JAMA Surgery | Original Investigation

Dynamic α-Fetoprotein Response and Outcomes After Liver Transplant for Hepatocellular Carcinoma

Karim J. Halazun, MD; Russell E. Rosenblatt, MD, MS; Neil Mehta, MD; Quirino Lai, MD; Kaveh Hajifathalian, MD, PhD; Andre Gorgen, MD; Gagan Brar, MD; Kazunari Sasaki, MD; Maria B. Majella Doyle, MD; Parissa Tabrizian, MD; Vatche G. Agopian, MD; Marc Najjar, MD; Tommy Ivanics, MD; Benjamin Samstein, MD; Robert S. Brown Jr, MD, MPH; Jean C. Emond, MD; Francis Yao, MD; Jan Lerut, MD, PhD; Massimo Rossi, MD; Gianluca Mennini, MD; Samuele Iesari, MD; Armin Finkenstedt, MD; Benedikt Schaefer, MD; Jans Mittler, MD; Maria Hoppe-Lotichius, MD; Cristiano Quintini, MD; Federico Aucejo, MD; William Chapman, MD; Gonzalo Sapisochin, MD

IMPORTANCE Accurate preoperative prediction of hepatocellular carcinoma (HCC) recurrence after liver transplant is the mainstay of selection tools used by transplant-governing bodies to discern candidacy for patients with HCC. Although progress has been made, few tools incorporate objective measures of tumor biological characteristics, resulting in inclusion of patients with high recurrence rates and exclusion of others who could otherwise be cured.

OBJECTIVE To externally validate the New York/California (NYCA) score, a recently published multi-institutional US HCC selection tool that was the first model incorporating a dynamic α-fetoprotein response (AFP-R) and compare the validated score with currently accepted HCC selection tools, namely, the Milan Criteria (MC), the French-AFP (F-AFP), and Metroticket 2.0 models.

DESIGN, SETTING, AND PARTICIPANTS A retrospective, multicenter prognostic analysis of prospectively collected databases of 2236 adults undergoing liver transplant for HCC was conducted at 3 US, 1 Canadian, and 4 European centers from January 1, 2001, to December 31, 2013. The AFP-R was measured as the difference between maximum and final pre-liver transplant AFP level. Cox proportional hazards regression and competing risk regression analyses examined recurrence-free and overall survival. Receiver operating characteristic analyses and net reclassification index were used to compare NYCA with MC, F-AFP, and Metroticket 2.0. Data analysis was performed from June 2019 to April 2020.

MAIN OUTCOMES AND MEASURES The primary study outcome was 5-year recurrence-free survival; overall survival was the secondary outcome.

RESULTS Of 2236 patients, 1808 (80.9%) were men; mean (SD) age was 58.3 (7.96) years. A total of 545 patients (24.4%) did not meet the MC. The NYCA score proved valid on competing risk regression analysis, accurately predicting recurrence-free and overall survival (5-year cumulative incidence of recurrence risk in NYCA risk categories was 9.5% for low-, 20.5%, for acceptable-, and 40.5% for high-risk categories; $P < .001$ for all). The NYCA also predicted recurrence-free survival on a center-specific level: 453 of 545 patients (83.1%) who did not meet MC, 213 of 308 (69.2%) who did not meet the French-AFP, 292 of 384 (76.1%) who did not meet Metroticket 2.0 would be recategorized into NYCA low- and acceptable-risk groups (>75% 5-year recurrence-free survival). The Harrell C statistic for the validated NYCA score was 0.66 compared with 0.59 for the MC and 0.57 for the F-AFP models ($P < .001$). The net reclassification index for NYCA was 8.1 vs MC, 12.9 vs F-AFP, and 10.1 vs Metroticket 2.0.

CONCLUSIONS AND RELEVANCE This study appears to externally validate the importance of AFP-R in the selection of patients with HCC for liver transplant. The AFP-R represents one of the truly objective measures of biological characteristics available before transplantation. Incorporation of AFP-R into selection criteria allows safe expansion of MC and other models, offering liver transplant to patients with acceptable tumor biological characteristics who would otherwise be denied potential cure.

JAMA Surg. doi:10.1001/jamasurg.2021.0954
Published online May 5, 2021.

Invited Commentary
Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Karim J. Halazun, MD, Division of Liver Transplantation and Hepatobiliary Surgery, Department of Surgery, Weill Cornell Medicine, 525 E 68th, F-763, New York, NY 10065 (kah7007@med.cornell.edu).

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

The adoption of the Milan Criteria (MC) for selecting patients with hepatocellular carcinoma (HCC) eligible for liver transplant has been the most robust intervention in HCC outcomes, improving survival for thousands of patients.[1,2] However, criticisms levied against the MC's dichotomous nature, restrictiveness, and lack of tumor biological indexes resulted in the emergence of several scores to improve patient selection.[3-9] Other scores have been validated and shown to be superior to the MC at predicting outcomes,[8,10,11] yet widespread adoption has lagged.

Some European centers have abandoned use of the MC, instead adopting criteria that include the α-fetoprotein (AFP) level, therefore adding valuable tumor biological information. The French-AFP (F-AFP) score, validated both in France and Italy,[8,10] is a 2-tiered model that has been applied in France in lieu of the MC. Similarly, the Metroticket 2.0 expands on the original up-to-7 criteria to include AFP levels.[4,11] Both tools supersede the MC's ability to predict outcomes and increase inclusiveness; however, both are reliant on static AFP levels. Recently, the New York/California (NYCA) score, which includes the dynamic AFP response (AFP-R),[12] was proposed by 3 US centers. This score was the first to incorporate both an objective AFP-R and use competing risk regression to assess the association between AFP-R and recurrence-free survival (RFS). The NYCA score was superior to the MC and F-AFP scores, increasing candidacy for liver transplant in patients beyond both sets of criteria without compromising RFS, and was a better predictor of overall survival (OS). Herein, we describe our external validation of the NYCA tool in a large international cohort of patients with HCC.

Key Points

Question Can the dynamic α-fetoprotein response predict outcomes of patients undergoing liver transplant for hepatocellular carcinoma in a heterogeneous population of patients from North America and Europe?

Findings In this multicenter international prognostic study of 2236 adults undergoing liver transplant for hepatocellular carcinoma at 4 North American and 4 European centers, the utility of the α-fetoprotein response was validated to predict recurrence and survival. In addition, the α-fetoprotein response was superior to currently available patient selection tools.

Meaning Incorporation of a dynamic α-fetoprotein response into currently used morphometric hepatocellular carcinoma selection criteria appears to improve the ability to predict recurrence and may increase access to liver transplant for patients with hepatocellular carcinoma who would otherwise be denied potential cure.

Methods

Study Population

Prospectively maintained databases of adults undergoing liver transplant for HCC from January 1, 2001, to December 31, 2013, at 4 North American and 4 European centers were analyzed. The participating North American centers were Toronto General Hospital (n = 461); University of California, San Francisco (n = 457); Cleveland Clinic (n = 384); and Washington University at St Louis (n = 315). The participating European centers were University of Innsbruck, Innsbruck, Austria (n = 197); Université Catholique de Louvain, Brussels, Belgium (n = 169); Goethe University (Johannes Gutenberg University Mainz), Mainz, Germany (n = 138); and Sapienza University, Rome, Italy (n = 115). The European Hepatocellular Cancer Liver Transplant Study Group provided data from European centers. Because the largest European center had less than 200 patients, and 3 of the 4 European centers had less than 50% of the patients of the largest North American center, for the analysis, the European centers were considered a single data set and the North American centers were analyzed individually. Data analyzed included demographic, donor-specific, laboratory, and tumor specific data. Information on pretransplant locoregional therapy use and explant pathologic data were also collected. Three patients with missing initial AFP values were excluded. Institutional review board approval to conduct the study was obtained in each center with waiver of informed consent because data were collected prospectively by each center.

AFP Response

As with the original NYCA score,[12] AFP levels at diagnosis (ie, the first AFP), maximum AFP at any point, and the immediate pretransplant AFP (ie, the final AFP) were recorded. The final AFP level was considered the last available AFP before the transplant episode and was not measured on the day of transplant. These points were chosen to validate their utility and standardize AFP measurements. First AFP provides a baseline measure and is used in the F-AFP scale to assess candidacy, the maximum AFP provides the best indicator of maximum tumor burden, and final AFP is essential in making decisions regarding transplant candidacy and is used by Metroticket 2.0. α-Fetoprotein levels greater than 200 ng/mL (to convert to micrograms per liter, multiply by 1) have previously been established as our cutoff.[9] This level is corroborated with correlations of AFP levels greater than 200 ng/mL to AFP messenger RNA expression in both HCC and healthy liver tissue, with resultant poor prognosis.[13,14] Levels of AFP increased to greater than 1000 ng/mL were used for categorization of marked increases based on previous literature,[15] with the F-AFP, Ontario Canada regional regulation, and the United Network for Organ Sharing regulations all using AFP greater than 1000 ng/mL as their upper limit.[8,10,16] Extreme AFP increase was considered when the maximum AFP level peaked at greater than 10 000 ng/mL. This extra cutoff was considered because 21 patients with a maximum AFP level 10 000 ng/mL were included, whereas in the original NYCA study, a few patients had AFP levels higher than 10 000 ng/mL. The originally established AFP-R (eFigure 1 in the Supplement) was validated using Cox proportional hazards regression and used for calculation of the NYCA score in this study.

NYCA Score Validation

The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

guideline statement was used to report study elements (TRIPOD V).[17] The primary outcome was 5-year RFS. Patients who had perioperative deaths within 90 days of liver transplant were therefore excluded. Recurrence was evaluated by imaging and biopsy if necessary. The NYCA score was calculated for all patients according to the previous model[12]; patients were assigned points accordingly and then risk-stratified into the 3 groups (**Table 1**). Competing risk regression was then performed for RFS, with death before recurrence as a competing risk. Once the NYCA score was validated for the entire cohort, center-level validation was conducted. Receiver operating characteristic curves were used to compare validated NYCA score with MC, F-AFP, and Metroticket 2.0. Net reclassification index was performed to evaluate whether the NYCA score would more accurately classify patients as high risk than previous models. The net reclassification index is a statistical test that compares whether risk scores more appropriately categorize patients as high risk depending on their outcome. The net reclassification index can be superior to the area under the curve, especially in cases where cutoffs are not strictly defined.[18] The net reclassification index was used to assess the utility of the NYCA score compared with the other models at predicting recurrence. A net reclassification of patients into the correct risk category compared with other prediction models (ie, reclassifying patients with an eventual recurrence into a higher-risk group or reclassifying patients without recurrence into a lower-risk group) would have resulted in a positive (favorable) net reclassification index for NYCA. The NYCA score was also assessed for its ability to predict OS using Cox regression to investigate factors associated with OS. The statistical analyses were performed with SPSS, version 26 (IBM Corp) and Stata, version 14.2 for Macintosh (StataCorp LLC). With 2-sided testing, $P = .05$ was considered statistically significant.

Table 1. Simplified NYCA Score Components and Point Allocation and Recurrence Risk Stratification According to NYCA Score

Factors affecting 5-y RFS	NYCA score
Maximum tumor size at diagnosis, cm	
0-3 (Reference)	0
>3-6	2
>6	4
Maximum tumor No. at diagnosis	
1 (Reference)	0
2-3	2
≥4	4
AFP response (maximum to final AFP level)	
AFP always <200 ng/mL	0
Responders	
Maximum >200-1000 to final <200 ng/mL	2
Maximum >1000 to final <1000 ng/mL (must be >50% decrease)	2
Nonresponders	
Maximum >200-400 to final >200 ng/mL	3
Maximum >400-1000 to final >200 ng/mL	4
Maximum >1000 to final >1000 ng/mL	6
Recurrence risk, NYCA score	
Low	0-2
Acceptable	3-6
High	≥7s

Abbreviations: AFP, α-fetoprotein; NYCA, New York/California; RFS, recurrence-free survival.

SI conversion factor: To convert AFP to micrograms per liter, multiply by 1.

Results

Patient Characteristics

A total of 2236 patients were included in the analyses, including 1808 men (80.9%) and 428 women (19.1%) (mean [SD] age, 58.3 [7.96] years). A total of 545 patients (24.4%) did not meet the MC at diagnosis. The baseline characteristics of the entire study population and differences between centers are given in **Table 2** and eTable 1 in the Supplement; demographic characteristics of individual European centers are reported in eTable 2 in the Supplement. The diagnosis of hepatitis C dominated in North America (972 of 1617 [60.1%]), whereas in Europe, alcohol-related liver disease was most common (230 of 619 [37.2%]) ($P < .001$). Significant differences in listing, locoregional therapy use, and transplant practices were observed. European centers (31%) and Toronto (32.8%) had the highest percentages of patients who did not meet MC. The University of California, San Francisco used locoregional therapy most aggressively (96.5% of patients), likely because of having the longest waiting times (median, 9.4 months). In contrast, only 58.5% of the patients received locoregional therapy in Cleveland, likely resulting from a median waiting time of 2 months. Recurrence rates ranged from 10.8% to 21%. No center had a 5-year RFS rate of less than 80%.

AFP-R Validation and Calibration

Significantly different recurrence risks over time were found, with maximum AFP level increases from 0 to 200, 200 to 1000, and greater than 1000 ng/mL (eFigure 2A in the Supplement). A second threshold analysis performed to examine whether patients with maximum AFP levels greater than 10 000 ng/mL represented a distinct survival category showed these patients to have significantly lower survival rates compared with those whose AFP levels were 1000 to 9999 ng/mL (eFigure 2B in the Supplement). The response for these patients with the extreme increase was therefore calibrated, requiring normalization of AFP levels rather than a reduction to less than 1000 ng/mL. Given the variability of values within reference ranges throughout the institutions, less than 30 ng/mL was considered to be the highest reference range value. **Figure 1** shows the progressive increase in recurrence risk with blunting of the AFP-R. As with the original NYCA score, patients with maximum AFP levels greater than 1000 ng/mL with more than 50% response and those with maximum AFP levels greater than 10 000 ng/mL with normalization of the final AFP value had RFS that did not differ significantly from those with maximum AFP levels of 200 to 1000 ng/mL with a final AFP level less than 200 ng/mL. For patients with AFP levels greater than 200 ng/mL, we found that the AFP starting point did not predict response rates (eTable 3 in the Supplement).

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

Table 2. Demographic, Clinicopathologic, Tumor, and Recurrence Data for Patients

Variable	European centers (n = 619)	Cleveland Clinic (n = 384)	Toronto General Hospital (n = 461)	UCSF (n = 457)	Washington University at St Louis (n = 315)	*P* value
Preoperatively available data (n = 2236)						
Patient demographic characteristic						
Men, No. (%)	516 (83.4)	304 (79.2)	385 (83.5)	356 (77.9)	247 (78.4)	.06
Women, No. (%)	103 (16.6)	80 (20.8)	76 (16.5)	101 (22.1)	68 (21.6)	
Age, mean (SD), y	58.3 (8.4)	59.4 (8.3)	57.2 (7.8)	59.1 (7.3)	57.4 (7.4)	<.001
Laboratory MELD at treatment, median (IQR)	11 (8-15)	11 (8-15)	11 (8-14)	10 (8-13)	14 (9-18)	<.001
Waiting time, median (IQR), mo	5.7 (2.6-10)	2.0 (0.8-5.1)	5.7 (2.7-11)	9.4 (5.9-14)	2.8 (0.8-5.8)	<.001
Diagnosis, %						
HCV	230 (37.2)	239 (62.2)	236 (51.2)	278 (60.8)	209 (66.3)	<.001
HBV	81 (13.1)	21 (5.5)	113 (24.5)	122 (26.7)	23 (7.3)	<.001
Alcohol-related liver disease	246 (39.7)	75 (19.5)	61 (13.2)	22 (4.8)	48 (15.2)	<.001
Cryptogenic/NASH	48 (7.8)	48 (12.5)	30 (6.5)	26 (5.7)	46 (14.6)	<.001
Other (PBC, PSC, AIH, hemochromatosis)	55 (8.9)	31 (8.1)	21 (4.6)	9 (2.0)	21 (6.7)	<.001
Tumor characteristics						
Size of largest tumor, mean (SD), cm	2.89 (2.01)	2.47 (1.54)	3.19 (2.06)	3.07 (1.19)	2.54 (1.79)	.001
Tumor size (largest mass) >3cm	31.7	28.4	40.3	36.5	29.9	.003
Multifocal HCC	48.6	39.6	47.1	29.2	36.5	<.001
T1 category tumors at diagnosis	173 (27.9)	35 (9.1)	78 (17)	1 (0.2)	26 (8.1)	<.001
AFP level at diagnosis, median (IQR), ng/mL	8.4 (4.5-28.5)	10 (5.2-38.4)	10 (5-39)	16 (5.7-77)	9.3 (5-27.4)	<.001
Maximum AFP level, median (IQR), ng/mL	12.0 (5.0-40.0)	21.8 (8.8-111.9)	16 (6-71.2)	24.5 (7.1-163)	12.8 (5.6-67.4)	<.001
Final AFP level, median (IQR), ng/mL	7.0 (3.9-18.5)	10.6 (5.6-39.8)	9 (5-38)	8.7 (4-37.3)	7.8 (5-27.1)	.023
Not meeting MC at diagnosis, No. (%)	192 (31.0)	84 (21.9)	151 (32.8)	66 (14.4)	52 (16.5)	<.001
French-AFP score >2, No. (%)	104 (16.8)	37 (9.6)	75 (16.3)	59 (12.9)	24 (10.8)	.007
Receiving LRT, No. (%)	545 (88.0)	224 (58.3)	75 (63.6)	441 (96.5)	229 (72.7)	<.001
Recurrence/survival data						
Recurrence rates, No. (%)	87 (14.1)	62 (16.1)	97 (21)	56 (12.3)	34 (10.8)	<.001
Recurrence-free survival						
1 y	96	94	93	96	96	ST[a]
3 y	89	87	84	89	93	
5 y	85	83	81	88	89	
Overall recurrence-free survival						
1 y	93	94	95	96	95	ST[a]
3 y	85	83	81	87	85	
5 y	77	75	73	81	76	
Time from recurrence to death, median (IQR), mo	12.5 (3.5-26.7)	10.1 (5.0-18.1)	15.2 (7.4-23.8)	13.7 (6.3-23.9)	14.6 (5.0-2)7.6	.57

Abbreviations: AFP, α-fetoprotein; AIH, autoimmune hepatitis; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; IQR, interquartile range; LRT, locoregional therapy; MC, Milan Criteria; MELD, model of end-stage liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; ST, supplement table; UCSF, University of California, San Francisco.

SI conversion factor: To convert AFP to micrograms per liter, multiply by 1.

[a] Log-rank tables for comparing 5-year recurrence-free and overall survival between the centers are available in eTable 1A and B in the Supplement.

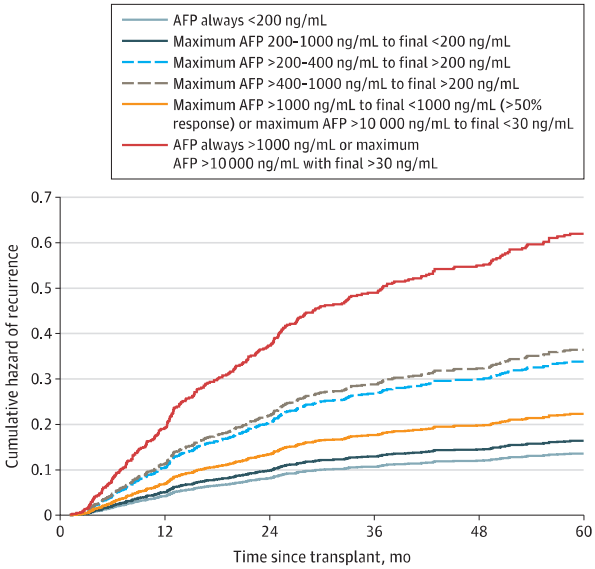
NYCA Score Validation

The original NYCA score included points for tumor size, number, and AFP-R (Table 1). These scores were assigned to patients in this study and the validated NYCA score was generated. Patients were again risk stratified into 3 recurrence risk categories. **Figure 2** shows the results of competing risk regression for the validated NYCA risks, using death before recurrence as a competing risk for HCC recurrence. The 5-year cumulative incidence of recurrence risk in NYCA risk categories was 9.5% for low risk (NYCA score, 0-2), 20.5% for acceptable risk (NYCA score 3-6), and 40.5% (NYCA score, ≥7) for high risk (all $P < .001$). Similar to the original score, we observed a progressive increase in recurrence risk with an increasing NYCA score, with the high-risk group having more than a 5-times increased recurrence risk compared with the low-risk group and more than 2 times the recurrence risk of the acceptable-risk group. The validated NYCA score competing risk regression model was then applied to each center (eTable 4 in the Supple-

© 2021 American Medical Association. All rights reserved.

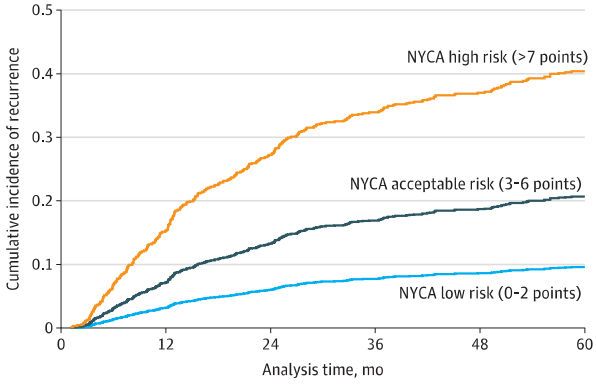
Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

Figure 1. Cumulative Hazard of Recurrence by Degree of Response From Maximum to Final α-Fetoprotein (AFP) Level in the Validated New York/California Score



To convert AFP to micrograms per liter, multiply by 1.

Figure 2. Competing Risk Regression Analysis by Validated New York/California (NYCA) Score Category Showing Significant Difference Between the 3 NYCA Categories



Five-year cumulative incidence of recurrence risk in NYCA risk categories was 9.5% for low risk (NYCA score, 0-2), 20.5% for acceptable risk (NYCA score 3-6), and 40.5% (NYCA score, ≥7) for high risk. Death was the competing risk to recurrence.

ment). The NYCA score was valid at predicting RFS at each risk category in 4 of the 5 centers. In the St Louis center, the recurrence risk for patients in the high-risk category did not differ significantly from that of the low and acceptable risk categories. The St Louis center had the lowest number of events of any of the centers studied (10.8% recurrence) and the highest 5-year RFS, and thus had a small number of patients in the high-risk group (3% vs 10% at the center with the highest number [Toronto]), likely accounting for this. A sensitivity analysis was performed to examine the validity of the NYCA score only in patients waiting for more than 6 months, which is the current situation for most candidates for liver transplant in the US. The NYCA remained valid in this subgroup of patients, showing almost identical cumulative incidence for recurrence for 3 validated NYCA risk categories (5-year cumulative incidence of recurrence per NYCA risk for patients waiting more than 6-months: NYCA low risk, 9.4%; acceptable risk, 20.8%; and high risk, 42.2%).

OS and Explant Characteristics

Cox proportional hazards regression analysis was performed to assess factors associated with OS. The validated NYCA score independently predicted OS after controlling for other factors, namely, diagnosis of hepatitis C and age, which were the only other factors found to significantly affect OS (NYCA acceptable risk, hazard ratio, 1.35; 95% CI, 1.14-1.60; $P = .001$ and NYCA high risk, hazard ratio, 2.36; 95% CI, 1.81-3.10; $P < .001$). As depicted in **Figure 3**, validated NYCA accurately stratified OS. Validated NYCA risk categories also correlated well with explant pathologic factors (eTable 5 in the Supplement). Vascular invasion was present in 37.6% of high-risk patients compared with 20.6% of acceptable-risk patients and 12.8% of low-risk patients ($P < .001$). Similarly, high-risk patients had a higher proportion (14.8%) of patients with poor differentiation (acceptable risk, 9.4%; low risk, 6.6%; $P < .001$). In addition, high-risk patients were less likely to have complete pathologic response (complete necrosis) of their tumors (9.6%) compared with patients at acceptable (15.8%) and low (24.0%) risk ($P < .001$).

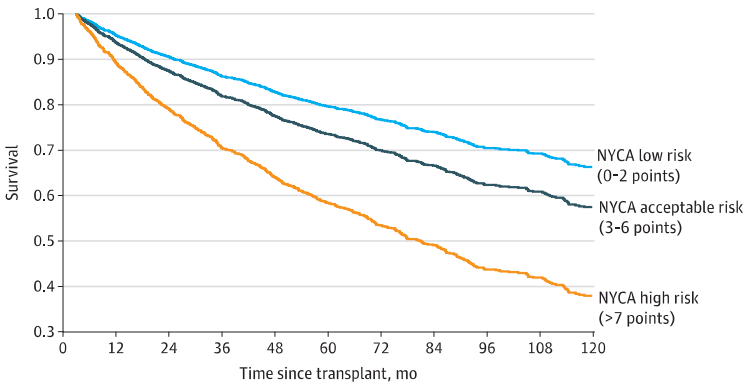
Comparison With Other Models and Reclassification

The validated NYCA score was compared with the F-AFP, MC, and Metroticket 2.0 models using receiver operating characteristic analysis. The validated NYCA score was significantly better at predicting recurrence than all other scores, with a Harrell C statistic 0.66 for validated NYCA vs 0.60 for Metroticket 2.0 (using final AFP level), 0.57 for F-AFP (using first AFP level), and 0.58 for MC ($P < .001$ for all). A total of 83.1% (453 of 545) of the patients who did not meet MC, 69.2% (213 of 308) of patients who did not meet F-AFP, and 76.1% (292 of 384) of patients who did not meet Metroticket 2.0 would be recategorized into either low-risk or acceptable-risk NYCA groups with 5-year RFS greater than 75% and a 5-year OS greater than 70%. Because both F-AFP and Metroticket 2.0 use a single AFP level and in order to observe whether there was variability in the predictive power of the scores using different AFP time points, the F-AFP level was recalculated using the maximum and final AFP levels, and the Metroticket 2.0 score was recalculated using the first and maximum AFP levels. Although the C statistic of the scores for predicting recurrence improved to 0.60 for F-AFP using both the maximum and final AFP levels (from 0.57 for the first AFP level) and to 0.61 for Metroticket 2.0 using the maximum AFP level (C statistic decreased to 0.58 for the first AFP level), these scores remained statistically significantly worse than using the NYCA-based AFP-R score (C statistic, 0.66; $P < .001$ for all). The net reclassification index for NYCA at predicting recurrence compared with the MC was 8.1 based on a correct reclassification

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

Figure 3. Log-Rank Testing According to New York/California (NYCA) Score Categories Showing Significant Differences in Overall Survival Between the 3 Groups



No. at risk

NYCA low risk	1562	1472	1376	1260	1136	979	789	645	493	362	236
NYCA acceptable risk	554	506	453	405	361	299	240	175	138	98	70
NYCA high risk	116	103	91	67	59	51	43	36	30	21	17

For the NYCA low-risk category, acceptable vs low risk, $P = .001$; for the NYCA acceptable-risk category, low vs acceptable risk, $P = .001$; and for the NYCA high-risk category, both comparisons, $P < .001$.

of events in 12.6% of events while overstratifying 4.4% nonevents. The positive net reclassification index (8.1) indicates that the increased sensitivity of NYCA over the MC of 12.6% is worth the decreased specificity of 4.4%. Similarly, compared with F-AFP (first AFP) the net reclassification index of NYCA for predicting recurrence was also positive at 12.9. The net reclassification index of NYCA vs F-AFP was 5.7 using the maximum and 7.7 with the final AFP. Compared with Metroticket 2.0, the net reclassification index for NYCA at predicting recurrence was 10.1 using the first, 4.1 using the maximum, and 6.7 using the final AFP levels (eTable 6 in the Supplement provides net reclassification index derivation).

Discussion

This study externally validates the only HCC selection tool to incorporate AFP-R, the NYCA score, on a large, heterogeneous cohort of patients from Europe and North America and underscores the importance of incorporating the AFP-R metric into selection of patients for liver transplant. α-Fetoprotein levels are known to predict response to locoregional therapy, waitlist outcomes, postresection outcomes, and liver transplant outcomes.[19-21] The degree of increase, absolute cutoff levels, and their application to liver transplant selection have, however, been a source of controversy. Novel seminal scoring systems, such as the F-AFP[8] and Metroticket 2.0,[11] have been designed to include single-snapshot AFP values at 3 cutoff levels (0-100, 100-1000, and >1000 ng/mL) into dichotomous selection criteria. What NYCA offers over these scores is insight into the importance of AFP kinetics in refining selection. Our data suggest that AFP-R is a more robust biological metric than simply using AFP at listing or transplant. Although one can argue that F-AFP and Metroticket 2.0 scores can be recalculated over time while patients are on waitlists, many patients with marked and extreme AFP increases, very large or numerous lesions, or a combination thereof would be excluded from transplant or downstaging pathways in many centers, despite the potential for good outcomes (>75% RFS for NYCA acceptable risk). To test how these scores perform, we recalculated F-AFP and Metroticket 2.0 scores using AFP levels at the 3 times available. Only NYCA was found to be significantly superior to MC at predicting RFS and OS. The Metroticket 2.0 was designed to predict HCC-specific mortality—not recurrence. Recurrence after transplant is considered terminal, with a median time to death after recurrence of approximately 12 to 18 months,[22-24] hence, our rationale for choosing recurrence as the end point in this study. Some patients, however, can attain longer survival times after aggressive management and may die from other causes.[21] The reason to use HCC-specific mortality in Metroticket 2.0 was to allow for the recent introduction of direct-acting antivirals, and the substantial decrease in mortality associated with HCV recurrence. However, even when accounting for this on competing risk regression, validated NYCA still outperformed Metroticket 2.0. To assess the utility of NYCA vs Metroticket 2.0 at predicting HCC-related death, we considered any death within 18 months of HCC recurrence as an HCC-related death, and all other deaths after recurrence or otherwise as non-HCC related. The Harrell C statistic for validated NYCA to predict HCC-related death effectively remained unchanged and significantly higher than Metroticket 2.0.

The original NYCA score was established on a large cohort of patients in New York and California; more than 61% of the patients had hepatitis C and less than 9% had alcohol-related liver disease, and the cohorts were largely homogeneous among the 3 centers.[12] In the present study, although 53% of the patients had hepatitis C, the participating European and Canadian centers allowed the study of a more heterogeneous population, with more than 20% of patients having alcohol-related liver disease. The ability of NYCA to be validated individually in each center further strengthens the case for widespread applicability to different patient populations, because European centers had an equal distribution of hepa-

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

titis C and adrenoleukodystrophy among their patients. In addition, further strengthening the study is the apparent variation in listing practices among centers—namely, use of locoregional therapy and waiting times. Almost all patients in the original study received locoregional therapy and had a median waiting time of 9 months.[12] In contrast, waiting times in this study were variable and ranged from 2 to more than 9 months; thus, locoregional therapy practices also differed. The ability of the NYCA to predict RFS independently of locoregional therapy use and waiting time is key to its potential for widespread applicability. We purposefully did not specifically examine the different locoregional therapy modalities and their individual effectiveness for 2 reasons: first, approximately 75% of the patients in our study received transarterial chemoembolization (as is the case in most centers), which is likely to skew the results, and second, 2 large multicenter studies have reported no correlation with locoregional therapy modality and outcome,[25,26] adding to the existing literature that demonstrated no statistically significant differences in outcomes based on locoregional therapy modality.

Notably, this validation effort includes patients who would have been deemed unfit for transplant at most centers, namely, those with maximum AFP levels greater than 10 000 ng/mL. These patients appear to have a unique entity of aggressive tumors that differ in their recurrence risk from those with AFP levels greater than 1000 but less than 10 000 ng/mL, with a more than 2-times risk of recurrence (eFigure 2B in the Supplement). Although it is widely accepted that AFP levels greater than 1000 ng/mL confer poorer outcomes,[15,27] few studies have examined patients with AFP levels greater than 10 000 ng/mL. Nomura et al[28] reported in 1989 that AFP level increases greater than 10 000 ng/mL conferred worse prognosis and larger tumor size than AFP levels of 1000 to 10 000 and less than 1000 ng/mL. Other studies have shown correlations of AFP levels greater than 10 000 ng/mL with poorer survival.[29,30] However, the effect of treatment in such a population, to our knowledge, has never been assessed. Our work suggests that an AFP-R decrease from greater than 10 000 ng/mL to AFP less than 30 ng/mL resulted in no recurrences (0 of 8), but 61.5% of patients (8 of 13) without this AFP-R experienced recurrence. Most patients with AFP levels greater than 10 000 ng/mL are excluded from transplant immediately for fear of infiltrative HCC and/or vascular involvement. Automatically excluding these patients without objective assessment of AFP-R may not be warranted based on our findings.

The superiority of including AFP-R into selection criteria is not limited to improved RFS or OS; it also allows for inclusion of patients who would have been excluded by both the MC and F-AFP. Reclassification of 83% of patients who did not meet MC and 69% of patients who did not meet F-AFP into the acceptable or low-risk NYCA groups in this study (5-year RFS > 75%), underlies the importance of moving away from dichotomous single AFP nonkinetic scores. Oncologic standards of care often assess or rely on responses to neoadjuvant therapy before surgery in various solid-organ cancers. The transplant community has lagged in implementing response-based scores for HCC selection, largely due to the success of MC and the relatively low recurrence rates. The judicious use of organs and ability to provide cure to a large cohort of patients, however, compels the search for criteria that are inclusive without increasing failures. Modified Response Evaluation Criteria in Solid Tumors (mRECIST) are response-driven criteria,[31] albeit radiology based, that have been validated and widely adopted in clinical evaluation of early/intermediate stage HCC and have been crucial to the evaluation of drugs in phase 2/3 clinical trials.[32] The mRECIST has been reported to predict waitlist and posttransplant outcomes in HCC, with failure of response or progression of disease being synonymous with increased risk of dropout and posttransplant recurrence.[33] Recently Cucchetti et al[34] combined mRECIST with Metroticket 2.0, improving the original scale's ability to predict HCC-related death, with a net reclassification index of 5.8. This addition of an assessment of radiologic response is similar to adding a measurable AFP-R, as is the case with NYCA, into morphometric selection tools to improve prediction of outcomes. Although it was beyond the scope of this study to add mRECIST, such an addition would perhaps further strengthen the ability of NYCA to predict outcomes. One advantage of using AFP-R compared with mRECIST, however, is that AFP is an objective measure of biology and not subject to variability in interpretation based on radiologic assessment. This is possibly reflected in the higher net reclassification indexes observed in our study, although, as pointed out by Cucchetti et al,[34] higher net reclassification index is not necessarily synonymous with increased accuracy, and thus we chose to also depict the directional changes.[29,35,36] Nevertheless, the work of Cucchetti et al[34]and the current validation of AFP-R together provide evidence to transplant-governing bodies to move away from single-point measures to predict success in HCC post transplant and add response assessments into selection tools.

Limitations

This study has limitations, including its retrospective, multicenter design, which results in inherent variability in individual centers' HCC management algorithms and introduces ascertainment bias, because all patients in the study received liver transplant, and thus were deemed to have appropriate tumor biological characteristics. Another limitation is the lack of information on timing of AFP values with respect to locoregional therapy, and time between the final measure of AFP levels and liver transplant. Although choosing maximum and final AFP levels simplified our model and removed the possibility of the AFP level increasing beyond its previous peak before liver transplant, it does not eliminate the possibility that the AFP level may have reached a nadir and then started increasing below the maximum level before transplant. Availability of this information would allow a more thorough assessment of response but would introduce complexities in using such a model for liver transplant selection. Our data are strengthened by comparisons with static scores using AFP levels at the 3 different time points; however, lack of availability of every AFP value is a limitation, because we were unable to assess the value of each result. Although the aim of this study was not to specifically examine the use of locoregional therapy, because locoregional therapy is currently standard of care in most US centers, it is intuitive that patients with AFP responses received

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

locoregional therapy, and our study noted that, among responders, 158 of 167 patients (95%) received at least 1 locoregional therapy before transplant, and among 331 patients with AFP levels greater than 200 ng/mL, 290 (88%) received locoregional therapy before liver transplant. It is unclear why few patients (n = 9) did not receive locoregional therapy but had a response; however, all centers participating are tertiary referral centers and the possibility exists that patients received treatment at another institution that was not documented at the transplant center. Although the reason for a decrease in the AFP level in these 9 remaining patients is unclear, they were included as responders to avoid bias. Another limitation was the lack of information on waitlisted patients who dropped out of the study. Assessment of AFP-R on an intent-to-treat basis would be of value in allowing a more thorough understanding of AFP kinetics and the association with waitlist mortality.

Conclusions

The findings of this prognostic study externally validate the NYCA score and highlight the utility of the dynamic AFP-R at predicting patient outcomes after liver transplant for HCC. Incorporation of such objective response-based biological assessment of patients with HCC before transplant into selection tools should become standardized as our understanding of their importance evolves.

ARTICLE INFORMATION

Accepted for Publication: February 5, 2021.

Published Online: May 5, 2021. doi:10.1001/jamasurg.2021.0954

Author Affiliations: Division of Liver Transplantation and Hepatobiliary Surgery, Department of Surgery, Weill Cornell Medicine, New York, New York (Halazun, Rosenblatt, Hajifathalian, Samstein); Center for Liver Disease and Transplantation, Columbia University Medical Center, New York Presbyterian Hospital, New York (Halazun, Rosenblatt, Hajifathalian, Najjar, Samstein, Brown, Emond); Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medical College, New York, New York (Rosenblatt, Hajifathalian, Brown); Division of Gastroenterology, Department of Medicine, University of California, San Francisco (Mehta, Yao); Sapienza University, Rome, Italy (Lai, Rossi, Mennini); Multi-Organ Transplant and HPB Surgical Oncology, Division of General Surgery, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada (Gorgen, Ivanics, Sapisochin); Center for Advanced Digestive Care, Division of Hematology and Oncology, New York-Presbyterian Hospital, Weill Cornell Medical College of Cornell University, New York (Brar); Department of General Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, Ohio (Sasaki, Quintini, Aucejo); Section of Abdominal Organ Transplant, Department of Surgery, Washington University School of Medicine, Washington University, St Louis, Missouri (Doyle, Chapman); Recanati/Miller Transplantation Institute, Department of Transplantation, Mount Sinai Medical Center, New York, New York (Tabrizian); Division of Liver and Pancreas Transplantation, Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles, California (Agopian); Institute for Experimental and Clinical Research (IREC) Université catholique de Louvain (UCL), Brussels, Belgium (Lerut, Iesari); Gastroenterology and Hepatology, Department of Internal Medicine I, Innsbruck Medical University, Innsbruck, Austria (Finkenstedt, Schaefer); Department of Transplantation and Hepatobiliary Surgery, University of Mainz, Mainz, Germany (Mittler, Hoppe-Lotichius).

Author Contributions: Dr Halazun had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Halazun, Gorgen, Tabrizian, Najjar, Brown, Emond, Lerut, Chapman, Sapisochin.
Acquisition, analysis, or interpretation of data: Halazun, Rosenblatt, Mehta, Lai, Hajifathalian, Gorgen, Brar, Sasaki, Doyle, Tabrizian, Agopian, Najjar, Ivanics, Samstein, Brown, Yao, Lerut, Rossi, Mennini, Iesari, Finkenstedt, Schaefer, Mittler, Hoppe-Lotichius, Quintini, Aucejo, Sapisochin.
Drafting of the manuscript: Halazun, Gorgen, Najjar, Quintini, Chapman.
Critical revision of the manuscript for important intellectual content: Halazun, Rosenblatt, Mehta, Lai, Hajifathalian, Gorgen, Brar, Sasaki, Doyle, Tabrizian, Agopian, Najjar, Ivanics, Samstein, Brown, Emond, Yao, Lerut, Rossi, Mennini, Iesari, Finkenstedt, Schaefer, Mittler, Hoppe-Lotichius, Quintini, Aucejo, Sapisochin.
Statistical analysis: Halazun, Rosenblatt, Hajifathalian, Gorgen, Agopian, Najjar.
Administrative, technical, or material support: Halazun, Doyle, Tabrizian, Ivanics, Brown, Lerut, Iesari, Schaefer, Mittler.
Supervision: Mehta, Lai, Agopian, Samstein, Emond, Rossi, Aucejo, Sapisochin.

Conflict of Interest Disclosures: Dr Rosenblatt reported a spouse briefly owning shares in Gilead Stock outside the submitted work. Dr Chapman reported receiving nonfinancial support from Novartis outside the submitted work; in addition, Dr Chapman had a patent for Pathfinder and received royalties. Dr Sapisochin reported receiving grants from Roche and Bayer and personal fees from Integra outside the submitted work. No other disclosures were reported.

REFERENCES

1. Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med*. 1996;334(11):693-699. doi:10.1056/NEJM199603143341104

2. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med*. 2011;365(12):1118-1127. doi:10.1056/NEJMra1001683

3. Yao FY, Ferrell L, Bass NM, et al. Liver transplantation for hepatocellular carcinoma: expansion of the tumor size limits does not adversely impact survival. *Hepatology*. 2001;33(6):1394-1403. doi:10.1053/jhep.2001.24563

4. Mazzaferro V, Llovet JM, Miceli R, et al; Metroticket Investigator Study Group. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol*. 2009;10(1):35-43. doi:10.1016/S1470-2045(08)70284-5

5. Toso C, Trotter J, Wei A, et al. Total tumor volume predicts risk of recurrence following liver transplantation in patients with hepatocellular carcinoma. *Liver Transpl*. 2008;14(8):1107-1115. doi:10.1002/lt.21484

6. Agopian VG, Harlander-Locke M, Zarrinpar A, et al. A novel prognostic nomogram accurately predicts hepatocellular carcinoma recurrence after liver transplantation: analysis of 865 consecutive liver transplant recipients. *J Am Coll Surg*. 2015;220(4):416-427. doi:10.1016/j.jamcollsurg.2014.12.025

7. Halazun KJ, Hardy MA, Rana AA, et al. Negative impact of neutrophil-lymphocyte ratio on outcome after liver transplantation for hepatocellular carcinoma. *Ann Surg*. 2009;250(1):141-151. doi:10.1097/SLA.0b013e3181a77e59

8. Duvoux C, Roudot-Thoraval F, Decaens T, et al; Liver Transplantation French Study Group. Liver transplantation for hepatocellular carcinoma: a model including α-fetoprotein improves the performance of Milan criteria. *Gastroenterology*. 2012;143(4):986-94.e3. doi:10.1053/j.gastro.2012.05.052

9. Halazun KJ, Najjar M, Abdelmessih RM, et al. Recurrence after liver transplantation for hepatocellular carcinoma: a new MORAL to the story. *Ann Surg*. 2017;265(3):557-564. doi:10.1097/SLA.0000000000001966

10. Notarpaolo A, Layese R, Magistri P, et al. Validation of the AFP model as a predictor of HCC recurrence in patients with viral hepatitis-related cirrhosis who had received a liver transplant for HCC. *J Hepatol*. 2017;66(3):552-559. doi:10.1016/j.jhep.2016.10.038

11. Mazzaferro V, Sposito C, Zhou J, et al. Metroticket 2.0 model for analysis of competing risks of death after liver transplantation for hepatocellular carcinoma. *Gastroenterology*. 2018;154(1):128-139. doi:10.1053/j.gastro.2017.09.025

12. Halazun KJ, Tabrizian P, Najjar M, et al. Is it time to abandon the Milan Criteria? results of a bicoastal US collaboration to redefine hepatocellular carcinoma liver transplantation selection policies. *Ann Surg*. 2018;268(4):690-699. doi:10.1097/SLA.0000000000002964

13. Peng SY, Chen WJ, Lai PL, Jeng YM, Sheu JC, Hsu HC. High alpha-fetoprotein level correlates

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021

with high stage, early recurrence and poor prognosis of hepatocellular carcinoma: significance of hepatitis virus infection, age, p53 and beta-catenin mutations. *Int J Cancer*. 2004;112(1):44-50. doi:10.1002/ijc.20279

14. Jeng YM, Chang CC, Hu FC, et al. RNA-binding protein insulin-like growth factor II mRNA-binding protein 3 expression promotes tumor invasion and predicts early recurrence and poor prognosis in hepatocellular carcinoma. *Hepatology*. 2008;48(4):1118-1127. doi:10.1002/hep.22459

15. Hameed B, Mehta N, Sapisochin G, Roberts JP, Yao FY. Alpha-fetoprotein level > 1000 ng/mL as an exclusion criterion for liver transplantation in patients with hepatocellular carcinoma meeting the Milan Criteria. *Liver Transpl*. 2014;20(8):945-951. doi:10.1002/lt.23904

16. OPTN/UNOS policy notice modification to hepatocellular carcinoma (HCC) extension criteria. Accessed April 15, 2020. https://optn.transplant.hrsa.gov/media/2411/modification-to-hcc-auto-approval-criteria_policy-notice.pdf

17. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. doi:10.7326/M14-0697

18. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157-172. doi:10.1002/sim.2929

19. Mehta N, Dodge JL, Goel A, Roberts JP, Hirose R, Yao FY. Identification of liver transplant candidates with hepatocellular carcinoma and a very low dropout risk: implications for the current organ allocation policy. *Liver Transpl*. 2013;19(12):1343-1353. doi:10.1002/lt.23753

20. Vibert E, Azoulay D, Hoti E, et al. Progression of alphafetoprotein before liver transplantation for hepatocellular carcinoma in cirrhotic patients: a critical factor. *Am J Transplant*. 2010;10(1):129-137. doi:10.1111/j.1600-6143.2009.02750.x

21. Toso C, Meeberg G, Hernandez-Alejandro R, et al. Total tumor volume and alpha-fetoprotein for selection of transplant candidates with hepatocellular carcinoma: a prospective validation. *Hepatology*. 2015;62(1):158-165. doi:10.1002/hep.27787

22. Fernandez-Sevilla E, Allard MA, Selten J, et al. Recurrence of hepatocellular carcinoma after liver transplantation: is there a place for resection? *Liver Transpl*. 2017;23(4):440-447. doi:10.1002/lt.24742

23. Sapisochin G, Goldaracena N, Astete S, et al. Benefit of treating hepatocellular carcinoma recurrence after liver transplantation and analysis of prognostic factors for survival in a large Euro-American series. *Ann Surg Oncol*. 2015;22(7):2286-2294. doi:10.1245/s10434-014-4273-6

24. Bodzin AS, Lunsford KE, Markovic D, Harlander-Locke MP, Busuttil RW, Agopian VG. Predicting mortality in patients developing recurrent hepatocellular carcinoma after liver transplantation: impact of treatment modality and recurrence characteristics. *Ann Surg*. 2017;266(1):118-125. doi:10.1097/SLA.0000000000001894

25. Agopian VG, Harlander-Locke MP, Ruiz RM, et al. Impact of pretransplant bridging locoregional therapy for patients with hepatocellular carcinoma within Milan Criteria undergoing liver transplantation: analysis of 3601 patients from the US multicenter HCC Transplant Consortium. *Ann Surg*. 2017;266(3):525-535. doi:10.1097/SLA.0000000000002381

26. Kardashian A, Florman SS, Haydel B, et al. Liver transplantation outcomes in a US multicenter cohort of 789 patients with hepatocellular carcinoma presenting beyond Milan Criteria. *Hepatology*. 2020;72(6):2014-2028. doi:10.1002/hep.31210

27. Hakeem AR, Young RS, Marangoni G, Lodge JP, Prasad KR. Systematic review: the prognostic role of alpha-fetoprotein following liver transplantation for hepatocellular carcinoma. *Aliment Pharmacol Ther*. 2012;35(9):987-999. doi:10.1111/j.1365-2036.2012.05060.x

28. Nomura F, Ohnishi K, Tanabe Y. Clinical features and prognosis of hepatocellular carcinoma with reference to serum alpha-fetoprotein levels: analysis of 606 patients. *Cancer*. 1989;64(8):1700-1707. doi:10.1002/1097-0142(19891015)64:8<1700::AID-CNCR2820640824>3.0.CO;2-Z

29. Chan SL, Chan AT, Yeo W. Role of alpha-fetoprotein in hepatocellular carcinoma: prognostication, treatment monitoring or both? *Future Oncol*. 2009;5(6):889-899. doi:10.2217/fon.09.64

30. Ikai I, Arii S, Kojiro M, et al. Reevaluation of prognostic factors for survival after liver resection in patients with hepatocellular carcinoma in a Japanese nationwide survey. *Cancer*. 2004;101(4):796-802. doi:10.1002/cncr.20426

31. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis*. 2010;30(1):52-60. doi:10.1055/s-0030-1247132

32. Llovet JM, Lencioni R. mRECIST for HCC: performance and novel refinements. *J Hepatol*. 2020;72(2):288-306. doi:10.1016/j.jhep.2019.09.026

33. Lee DD, Samoylova M, Mehta N, et al. The mRECIST classification provides insight into tumor biology for patients with hepatocellular carcinoma awaiting liver transplantation. *Liver Transpl*. 2019;25(2):228-241. doi:10.1002/lt.25333

34. Cucchetti A, Serenari M, Sposito C, et al. Including mRECIST in the Metroticket 2.0 criteria improves prediction of hepatocellular carcinoma-related death after liver transplant. *J Hepatol*. 2020;73(2):342-348. doi:10.1016/j.jhep.2020.03.018

35. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. 2014;160(2):122-131. doi:10.7326/M13-1522

36. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology*. 2014;25(1):114-121. doi:10.1097/EDE.0000000000000018

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a KU Leuven 2Bergen Biomedical Library User on 05/10/2021