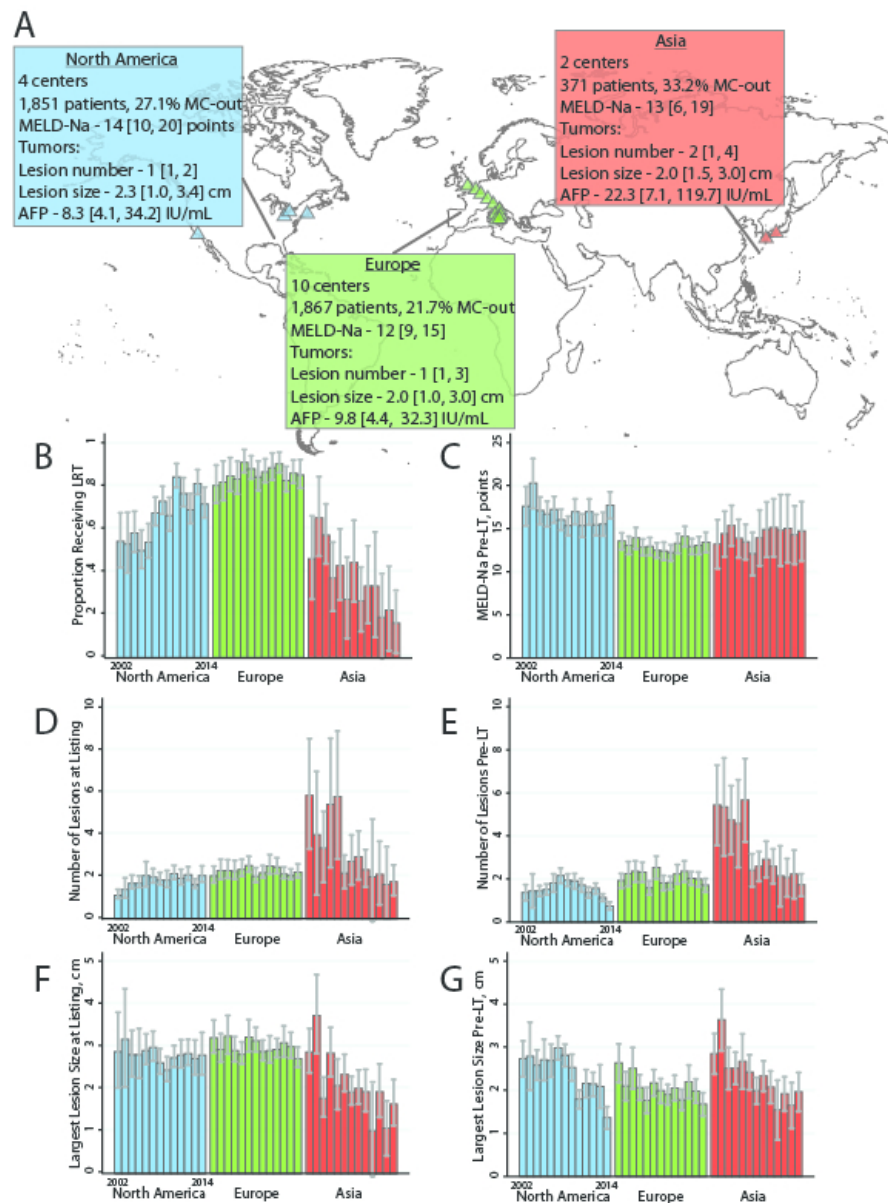


Figure 1

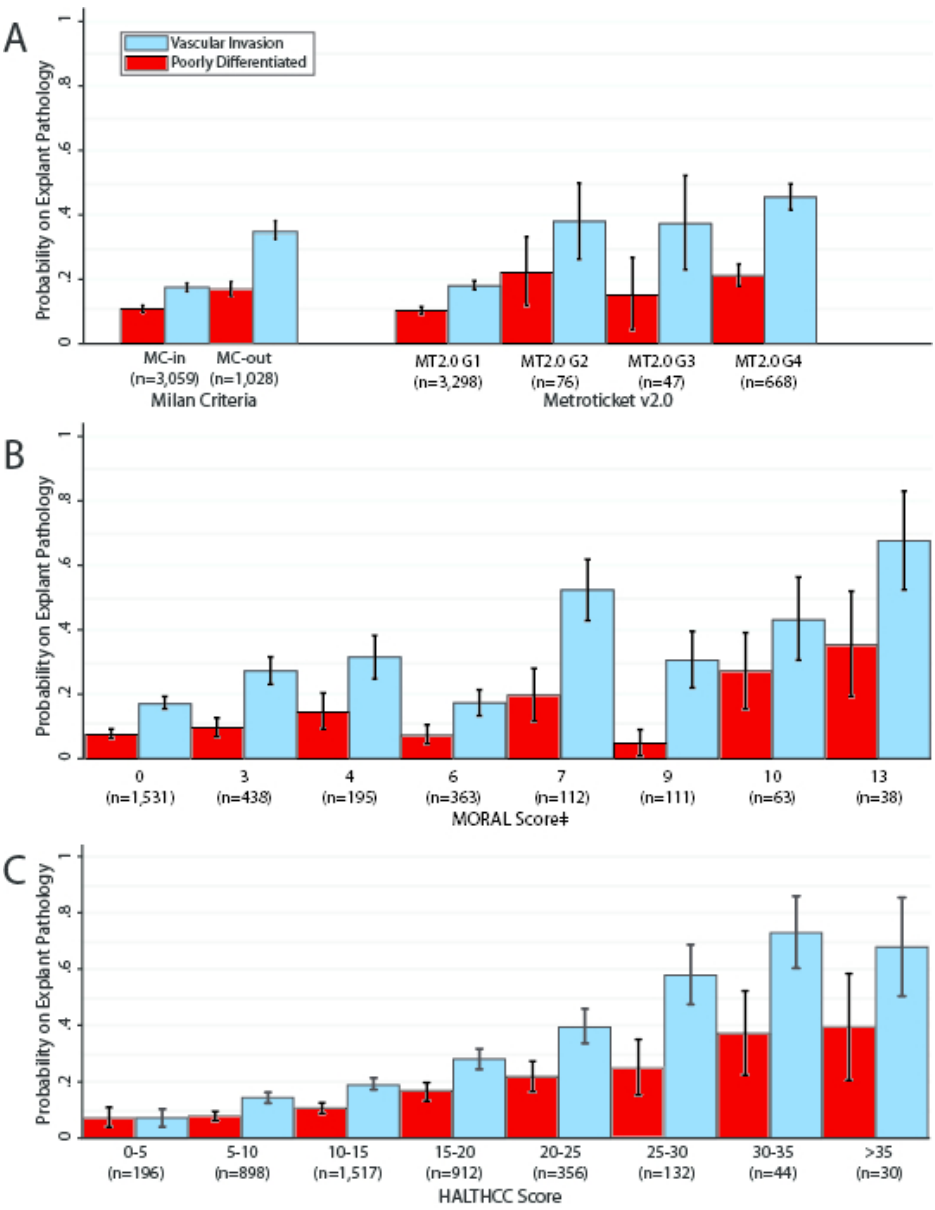


Figure 2

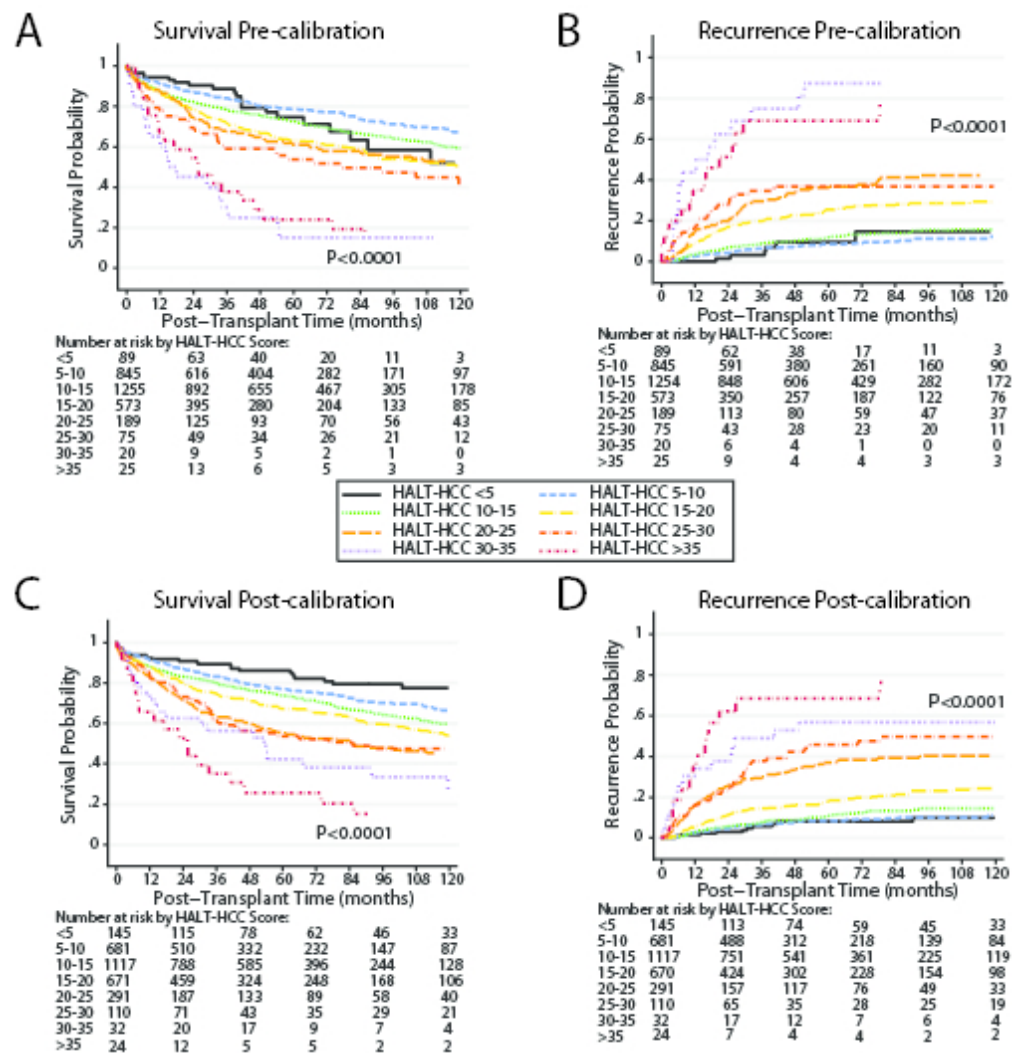


Figure 3

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

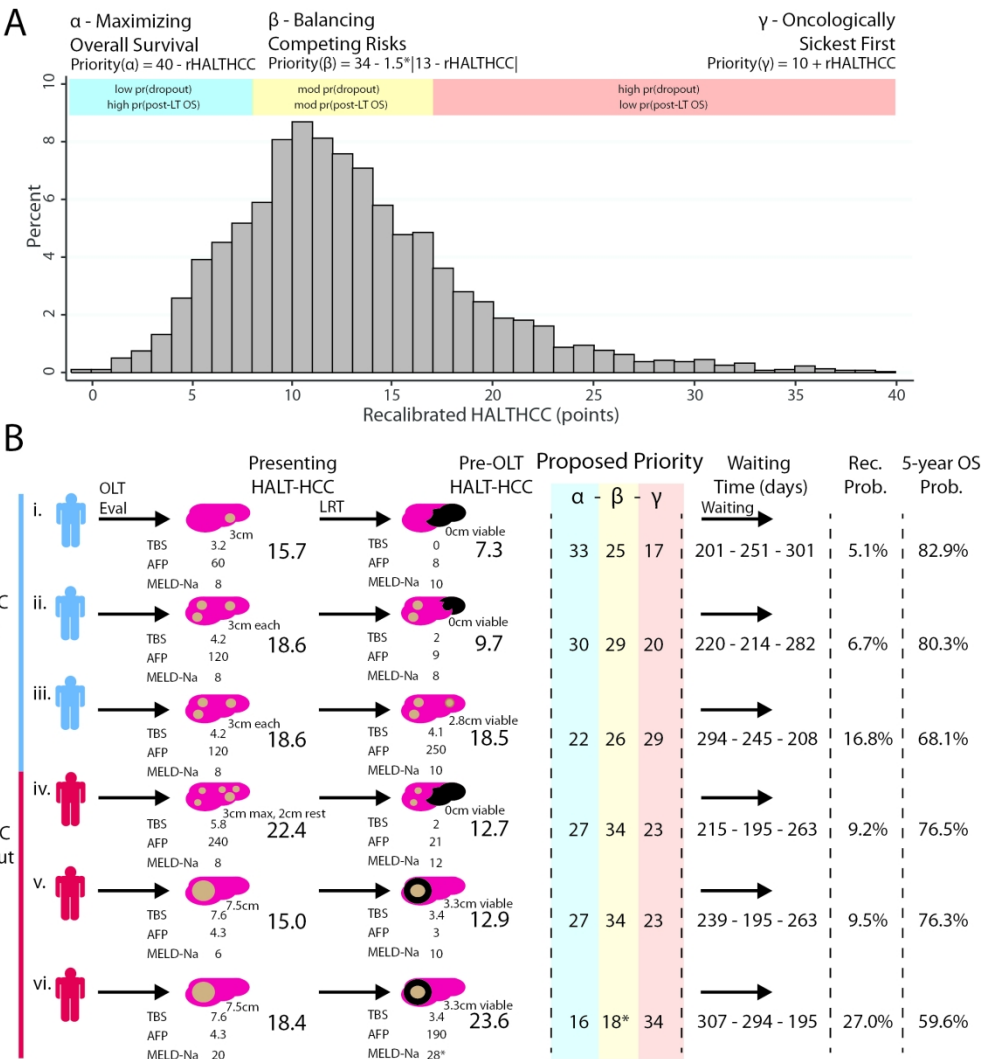


Figure 4

# Charting the Path Forward for Risk Prediction in Liver Transplant for HCC: International Validation of HALTHCC amongst 4,089 patients

Daniel J. Firl<sup>1,§</sup>, Kazunari Sasaki<sup>1,§</sup>, Vatche G. Agopian<sup>2</sup>, Andre Gorgen<sup>3</sup>, Shoko Kimura<sup>4</sup>, Withit Dumronggittigule<sup>2</sup>, John C. McVey<sup>1</sup>, Samuele Iesari<sup>5</sup>, Gianluca Mennini<sup>6</sup>, Alessandro Vitale<sup>7</sup>, Armin Finkenstedt<sup>8</sup>, Simona Onali<sup>9</sup>, Maria Hoppe-Lotichius<sup>10</sup>, Giovanni Vennarecci<sup>11</sup>, Tommaso M. Manzia<sup>12</sup>, Daniele Nicolini<sup>13</sup>, Alfonso W. Avolio<sup>14</sup>, Salvatore Agnes<sup>14</sup>, Marco Vivarelli<sup>13</sup>, Giuseppe Tisone<sup>12</sup>, Giuseppe M. Ettorre<sup>11</sup>, Gerd Otto<sup>10</sup>, Emmanuel Tsochatzis<sup>9</sup>, Massimo Rossi<sup>6</sup>, Andre Viveiros<sup>8</sup>, Umberto Cillo<sup>7</sup>, Jan Lerut<sup>5</sup>, James F Markmann<sup>4</sup>, Toru Ikegami<sup>15</sup>, Toshimi Kaido<sup>16</sup>, Gonzalo Sapisochin<sup>3</sup>, Quirino Lai<sup>5,6</sup> on behalf of the European Hepatocellular Cancer Liver Transplant Study Group, and Federico N. Aucejo<sup>1</sup>.

**Affiliations:** <sup>1</sup>Department of General Surgery and Cleveland Clinic Lerner College of Medicine, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>§</sup>Current Address: Department of Surgery, Duke University Hospital, Durham, NC, USA. <sup>2</sup>Dumont-UCLA Transplant and Liver Cancer Center, Department of Surgery, Ronald Reagan UCLA Medical Center and David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>3</sup>Department of Abdominal Transplant and HPB Surgical Oncology, University Health Network and University of Toronto, Toronto, Canada. <sup>4</sup>Transplant Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>5</sup>Starzl Unit of Abdominal Transplantation, St. Luc University Hospital, Université Catholique Louvain, Brussels, Belgium. <sup>6</sup>Department of General Surgery and Organ Transplantation, Umberto I Hospital, Sapienza University, Rome, Italy. <sup>7</sup>Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy. <sup>8</sup>Department of Medicine I, Medical University Innsbruck, Innsbruck, Austria. <sup>9</sup>UCL Institute for Liver and Digestive Health and Royal Free Sherlock Liver Centre, Royal Free Hospital and UCL, London, UK. <sup>10</sup>Department of Transplantation and Hepatobiliary Surgery, University of Mainz, Mainz, Germany. <sup>11</sup>Division of General Surgery and Liver Transplantation, San Camillo Hospital, Rome, Italy. <sup>12</sup>Department of Transplant Surgery, Polyclinic Tor Vergata Foundation, Tor Vergata University, Rome, Italy. <sup>13</sup>Unit of Hepatobiliary Surgery and Transplantation, Azienda Ospedaliero-Universitaria Ospedali Riuniti, Torrette Ancona, Italy. <sup>14</sup>Liver Unit, Department of Surgery, Agostino Gemelli Hospital, Catholic University, Rome, Italy. <sup>15</sup>Kyoto University Hospital, Kyoto, Japan. <sup>16</sup>Kyushu University Hospital, Kyushu, Japan.

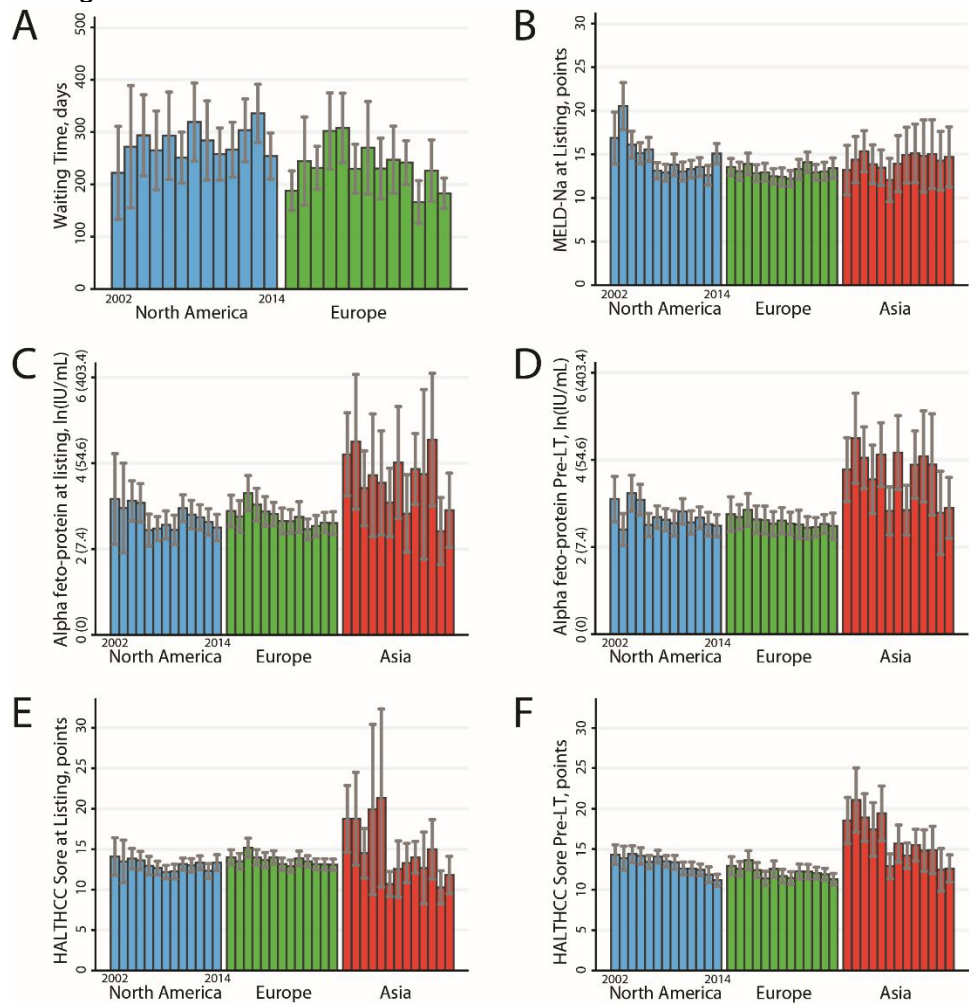
**§Corresponding Author:** Kazunari Sasaki MD, Department of General Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, OH, USA. E: sasakik@ccf.org

**Funding Source:** KS received an American Society of Transplant Surgeons Fellowship related to this work. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

**Disclosures:** The authors declare no conflicts or competing interests as described by *Hepatology* guidelines.

**Keywords:** liver transplantation, risk models, selection criteria, international study

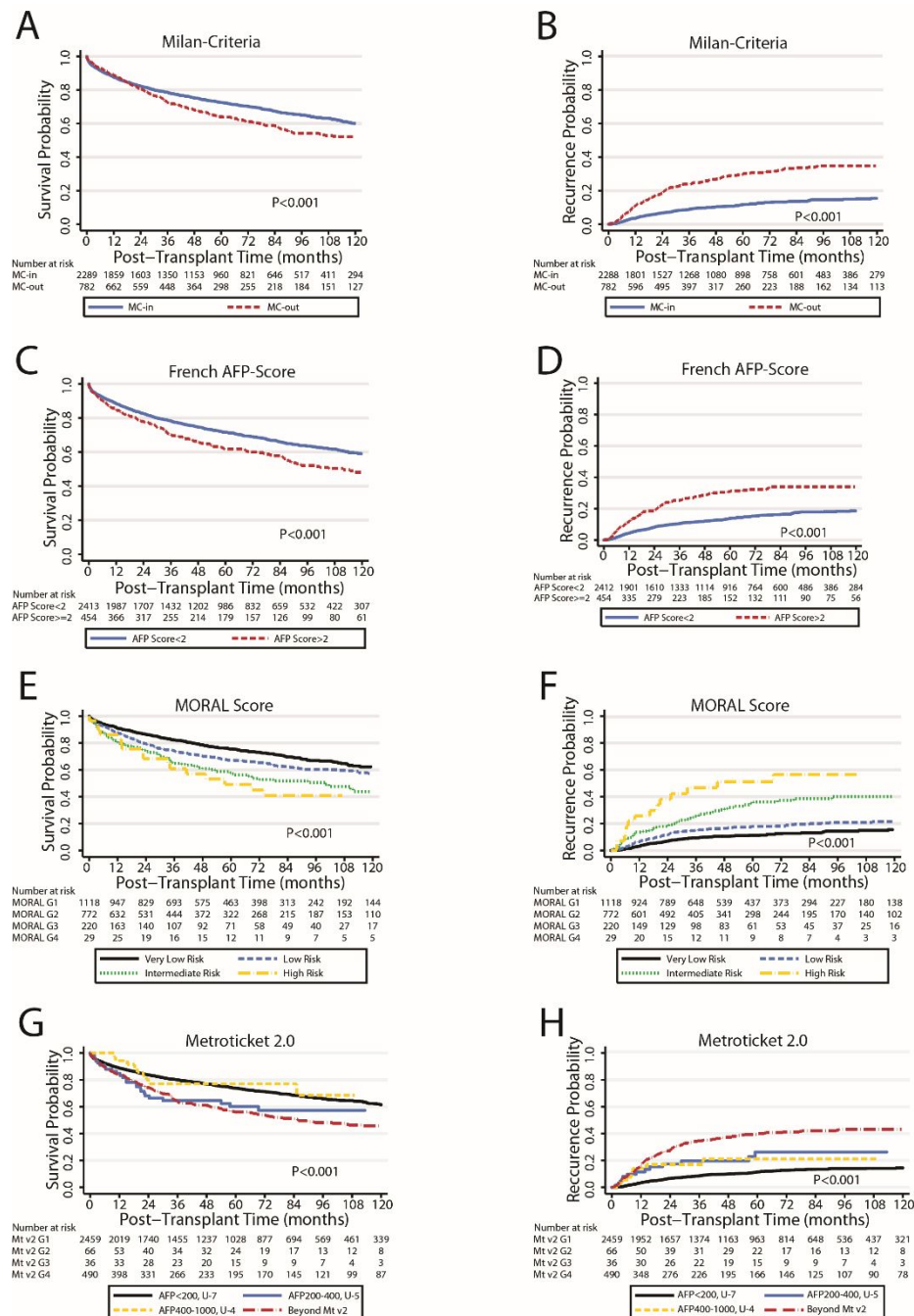
Supplemental Figure 1: Extended Trends in Worldwide LT for HCC



A) Waiting time (days from listing for LT to transplantation) by year and center (note: no Asia centers are graphed since there were only living donation LT). B) MELD-Na measured at candidate listing for LT, by year of LT and center region. C) Alpha fetoprotein level measured at candidate listing for LT, by year of LT and center region. D) Alpha fetoprotein level measured just prior to LT, by year of LT and center region. E) Hazard associated with liver transplantation for HCC (HALTHCC) score at time of candidate listing, by year of LT and center region. F) HALTHCC score just prior to LT, by year of LT and candidate region. **NOTE:** Panels A-F – Bars represent means with 95% confidence intervals.

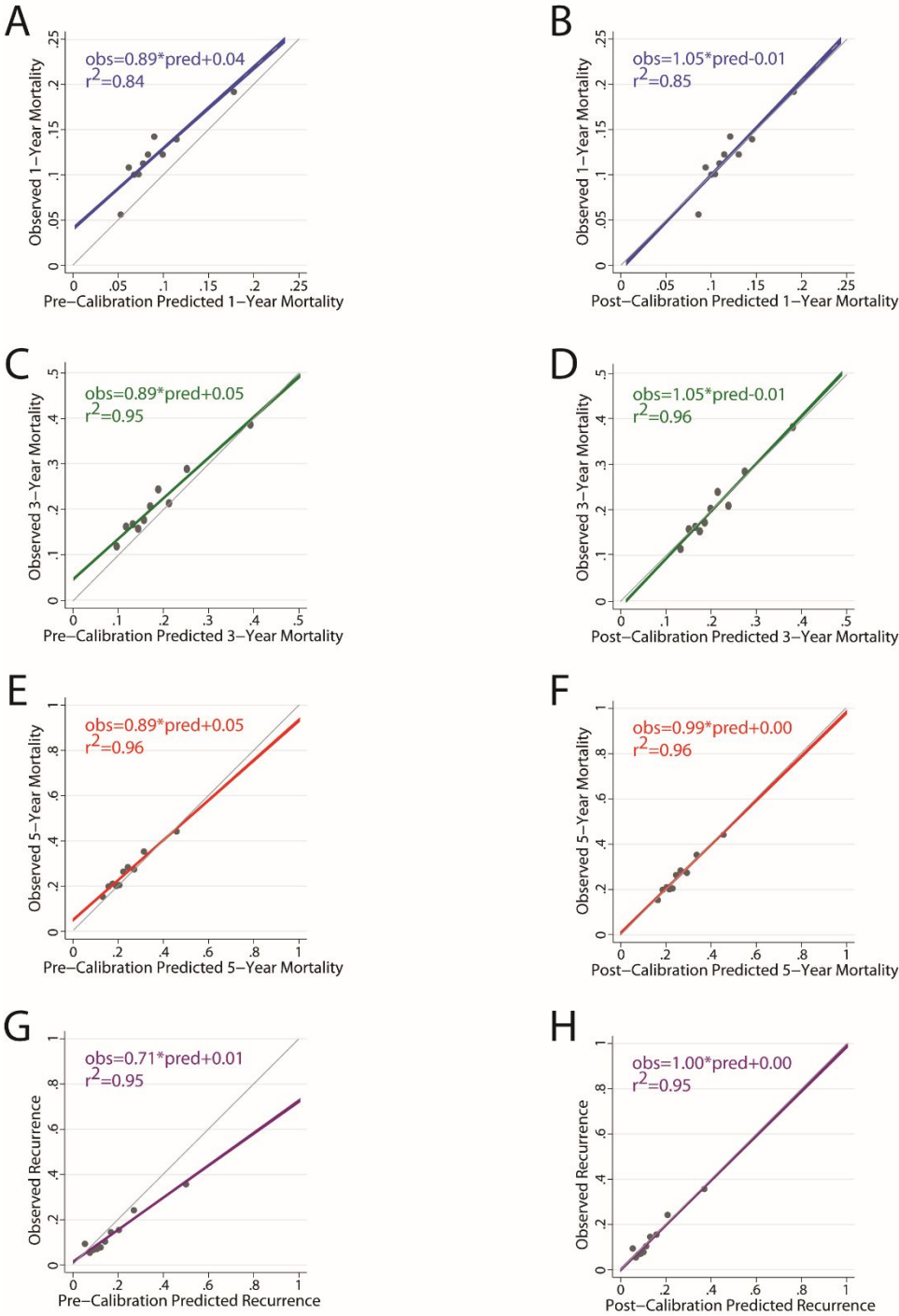


Supplementary Figure 2: Post-LT Outcomes by Competitor Risk Scores



A) Kaplan-Meier (KM) survival estimates by Milan Criteria. B) As in A but for recurrence estimates. C) KM survival estimates for French AFP-Score. D) As in C but for recurrence estimates. E) KM survival estimates for MORAL Score. F) As in E but for recurrence estimates. G) KM survival estimates for Metroticket 2.0 criteria. H) As in G but for recurrence estimates. **Abbreviations:** AFP: alpha feto-protein, G: group, MORAL: model of recurrence after liver transplantation, Mt: metroticket. **NOTE:** P-values in panels indicate log rank global test.

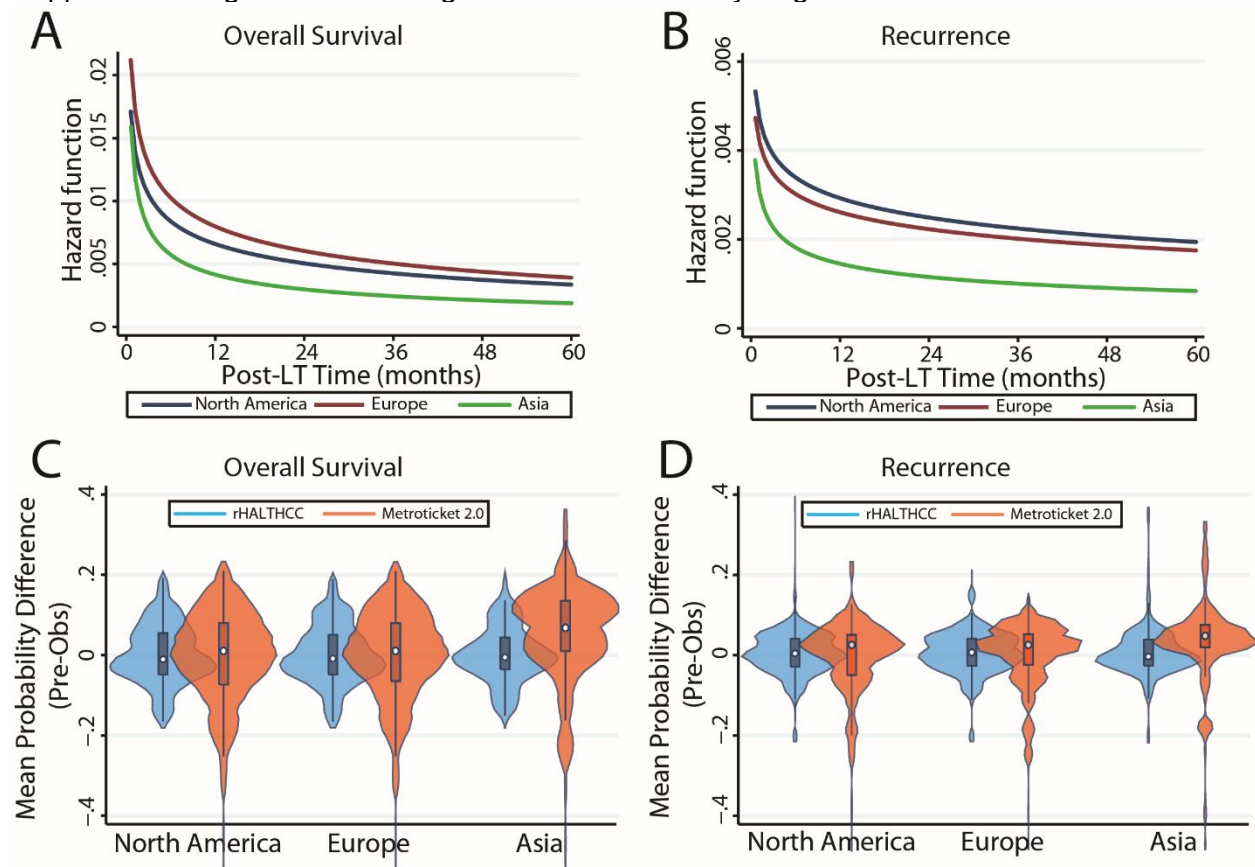
Supplementary Figure 3: Assessing HALTHCC Accuracy Using Calibration Curves



A) Calibration plot of overall survival (OS) at 1 year observed vs predicted (using original HALTHCC estimates). B) Calibration plot of OS at 1 year observed vs predicted (after HALTHCC recalibration). C) As in A but for 3-year OS. D) As in B but for 3-year OS. E) As in A but for 5-year OS. F) As in B but for 5-year OS. G) As in A but for cumulative recurrence rate. H) As in B but for cumulative recurrence rate. **Abbreviations:** obs: observed, pred: predicted, r²: coefficient of determination. **NOTE:** Displayed equations represent linear regressions of deciles of the study population for a given observed vs predicted plot (as described in the methods).  
Supplemental Table 1: Competitor Model Search and Inclusion



Supplemental Figure 4: Assessing Risk of Model Bias by Region



A) Generalized linear models are used to estimate the baseline hazard function for each of the cohorts in North America, Europe and Asia which demonstrates proportionality across all 5 years (curves do not cross). B) As in A) but for recurrence estimates. C) Calibration plot of OS by region for rHALTHCC vs Metroticket 2.0 (predicted-observed for percentiles of patients). Demonstrates deviation from normal distribution about 0 for Metroticket 2.0 but not rHALTHCC, confirming that estimates are not systematically biased. D) As in C but for recurrence.

**Abbreviations:** obs: observed, pred: predicted.

Supplemental Table 1: Competitor Model Search and Inclusion

Supplemental Table 1: Review of the Literature and Identification of Competitor Models					
Search Terms	Criteria/Model Name	Criteria/Model Overview	Citation	Included	Reason(s)
PUBMED from 1996-2018, English language articles only:	Advanced HCC selection by tumor markers	Patients stratified by PIVKA and AFP at various levels, even >10cm or 10 or more tumors had reasonable outcomes	Lee et al <sup>1</sup>	No	No PIVKA data available
Hepatocellular carcinoma	mRECIST incorporation	mRECIST responders (CR and PR) as criteria for LT	Lei et al <sup>2</sup>	No	No intention to treat population raising concerns of accuracy of estimates, mRECIST data not reliably reported by all centers
HCC					
Liver transplantation for HCC					
Hepatocellular carcinoma risk score	Metroticket 2.0	Presented as similar bins of risk within combined tumor burden and AFP levels	Mazzaferro et al <sup>3</sup>	Yes	-
Risk index for hepatocellular carcinoma	Kyushu Criteria	Maximum tumor diameter 5cm OR DCP<300mU/mL	Uchiyama et al <sup>4</sup>	No	No DCP data available
	HALTHCC Score	Continuous risk score incorporating: MELD-Na, AFP, and tumor burden	Sasaki and Firl et al <sup>5</sup>	Yes	-
Selection criteria for liver transplant and HCC	Hangzhou Criteria with PLR	Combined previously reported Hangzhou criteria with platelet lymphocyte ratio	Xia et al <sup>6</sup>	No	No platelet to lymphocyte ratio data available, single center, no separation of training the PLR cutoff from validating the score
Tumor markers in HCC					
Scoring systems for HCC					
	Columbia MORAL	Points for each of, NLR>5 (6pts), AFP>200 (4pts), and tumor size>3cm (3pts)	Halazun et al <sup>7</sup>	Yes	-
	TRAIN Score	0.988 (if mRECIST PD) + 0.838 (if AFP slope >= 15.0 ng/mL/month) + 0.452 (if NLR >= 5:0) – 0.03*WT(in months)	Lai et al <sup>8</sup>	No	Radiological response to bridging therapy not precisely recorded from all centers, delta AFP not available from all centers
	Warsaw Criteria	Combined UCSF and Up-to-seven with AFP<100	Grat et al <sup>9</sup>	No	Single center, small study, unclear methodology re combination of other criteria
	Toronto Criteria	No systemic cancer-related symptoms, no concerning pathologic features on mandatory protocol biopsy	Sapisochin et al <sup>10</sup>	No	Biopsy not routinely performed in pre-LT phase
	Korean MoRAL	Serum levels of protein induced by vitamin k absence-II (PIVKA-II) and AFP	Lee et al	No	No serum levels of PIVKA-II available.
	Delta-slope of AFP	Slope of AFP >15ng/mL/month associated with worse post-LT outcomes	Lai et al <sup>11</sup>	No	Small sample size, single center, precise longitudinal data was not available from all centers

Firl DJ et al.

Page 7

MELD-EQ	Score to equalize risk of dropout between HCC and nonHCC candidates	Marvin et al <sup>12</sup>	No	MELD without sodium not available in the present dataset, focus on dropout
Kyoto Criteria	Tumor #<10, maximal diameter 5cm, serum DCP<400	Kaido et al <sup>13</sup>	No	Single center, small sample (n=198), DCP not available (not commonly used in clinical practice)
TTV and L%	TTV<172cm <sup>3</sup> , L%>30	Li et al <sup>14</sup>	No	Single center, small sample (n=216), not validated
French AFP-Score	AFP Score<2	Duvoux et al <sup>15</sup>	Yes	-
Zhang Neural Net Model	Neural net derived, nonlinear estimates	Zhang et al <sup>16</sup>	No	Single center, small sample (n=290), complex (24 variables) and not validated
Hangzhou Criteria	Total Tumor diameter <=8, OR >8 but histopathology grade I or II and AFP<400	Zheng et al <sup>17</sup>	No	Includes post-LT factors, small sample (n=195), single center, resembles UCSF
TTV/AFP	TTV<115 and AFP<400	Toso et al <sup>18</sup>	No	Do not have all tumor sizes for some centers
Up-to-seven	Number+maximum tumor size <=7	Mazzaferro et al <sup>19</sup>	No	Updated metroticket is included for comparison
Dallas Criteria	1 up to 6 or 4 less than 5	Onaca et al <sup>20</sup>	No	Similar to UCSF, not validated
UCSF Criteria	1 up to 6.5cm or 3 less than 4.5; total tumor diameter<8	Yao et al <sup>21</sup>	Yes	-
Pittsburgh Criteria	Neural net based calculation of 9 risk factors	Marsh et al <sup>22</sup>	No	Many factors were only determinable at surgery or with pathology, ie node status or surgical margin
Milan Criteria	1 up to 5 or 3 less than 3	Mazzaferro et al <sup>23</sup>	Yes	-

Abbreviations: AFP: alpha fetoprotein, DCP: des-gamma-carboxy prothrombin

Supplemental Table 2: Comparison of international cohort to original US HALTHCC cohort			
	Derivation Cohort (n=420)	International Cohort (n=4,089)	P-value
Transplant Year	2009 [2007, 2012]	2009 [2005, 2012]	<0.0001
World Region			<0.0001
North America	420 (100)	1851 (45.3)	
Europe	0 (0)	1867 (45.7)	
Asia	0 (0)	371 (9.0)	
Sex, male	270 (64.3)	2822 (69.0)	0.008
Age, years	59 [55, 65]	58 [52, 63]	<0.0001
Underlying Liver Disease*			<0.0001
HCV	253 (60.2)	2243 (54.9)	
HBV	27 (6.4)	693 (16.9)	
EtOH	81 (19.3)	826 (20.2)	
NASH	49 (11.7)	239 (5.8)	
Other	57 (13.6)	325 (7.9)	
Listing Characteristics			
Laboratory MELD-Na, points	11.0 [7.0, 17.0]	12.0 [9.0, 16.0]	0.003
Tumor Number	1.0 [1.0, 2.0]	1.0 [1.0, 3.0]	<0.0001
Tumor Size, cm	2.3 [1.5, 3.2]	2.5 [1.9, 3.6]	<0.0001
TBS, points	3.0 [2.2, 4.0]	3.4 [2.4, 4.6]	<0.0001
AFP, IU,mL	9.0 [4.7, 39.3]	10.0 [4.7, 39.0]	0.981
Patients meeting MC HALTHCC, points	339 (80.7)	2927 (71.6)	<0.0001
	12.5 [9.6, 14.8]	12.6 [10.2, 15.8]	<0.0001
Pre-LT Characteristics			
History of bridging, %	223 (53.1)	2914 (71.3)	<0.0001
Laboratory MELD-Na, points	12.0 [8.0, 18.0]	13.0 [9.0, 17.6]	0.010
Tumor Number	1.0 [1.0, 2.0]	1.0 [1.0, 3.0]	0.756
Tumor Size, cm	2.1 [1.2, 3.0]	2.1 [1.0, 3.0]	0.484
TBS, points	2.9 [1.9, 3.8]	3.1 [1.6, 4.2]	0.148
AFP, IU,mL	8.9 [4.5, 32.9]	9.8 [4.3, 36.8]	0.933
NLR	2.9 [1.8, 4.7]	2.7 [1.7, 4.4]	0.034
Patients meeting MC HALTHCC, points	346 (82.4)	3060 (74.8)	<0.0001
	11.7 [9.3, 14.5]	12.1 [9.4, 15.6]	0.050
<b>Note:</b> Continuous variables: median [interquartile range]. *May not sum to unity due to overlapping etiologies. <b>Abbreviations:</b> AFP: preoperative alpha-fetoprotein, HALT-HCC: Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, HBV: hepatitis B virus, HCC: hepatocellular carcinoma, HCV: hepatitis C virus, MELD-Na: Model of End Stage Liver Disease Sodium, NASH: non-alcoholic steatohepatitis, SRTR: Scientific Registry of Transplant Recipients, TBS: Tumor Burden Score			

Supplemental Table 3: Descriptive Statistics by World Region				
	North America (n=1851)	Europe (n=1867)	Asia (n=371)	P-value
Transplant Year	2009 [2006, 2012]	2008 [2004, 2012]	2007 [2004, 2010]	<0.0001
Sex, male	1045 (56.5)	1572 (84.2)	205 (55.3)	<0.0001
Age, years	59 [54, 64]	57 [49, 62]	58 [53, 63]	<0.0001
Underlying Liver Disease*				<0.0001
HCV	1105 (59.7)	895 (47.9)	243 (65.5)	
HBV	263 (14.2)	343 (18.4)	87 (23.5)	
EtOH	252 (13.6)	552 (29.6)	22 (5.9)	
NASH	130 (7.0)	101 (5.4)	8 (2.2)	
Other	170 (9.2)	141 (7.6)	14 (3.8)	
Listing Characteristics				
Laboratory MELD-Na, points	12.0 [9.0, 17.0]	12.0 [9.0, 15.0]	9.0 [4.0, 15.0]	<0.0001
Tumor Number	1 [1, 2]	2 [1, 3]	2 [1, 4]	<0.0001
Tumor Size, cm	2.5 [1.8, 3.5]	2.7 [2, 4]	2 [1, 3]	<0.0001
TBS, points	3.2 [2.3, 4.4]	3.6 [2.6, 4.8]	3.4 [1.8, 4.9]	<0.0001
AFP, IU/mL	8.3 [4.1, 34.2]	10.0 [5.0, 38.0]	17.8 [6.8, 102.9]	<0.0001
Patients meeting MC	1404 (75.9)	1293 (69.3)	258 (69.5)	<0.0001
HALTHCC, points	12.0 [9.9, 15.1]	12.9 [10.4, 16.1]	12.2 [9.6, 16.7]	<0.0001
Pre-LT Characteristics				
Bridging therapy, %	1244 (67.2)	1533 (82.1)	137 (36.9)	<0.0001
Laboratory MELD-Na, points	14.0 [10.0, 20.0]	12.0 [9.0, 15.0]	12.7 [6.1, 18.6]	<0.0001
Tumor Number	1 [1, 2]	1 [1, 3]	2 [1, 4]	<0.0001
Tumor Size, cm	2.3 [1.0, 3.4]	2.0 [1.0, 3.0]	2.0 [1.5, 3.0]	0.0013
TBS, points	2.9 [1.4, 4.1]	3.2 [1.6, 4.2]	3.4 [2.3, 5.3]	<0.0001
AFP, IU/mL	8.3 [4.1, 34.2]	9.8 [4.4, 32.3]	22.3 [7.1, 119.7]	<0.0001
NLR	2.7 [1.8, 4.5]	2.8 [1.9, 4.2]	1.8 [0.5, 3.7]	<0.0001
Patients meeting MC	1350 (72.9)	1462 (78.3)	248 (66.8)	<0.0001
HALTHCC, points	12.2 [9.5, 15.5]	11.7 [8.9, 15.1]	14.4 [10.9, 19.2]	<0.0001

**Note:** Continuous variables: median [interquartile range]. \*May not sum to unity due to overlapping etiologies. **Abbreviations:** AFP: preoperative alpha-fetoprotein, HALT-HCC: Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, HBV: hepatitis B virus, HCC: hepatocellular carcinoma, HCV: hepatitis C virus, MELD-Na: Model of End Stage Liver Disease Sodium, NASH: non-alcoholic steatohepatitis, SRTR: Scientific Registry of Transplant Recipients, TBS: Tumor Burden Score

Supplemental Table 4: Model comparisons for OS (Harrell's C, NRI, IDI, AIC) using Validation set					
	Harrells C	NRI	IDI	AIC	P-value
Milan Criteria	0.55 (0.53-0.56)	GS	GS	14651	<0.0001
UCSF Criteria	0.54 (0.53-0.56)	0.167 (0.123-0.211)	0.008 (0.003-0.015)	14639	<0.0001
MORAL-Score	0.56 (0.53-0.58)	0.235 (0.117-0.305)	0.013 (0.006-0.023)	9301#	<0.0001
AFP-Score	0.56 (0.54-0.58)	0.173 (0.124-0.223)	0.008 (0.003-0.014)	14701	<0.0001
Metroticket 2.0	0.57 (0.55-0.58)	0.260 (0.201-0.315)	0.016 (0.009-0.025)	14401	<0.0001
oHALTHCC	0.62 (0.60-0.64)	0.284 (0.212-0.347)	0.036 (0.025-0.049)	14365	0.0511
rHALTHCC	0.63 (0.61-0.65)	0.330 (0.261-0.391)	0.043 (0.031-0.057)	14332	-
Note: %For Metroticket 2.0 above AP indicates as presented, meaning bins of risk whereas RE indicates raw estimates in continuous form. Abbreviations: AIC: Akaike information criterion, AFP: alpha feto-protein, GS: Gold standard, IDI: Integrated discrimination improvement, MORAL: Model Of Recurrence After Liver Transplantation, NRI: Net reclassification improvement, oHALT-HCC: Original Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, rHALT-HCC: Recalibrated Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, UCSF: University of California San Francisco					

Supplemental Table 5: Model comparisons for Recurrence (Harrell's C, NRI, IDI, AIC) using Validation set

	Harrells C	NRI	IDI	AIC	P-value
Milan Criteria	0.61 (0.59-0.64)	GS	GS	6283	<0.0001
UCSF Criteria	0.58 (0.56-0.60)	0.368 (0.292-0.441)	0.013 (0.005-0.023)	6278	<0.0001
MORAL-Score	0.61 (0.58-0.65)	0.174 (0.102-0.403)	0.017 (0.007-0.030)	4184#	<0.0001
AFP-Score	0.64 (0.61-0.67)	0.341 (0.255-0.419)	0.012 (0.005-0.022)	6148	<0.0001
Metroticket 2.0	0.65 (0.62-0.67)	0.540 (0.439-0.639)	0.033 (0.021-0.049)	6108	<0.0001
oHALTHCC	0.69 (0.66-0.72)	0.429 (0.340-0.519)	0.045 (0.030-0.065)	6110	0.0003
rHALTHCC	0.71 (0.68-0.74)	0.518 (0.419-0.646)	0.055 (0.036-0.078)	6072	-

Note: %For Metroticket 2.0 above AP indicates as presented, meaning bins of risk whereas RE indicates raw estimates in continuous form. Abbreviations: AIC: Akaike information criterion, AFP: alpha feto-protein, GS: Gold standard, IDI: Integrated discrimination improvement, MORAL: Model Of Recurrence After Liver Transplantation, NRI: Net reclassification improvement, oHALT-HCC: Original Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, rHALT-HCC: Recalibrated Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, UCSF: University of California San Francisco

Supplemental Table 6: Model comparisons for HCC-related mortality (Harrell's C, NRI, IDI, AIC) using Validation set

	Harrells C	NRI	IDI	AIC	P-value
Milan Criteria	0.62 (0.59-0.64)	GS	GS	4815	<0.0001
UCSF Criteria	0.60 (0.57-0.62)	0.411 (0.323-0.503)	0.014 (0.006-0.026)	4806	<0.0001
MORAL-Score	0.62 (0.58-0.66)	0.167 (0.070-0.336)	0.019 (0.008-0.034)	3246	<0.0001
AFP-Score	0.64 (0.61-0.68)	0.375 (0.276-0.471)	0.012 (0.005-0.023)	4716	<0.0001
Metroticket 2.0	0.66 (0.63-0.69)	0.659 (0.556-0.750)	0.034 (0.021-0.050)	4649	<0.0001
oHALTHCC	0.72 (0.69-0.75)	0.471 (0.370-0.570)	0.059 (0.038-0.079)	4629	0.003
rHALTHCC	0.74 (0.71-0.77)	0.546 (0.435-0.655)	0.066 (0.047-0.088)	4598	-

Note: %For Metroticket 2.0 above AP indicates as presented, meaning bins of risk whereas RE indicates raw estimates in continuous form. Abbreviations: AIC: Akaike information criterion, AFP: alpha feto-protein, GS: Gold standard, IDI: Integrated discrimination improvement, MORAL: Model Of Recurrence After Liver Transplantation, NRI: Net reclassification improvement, oHALT-HCC: Original Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, rHALT-HCC: Recalibrated Hazard Associated with Liver transplantation for Hepatocellular Carcinoma, UCSF: University of California San Francisco



Supplemental Table 7: Validation Set Stratified by MC and rHALTHCC

	rHALTHCC ( $<10$ points)	rHALTHCC (10-15 points)	rHALTHCC (15-20 points)	rHALTHCC (20-25 points )	rHALTHCC ( $>25$ points)	Total
Milan- in	950 (79.2 [76.2-81.9])	934 (70.3 [66.9-73.5])	316 (66.1 [60.0-71.5])	71 (48.8 [35.0- 61.2])	20 (35.6 [14.6-57.3])	2291 (72.3% [70.2- 74.2])
Milan- out	110 (77.5 [65.7-85.6])	245 (73.5 [66.9-79.1])	214 (60.4 [52.8-67.1])	104 (52.3 [41.0-62.4])	90 (41.6 [31.0-51.9])	763 (62.8% [58.9- 66.5])
Total	1060(79.1 [76.2-81.7])	1179 (71.0 [68.0-73.7])	530 (63.7 [59.0-68.0])	175 (50.9 [42.3-59.0])	110 (40.4 [30.8-50.0])	

**Note: ( ) Denotes** 5-year overall survival by the Kaplan-Meier method with 95% confidence intervals.

**Abbreviations:** MC: Milan Criteria, rHALTHCC: recalibrated Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma

Supplemental Table 8: Ordinal Logistic Model for Recalibrating HALTHCC using Training Set No Asia

(n=928)	Coefficient	Standard error	Wald test	Odds Ratio (95% CI)	P-value*
MELD-NA, per point	0.0245	0.0091	2.70	1.025 (1.007-1.043)	0.007
TBS, pre point	0.1332	0.0265	5.03	1.142 (1.085-1.203)	$<0.0001$
AFP, per ln(IU/mL)	0.2337	0.0381	6.13	1.263 (1.172-1.361)	$<0.0001$
World Region					
North America	-Base case-	-	-	-	-
Europe	0.1817	0.141	1.29	1.199 (0.909-1.582)	0.198

**\*Based** on likelihood est adjusted for the other factors in the final model.

**Abbreviations:** AFP: preoperative alpha-fetoprotein (natural log transformed), HALTHCC: Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma, MELD-Na: Model of End Stage Liver Disease Sodium, TBS: Tumor burden score

Supplemental Table 9: Ordinal Logistic Model for Recalibrating HALTHCC using Training Set ASIA ALONE

(n=93)	Coefficient	Standard error	Wald test	Odds Ratio (95% CI)	P-value*
MELD-NA, per point	0.0272	0.0330	0.83	1.028 (0.963-1.096)	0.409
TBS, pre point	0.1344	0.0504	2.67	1.144 (1.036-1.263)	0.008
AFP, per ln(IU/mL)	0.2093	0.1034	2.02	1.233 (1.007-1.510)	0.043

**\*Based** on likelihood est adjusted for the other factors in the final model.

**Abbreviations:** AFP: preoperative alpha-fetoprotein (natural log transformed), HALTHCC: Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma, MELD-Na: Model of End Stage Liver Disease Sodium, TBS: Tumor burden score

Supplemental Table 10: Sensitivity Analysis of Harrell's C for Outcomes of Interest with Asian Centers Excluded

	OS - Harrells C	Recurrence – Harrells C	HCC Related Mortality Harrells C
Milan Criteria	0.54 (0.52-0.55)***	0.62 (0.59-0.64)***	0.62 (0.59-0.65)***
UCSF Criteria	0.54 (0.52-0.55)***	0.58 (0.56-0.60)***	0.58 (0.56-0.61)***
MORAL-Score	0.56 (0.54-0.58)***	0.60 (0.57-0.63)***	0.61 (0.57-0.64)***
AFP-Score	0.54 (0.53-0.55)***	0.58 (0.56-0.60)***	0.59 (0.56-0.61)***
Metroticket 2.0	0.56 (0.55-0.58)***	0.64 (0.62-0.66)***	0.66 (0.63-0.68)***
oHALTHCC	0.62 (0.60-0.64)NS	0.68 (0.66-0.70)**	0.70 (0.67-0.72)**
rHALTHCC ALL REG	0.62 (0.60-0.64)NS	0.69 (0.66-0.72)NS	0.71 (0.69-0.74)*
rHALTHCC w/NO ASIA	0.62 (0.60-0.64)	0.69 (0.66-0.72)	0.72 (0.69-0.75)

NOTE: \*,P<0.05;\*\*P<0.01;\*\*\*P<0.001;NS,P>0.05; for comparisons between rHALTHCC w/NO ASIA and each other model.

References:

1. Lee, H. W., Song, G.-W., Lee, S.-G., Kim, J. M., Joh, J.-W., Han, D. H., Kim, S. I., Kim, S. H., Kim, D.-S., Cho, J. Y. & Suh, K.-S. Patient selection by tumor markers in liver transplantation for advanced hepatocellular carcinoma. *Liver Transpl.* (2018). doi:10.1002/lt.25056

2. Lei, J., Zhong, J., Luo, Y., Yan, L., Zhu, J., Wang, W., Li, B., Wen, T. & Group, J. Y. L. S. Response to transarterial chemoembolization may serve as selection criteria for hepatocellular carcinoma liver transplantation. *Oncotarget* **8**, (2017).

3. Mazzaferro, V., Sposito, C., Zhou, J., Pinna, A. D., De Carlis, L., Fan, J., Cescon, M., Di Sandro, S., Yi-Feng, H., Lauterio, A., Bongini, M. & Cucchetti, A. Metroticket 2.0 Model for Analysis of Competing Risks of Death Following Liver Transplantation for Hepatocellular Carcinoma. *Gastroenterology* (2017). doi:10.1053/j.gastro.2017.09.025

4. Uchiyama, H., Itoh, S., Yoshizumi, T., Ikegami, T., Harimoto, N., Soejima, Y., Harada, N., Morita, K., Toshima, T., Motomura, T. & Maehara, Y. Living donor liver transplantation for hepatocellular

- carcinoma: results of prospective patient selection by Kyushu University Criteria in 7 years. *HPB* **19**, 1082–1090 (2017).
5. Sasaki, K., Firl, D. J., Hashimoto, K., Fujiki, M., Diago-Uso, T., Quintini, C., Egtesad, B., Fung, J. J., Aucejo, F. N. & Miller, C. M. Development and validation of the HALT-HCC score to predict mortality in liver transplant recipients with hepatocellular carcinoma: a retrospective cohort analysis. *Lancet Gastroenterol. Hepatol.* **2**, 595–603 (2017).
6. Xia, W., Ke, Q., Guo, H., Wang, W., Zhang, M., Shen, Y., Wu, J., Xu, X., Yan, S., Yu, J., Zhang, M. & Zheng, S. Expansion of the Milan criteria without any sacrifice: combination of the Hangzhou criteria with the pre-transplant platelet-to-lymphocyte ratio. *BMC Cancer* **17**, (2017).
7. Halazun, K. J., Najjar, M., Abdelmessih, R. M., Samstein, B., Griesemer, A. D., Guarrera, J. V., Kato, T., Verna, E. C., Emond, J. C. & Brown, R. S. Recurrence After Liver Transplantation for Hepatocellular Carcinoma: A New MORAL to the Story. *Ann. Surg.* **1** (2016). doi:10.1097/SLA.0000000000001966
8. Lai, Q., Nicolini, D., Inostroza Nunez, M., Iesari, S., Goffette, P., Agostini, A., Giovagnoni, A., Vivarelli, M. & Lerut, J. A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) Score. *Ann. Surg.* **264**, 787–796 (2016).
9. Grąt, M., Wronka, K. M., Stypułkowski, J., Bik, E., Krasnodębski, M., Masior, Ł., Lewandowski, Z., Grąt, K., Patkowski, W. & Krawczyk, M. The Warsaw Proposal for the Use of Extended Selection Criteria in Liver Transplantation for Hepatocellular Cancer. *Ann. Surg. Oncol.* **24**, 526–534 (2017).
10. Sapisochin, G., Goldaracena, N., Laurence, J. M., Dib, M., Barbas, A., Ghanekar, A., Cleary, S. P., Lilly, L., Cattal, M. S., Marquez, M., Selzner, M., Renner, E., Selzner, N., McGilvray, I. D., Greig, P. D. & Grant, D. R. The extended Toronto criteria for liver transplantation in patients with hepatocellular carcinoma: A prospective validation study. *Hepatology* **64**, 2077–2088 (2016).

11. Lai, Q., Inostroza, M., Rico Juri, J. M., Goffette, P. & Lerut, J. Delta-slope of alpha-fetoprotein improves the ability to select liver transplant patients with hepatocellular cancer. *HPB* **17**, 1085–1095 (2015).

12. Marvin, M. R., Ferguson, N., Cannon, R. M., Jones, C. M. & Brock, G. N. MELDEQ : An alternative Model for End-Stage Liver Disease score for patients with hepatocellular carcinoma. *Liver Transplant. Off. Publ. Am. Assoc. Study Liver Dis. Int. Liver Transplant. Soc.* **21**, 612–622 (2015).

13. Kaido, T., Ogawa, K., Mori, A., Fujimoto, Y., Ito, T., Tomiyama, K., Takada, Y. & Uemoto, S. Usefulness of the Kyoto criteria as expanded selection criteria for liver transplantation for hepatocellular carcinoma. *Surgery* **154**, 1053–1060 (2013).

14. Li, C., Wen, T.-F., Yan, L.-N., Li, B., Yang, J.-Y., Xu, M.-Q., Wang, W.-T. & Wei, Y.-G. Scoring Selection Criteria Including Total Tumour Volume and Pretransplant Percentage of Lymphocytes to Predict Recurrence of Hepatocellular Carcinoma after Liver Transplantation. *PLoS ONE* **8**, e72235 (2013).

15. Duvoux, C., Roudot–Thoraval, F., Decaens, T., Pessione, F., Badran, H., Piardi, T., Francoz, C., Compagnon, P., Vanlemmens, C., Dumortier, J., Dharancy, S., Gugenheim, J., Bernard, P., Adam, R., Radenne, S., Muscari, F., Conti, F., Hardwigsen, J., Pageaux, G., Chazouillères, O., Salame, E., Hilleret, M., Lebray, P., Abergel, A., Debette–Gratien, M., Kluger, M. D., Mallat, A., Azoulay, D. & Cherqui, D. Liver Transplantation for Hepatocellular Carcinoma: A Model Including  $\alpha$ -Fetoprotein Improves the Performance of Milan Criteria. *Gastroenterology* **143**, 986-994.e3 (2012).

16. Zhang, M., Yin, F., Chen, B., Li, B., Li, Y. P., Yan, L. N. & Wen, T. F. Mortality risk after liver transplantation in hepatocellular carcinoma recipients: A nonlinear predictive model. *Surgery* **151**, 889–897 (2012).

17. Zheng, S.-S., Xu, X., Wu, J., Chen, J., Wang, W.-L., Zhang, M., Liang, T.-B. & Wu, L.-M. Liver Transplantation for Hepatocellular Carcinoma: Hangzhou Experiences: *Transplantation* **85**, 1726–1732 (2008).

18. Toso, C., Asthana, S., Bigam, D. L., Shapiro, A. M. J. & Kneteman, N. M. Reassessing selection criteria prior to liver transplantation for hepatocellular carcinoma utilizing the scientific registry of transplant recipients database. *Hepatology* **49**, 832–838 (2009).
19. Mazzaferro, V., Llovet, J. M., Miceli, R., Bhoori, S., Schiavo, M., Mariani, L., Camerini, T., Roayaie, S., Schwartz, M. E., Grazi, G. L., Adam, R., Neuhaus, P., Salizzoni, M., Bruix, J., Forner, A., De Carlis, L., Cillo, U., Burroughs, A. K., Troisi, R., Rossi, M., Gerunda, G. E., Lerut, J., Belghiti, J., Boin, I., Gugenheim, J., Rochling, F., Van Hoek, B. & Majno, P. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol.* **10**, 35–43 (2009).
20. Onaca, N., Davis, G. L., Goldstein, R. M., Jennings, L. W. & Klintmalm, G. B. Expanded criteria for liver transplantation in patients with hepatocellular carcinoma: A report from the International Registry of Hepatic Tumors in Liver Transplantation. *Liver Transpl.* **13**, 391–399 (2007).
21. Yao, F. Liver transplantation for hepatocellular carcinoma: Expansion of the tumor size limits does not adversely impact survival. *Hepatology* **33**, 1394–1403 (2001).
22. Marsh, J. W., Dvorchik, I., Subotin, M., Balan, V., Rakela, J., Popechitelev, E. P., Subbotin, V., Casavilla, A., Carr, B. I., Fung, J. J. & Iwatsuki, S. The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation: A pilot study. *Hepatology* **26**, 444–450 (1997).
23. Mazzaferro, V., Regalia, E., Doci, R., Andreola, S., Pulvirenti, A., Bozzetti, F., Montalto, F., Ammatuna, M., Morabito, A. & Gennari, L. Liver Transplantation for the Treatment of Small Hepatocellular Carcinomas in Patients with Cirrhosis. *N. Engl. J. Med.* **334**, 693–700 (1996).