Is it Time to Abandon the Milan Criteria?

Results of a Bicoastal US Collaboration to Redefine Hepatocellular Carcinoma Liver Transplantation Selection Policies

Karim J. Halazun, MD, FACS,† Parissa Tabrizian, MD,‡ Marc Najjar, MD,† Sander Florman, MD,‡ Myron Schwartz, MD,‡ Fabrizio Michelassi, MD,* Benjamin Samstein, MD,*† Robert S. Brown Jr, MD, MPH,*†§ Jean C. Emond, MD,*† Ronald W. Busuttil, MD, PhD,|| and Vatche G. Agopian, MD, FACS||*

Objectives: European liver transplant (LT) centers have moved away from using the Milan Criteria (MC) for hepatocellular carcinoma (HCC) patient selection, turning to models including tumor biological indices, namely alpha-fetoprotein (AFP). We present the first US model to incorporate an AFP response (AFP-R), with comparisons to MC and French-AFP models (F-AFP).

Methods: AFP-R was measured as differences between maximum and final pre-LT AFP in HCC patients undergoing LT at 3 US centers (2001 to 2013). Cox and competing risk-regression analyses identified predictors of recurrence-free survival (RFS).

Results: Of 1450 patients, 235 (16.2%) were outside MC. Tumor size, number, and AFP-R were independent predictors of RFS and were assigned weighted points based on Cox-regression analysis. An AFP-R consistently < 200 ng/mL predicted the best outcome. A 3-tiered competing-risk RFS model, the New York/California (NYCA) score, was developed, accurately discriminating between groups ($P < 0.001$), and correlating with overall survival ($P < 0.001$). Two hundred one of 235 patients outside MC (85.5%) would be recategorized into NYCA low/acceptable-risk groups. The c-statistic for our NYCA score is 0.731 compared with 0.613 for MC and 0.658 for F-AFP ($P < 0.0001$).

Conclusion: Incorporation of AFP-R into HCC selection criteria allows for MC expansion. As United Network for Organ Sharing considers adding AFP to selection algorithms, the NYCA score provides an objective, user-friendly tool for centers to appropriately risk-stratify patients.

Keywords: alpha-fetoprotein, hepatocellular carcinoma, Milan criteria, tumor biology

(*Ann Surg* 2018;268:690–699)

Over 2 decades ago, Mazzaferro et al developed the Milan criteria,[1] vastly changing the landscape of liver transplantation (LT) for hepatocellular carcinoma (HCC) around the globe by setting clear guidelines for which patients with HCC benefit from LT with an ''acceptable'' recurrence-free survival (RFS). These criteria have been adopted globally by several transplant authorities to prioritize or even preclude LT for patients with HCC. Despite their utility and reproducibility, the Milan criteria have been repeatedly criticized for their restrictiveness, and more importantly, the lack of tumor biological indices to help dictate best oncological practice when transplanting HCC patients. Many groups have proposed the transplantation of patients with larger and more numerous tumors,[2–5] achieving results comparable to Milan. Most of these studies, however, followed original methodology by Mazzaferro, essentially retrospectively staging patients based on explant pathological data. As there is a high correlation between morphology and outcomes, selection of HCC patients based on morphology of the tumor has become the standard approach. More recently, several centers, including our own, have developed criteria that include tumor biological indices such as AFP to predict outcome.[6–10] Many of these scores and criteria have been shown to outperform the Milan criteria's ability to predict recurrence and survival; however, few have received widespread adoption by transplant regulatory agencies or been used to prioritize patients on the transplant waiting list and most remain confined to acting as selection tools in individual institutions. In the US, despite the evidence for the efficacy of scores based on biologic indices in prognostication, the United Network for Organ Sharing (UNOS) has been slow to move away from morphologic criteria for patient selection. In France, however, liver transplant centers have adopted a scoring system that includes a (AFP) in their selection algorithm based on research conducted by the French Liver Transplantation Study group.[11] The French AFP score not only includes tumor biology but also expands on the size and number criteria traditionally used, allowing for more patients with out-of-criteria (Milan) HCCs to be transplanted and cured.[11]

AFP has long been recognized as a biological predictor of prognosis in HCC and has been central to many of the new models developed for the transplant setting. Unfortunately, many systems, including the French AFP score, have used AFP at a single time point even though patients usually wait many months for transplantation during which time neoadjuvant therapies are administered. We hypothesized that the response of AFP over time serves as a better predictor of recurrence and survival and sought to investigate the importance of a dynamic AFP response through a multi-institutional collaboration from 3 of the largest liver transplant centers in the United States.

METHODS

Study Population

Prospectively maintained databases of all adult patients undergoing LT for HCC between January 2001 and December 2013 from 3

From the *Division of Liver Transplantation and Hepatobiliary Surgery, Department of Surgery, Weill Cornell Medicine, New York, NY; †Center for Liver Disease and Transplantation, Columbia University Medical Center, NY Presbyterian Hospital, New York, NY; ‡Department of Transplantation, Recanati/Miller Transplantation Institute, Mount Sinai Medical Center, New York, NY; §Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medical College, New York, NY; and ||Division of Liver and Pancreas Transplantation, Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA.

The authors report no conflicts of interest.

Reprints: Karim J. Halazun, MD, FACS, Assistant Professor of Surgery, Liver Transplant and HPB Surgeon, Division of Liver Transplantation and Hepatobiliary Surgery, Department of Surgery, Weill Cornell Medical College, 525 East 68th, F-763, New York, NY 10065.
E-mail: kah7007@med.cornell.edu; Vatche G. Agopian, MD, FACS, Associate Professor of Surgery, Department of Surgery, David Geffen School of Medicine at UCLA, Ronald Reagan Medical Center, Suite 8501-B, Los Angeles, CA 90095; E-mail: vagopian@mednet.ucla.edu.

ISSN: 0003-4932/18/26804-0690
DOI: 10.1097/SLA.0000000000002964

institutions in the US were analyzed. The centers participating in the study were New York Presbyterian Hospital Center for Liver Disease and Transplantation (Columbia University Medical Center and Weill Cornell Medicine), Mount Sinai School of Medicine, and UCLA. Data analyzed included basic demographic patient data, donor-specific data, laboratory data as well as tumor-specific data. Data were also collected on the use of pretransplant ablative/bridging therapies as well explant pathological data.

Alpha-Fetoprotein (AFP) Cutoff and AFP-Response

AFP levels at diagnosis, maximum AFP (Max-AFP) at any time point, and the final immediate pretransplant AFP (Final-AFP) were recorded. AFP levels of >200 were used as our cutoff for marked elevation based on previously observed correlation with AFP mRNA expression in both HCC and normal liver tissue, which are associated with poor prognosis.[8,12,13] AFP levels >1000 were used as our cutoff for extreme elevation based on previous work[14] with the French AFP score and the new UNOS regulations both using AFP >1000 as their upper limit for patient selection.[11,14,15] In order to confirm the clinical utility of these cut offs, we assessed Max-AFP and Final-AFP at levels <200, 200 to 1000, and >1000 to ascertain their ability to predict 5-year RFS using Kaplan-Meier curves. We then performed a threshold analysis through serial Cox-regression modeling to ascertain the AFP response that best predicted 5-year RFS, based on the degree of change of AFP from maximum value to the final pretransplant value.

Survival, Regression, and ROC Analyses

The primary outcome measure of the study was 5-year RFS, and we used Kaplan-Meier survival analysis to measure RFS in our patients. Patients who had perioperative deaths within 90 days of transplant were therefore excluded. A Cox regression analysis was then performed in a stepwise manner to conduct a multivariable analysis of pretransplant clinic-pathological factors that had a significant impact on 5-year RFS and in order to establish appropriate AFP cutoff as detailed above. Tumor size and number were converted to categorical values, with 3 categories in each variable [Size: 0 to 3 cm (reference), >3 to 6 cm, and >6 cm; Number: 1 tumor, 2 to 3 tumors, and ≥4 tumors]. Independent predictors of poor RFS were used to construct a new score, which we termed the NYCA (New York/California) score. IBMSPSS Version 25 for the Macintosh was used to perform the analyses. Receiver operative curves (ROCs) were then used to compare the NYCA score to the Milan Criteria and the French AFP score.

RESULTS

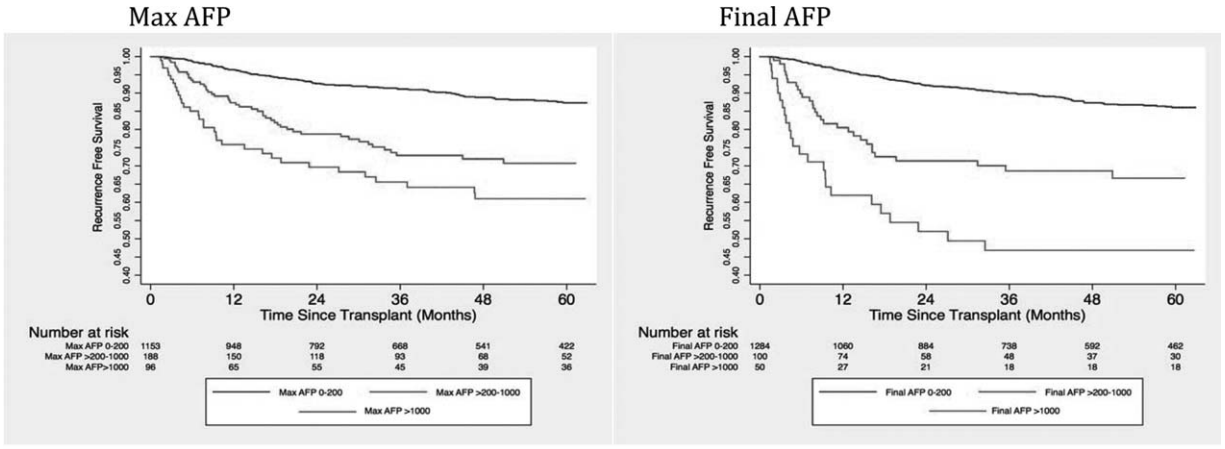
General Demographics and Tumor Characteristics

A total of 1450 patients were included with patient demographics and tumor characteristics in Table 1. Several center-specific differences are noted. A significantly higher proportion of patients transplanted in the NY centers had hepatitis C virus (HCV; 69.8% NYP, 65.4% Mt Sinai, 58.2% UCLA, $P < 0.0001$). A significantly higher proportion of patients with underlying alcoholic liver disease were transplanted at NYP and UCLA than Mt. Sinai (11.1% NYP, 9.0% UCLA, and 6.0% Mt Sinai, $P = 0.042$). Patients transplanted at UCLA had significantly larger mean tumors diameters at diagnosis than patients in the New York centers (3.31 cm UCLA vs 2.86 cm NYP and 2.73 cm Mt Sinai, $P < 0.0001$); however, a significantly higher proportion of patients in New York had multifocal HCCs than patients at UCLA (38.4% NYP, 36.6% Mt Sinai vs 24.2% UCLA, $P = 0.008$). Despite difference in tumor size and number, there were no significant differences in the proportions of patients transplanted outside Milan criteria in the 3 centers (18.0% NYP, 16.0% Mt Sinai, 15.6% UCLA, $P = 0.565$), and no differences in recurrence rates

TABLE 1. Demographic, Clinic-pathologic, Tumor, and Recurrence Data for Patients at the 3 Institutions Included in the Study With Comparisons Between the Institutions

Preoperatively Available Data (n = 1450)				
	NYP (n = 388)	**Mt Sinai (n = 382)**	**UCLA (n = 680)**	***P***
Patient demographics				
Male sex	77.3% (n = 300)	78.5% (n = 300)	72.9% (n = 496)	0.082
Mean age, y (SD)	58.18 ± 7.9	58.83 ± 8.4	58.2 ± 7.8	0.412
Median laboratory MELD at Tx (IQR)	12 (8–17)		13 (10–22)	
Mean BMI	27.7 ± 8.1		27.7 ± 5.5	
Diagnosis				
HCV	69.8% (n = 240)	65.4% (n = 250)	58.2% (n = 396)	0.001*
HBV	14.7% (n = 56)	16.0% (n = 61)	14.1% (n = 96)	0.716
Alcohol	11.1% (n = 43)	6.0% (n = 23)	9.0% (n = 61)	0.042*
Cryptogenic/NASH	5.7% (n = 22)	3.1% (n = 12)	6.5% (n = 44)	0.067
Other (PBC, PSC, AIH, hemochromatosis)	10.1% (n = 39)	9.2% (n = 35)	6.5% (n = 44)	0.082
Tumor characteristics				
Mean size of largest tumor, cm (SD)	2.86 ± 1.58	2.73 ± 1.33	3.31 ± 1.79	<0.0001*
Tumor size (Largest mass) > 3 cm	31.7% (n = 123)	28.2% (n = 105)	37.2% (n = 253)	0.008
Multifocal HCC	38.4% (n = 149)	36.6% (n = 138)	24.2% (n = 165)	<0.0001*
% Outside Milan at diagnosis	18% (n = 70)	16% (n = 61)	15.6% (n = 106)	0.565
Median AFP at diagnosis (IQR)	12.6 (6–52)	16.8 (6–61)	12.1 (4.8–49.2)	0.063
Median Maximum AFP (IQR)	24.4 (9–137)	30.0 (11–151)	20.2 (6–109)	0.046*
Median Final AFP (IQR)	11.5 (5–37)	10.6 (5–45)	9.1 (4–41)	0.241
NLR ≥ 5	13.7% (n = 53)	22.3% (n = 74)	24.4% (n = 161)	<0.0001*
Locoregional therapy				
% of patients receiving LRT	88.1%	82.0%	82.7%	0.05*
Recurrence data				
Explant pathology characteristics	NYP	Mt Sinai	UCLA	
Recurrence rates	13.4% (n = 52)	16.1% (n = 69)	11.3% (n = 77)	0.092
1, 3, 5-y recurrence-free survival	93%, 87%, 83%	94%, 86%, 80%	94%, 87%, 85%	NS for all

*Statistically significant.



A

LOG RANK TESTING	Max AFP 0-200	Max AFP >200-1000
Max AFP 0-200	-	p<0.0001
Max AFP >200-1000	p<0.0001	-
Max AFP > 1000	p<0.0001	p=0.032

B

LOG RANK TESTING	Final AFP 0-200	Final AFP >200-1000
Final AFP 0-200	-	p<0.0001
Final AFP >200-1000	p<0.0001	-
Final AFP > 1000	p<0.0001	p=0.008

FIGURE 1. Kaplan-Meier survival curves and log rank testing showing significant difference between the 3 levels of (A) Max AFP and (B) Final AFP on recurrence-free survival with significantly worse RFS at each stratum of rising AFP.

(13.4% NYP, 16.1% Mt Sinai, and 11.3% UCLA, $P = 0.092$) or 1, 3, and 5-year RFS between the centers (5-year RFS 83% NYP vs 80% Mt Sinai, $P = 0.688$, 5-year RFS 83% NYP vs 85% UCLA, $P = 0.416$, 5-year RFS 80% Mt Sinai vs 85% UCLA, $P = 0.181$).

AFP Cutoff and AFP Response

Figure 1 shows the effect of the 3 chosen AFP cutoffs on RFS for both Max-AFP (Fig. 1A) and Final-AFP (Fig. 1B) showing significant differences in 5-year RFS ($P < 0.05$ for all) and confirming the clinical utility of the chosen cutoffs (<200, >200 to 1000, and >1000) at differentiating the likelihood of poor 5-year RFS. Table 2 summarizes the results of the most predictive Cox model of AFP response predictive of poor 5-year RFS based on the above cutoffs, with Fig. 2 showing the cumulative hazard of recurrence for each level of AFP response from Max-AFP to Final-AFP. An AFP < 200 was the reference value, and levels that fell persistently to <200 predicted the best 5-year RFS. Interestingly, patients with Max-AFP that exceeded 1000 but fell to <1000 before transplantation with a response that exceeded 50% did just as well as patients with a Max-AFP of 200 to 1000 that fell to <200 before transplant (cumulative hazard of recurrence 24.5% and 24.1%, respectively).

Cox Regression Analysis and Creation of NYCA Score

All factors listed in Table 1, as well as the response to AFP described above were entered into a regression analysis to elucidate predictors of poor 5-year RFS. Four factors (maximum tumor size at diagnosis, maximum tumor number at diagnosis, AFP response from Max-AFP to Final-AFP, and an NLR ≥5) were independent predictors of 5-year RFS on both univariate and multivariate analysis (Table 3). A simple integer scoring system based on the hazard ratios (HRs) of the independent predictors of 5-year RFS (apart from NLR) was used, as has previously described.[6,11,16] The HRs of the factors derived from the Cox regression model were rounded to the nearest integer, and patients were then respective assigned points (Table 3). These points were used in an additive fashion to derive a score for each patient, giving a minimum score of 0 for those patients with none of the factors, and a maximum score of 14 for those with all 3 factors. Patients were then risk stratified according to their risk of recurrence based on the NYCA score, where patients in the Low-Risk Group had a score of 0 to 2, the Acceptable-Risk Group had a score of 3 to 6, and the High-Risk group had a score of >7 (Table 4). Note that the NLR was excluded due to the potential variability in capture of this variable between the 3 institutions and its variability with

TABLE 2. Cox Regression Analysis of AFP Response From Max AFP to Final AFP Best Predicting Poor 5-year RFS With AFP <200 Acting as the Reference Value

AFR Response	*P*	Hazard Ratio (CI)
AFP always <200	–	–
Max AFP 200–1000 to Final AFP <200	0.004	2.05 (1.26–3.34)
Max AFP >200–400 to Final AFP >200	<0.0001	3.34 (1.76–6.44)
AFP > 400–1000 to Final AFP >200	<0.0001	4.24 (2.57–6.98)
Max AFP >1000 to Final AFP < 1000 (>50% response)	0.036	2.07 (1.05–4.08)
AFP Always > 1000	<0.0001	7.28 (4.68–11.3)

Significant differences are observed at every response level.

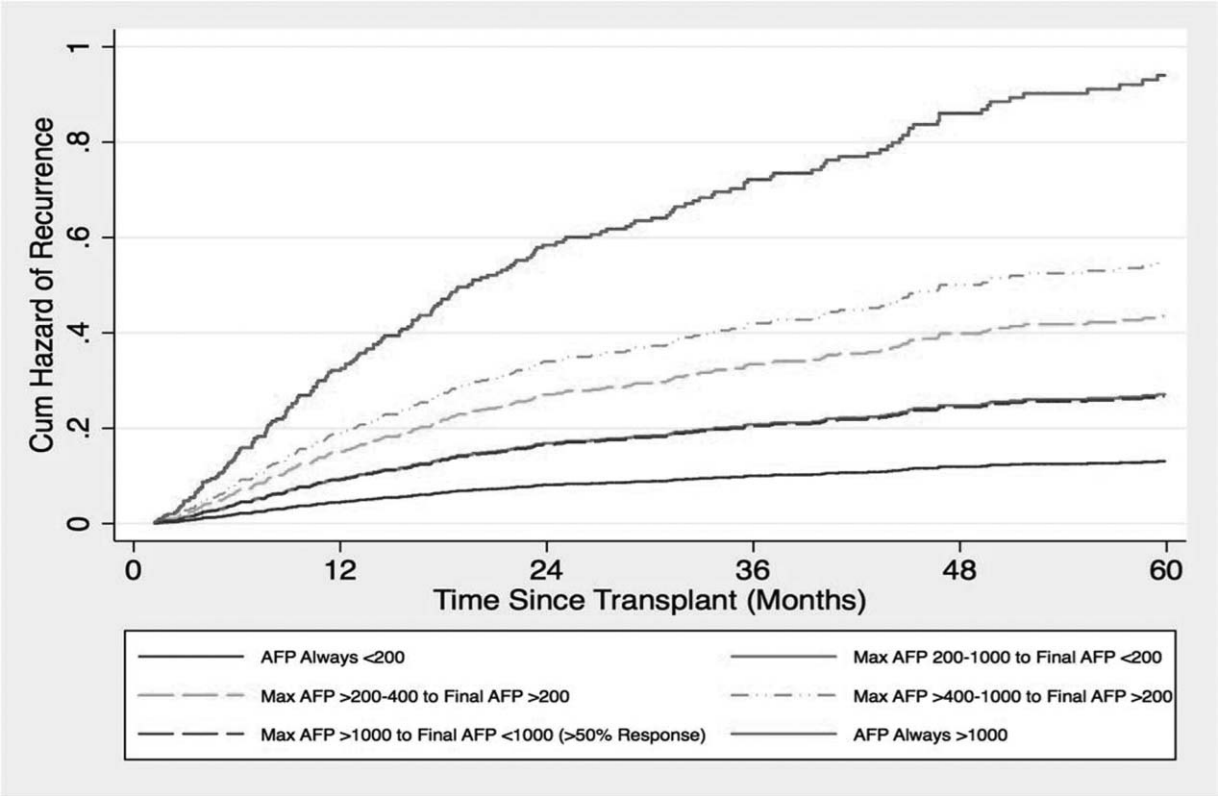


FIGURE 2. Cumulative hazard of recurrence by degree of response from Max AFP to Final AFP.

TABLE 3. Results of Univariable and Multivariable Cox regression Analysis to Analyze Predictors of Poor 5-year RFS as well as Point Allocation for NYCA Score

	Univariable Analysis		Multivariable Analysis		
Factor Affecting 5-y RFS	***P***	**Hazard Ratio (CI)**	***P***	**Hazard Ratio (CI)**	**NYCA Score Points**
Maximum tumor size at diagnosis, cm					
0–3 (Ref)	–		–	–	0
>3–6	<0.0001	2.08 (1.54–2.80)	<0.0001	1.96 (1.42–2.68)	2
>6	<0.0001	4.29 (2.56–7.19)	<0.0001	3.66 (2.13–6.34)	4
Maximum tumor number at diagnosis					
1 (Ref)	–	–	–	–	0
2–3	0.003	1.56 (1.15–2.13)	0.002	1.63 (1.19–2.26)	2
≥ 4	<0.0001	3.96 (2.36–6.63)	<0.0001	4.24 (2.47–7.27)	4
AFP response (Max AFP to Final AFP)					
AFP Always <200	–	–	–	–	0
Responders					
Max >200–1000 to Final <200	0.004	2.05 (1.26–3.34)	0.028	1.77 (1.07–2.96)	2
Max >1000 to Final <1000 (Must be >50% Drop)	0.036	2.07 (1.05–4.08)	0.107	1.75 (0.90–3.48)	2
Nonresponders					
Max >200–400 to Final >200	<0.0001	3.34 (1.76–6.44)	0.001	3.30 (1.67–6.55)	3
Max >400–1000 to Final > 200	<0.0001	4.24 (2.57–6.98)	<0.0001	3.82 (2.30–6.33)	4
Max >1000 to Final >1000	<0.0001	7.28 (4.68–11.3)	<0.0001	5.81 (3.63–9.31)	6
neutrophil-lymphocyte ratio ≥ 5	<0.0001	1.73 (1.26–2.37)	<0.0001	1.91 (1.38–2.66)	N/I

AFP indicates alpha fetoprotein; Max, maximum; N/I, not included.

TABLE 4. Recurrence Risk Stratification According to NYCA Score

NYCA Score	Recurrence Risk	5-y RFS
0–2	Low risk	90%
3–6	Acceptable risk	70%
7 or more points	High risk	42%

differing inflammatory states, as well as the availability of data only immediately before transplant. As were attempting to formulate an objective response for patients with HCC over time, we did not feel including the NLR into the score as a single available value would be of clinical utility and would make immediate applicability of the score as a selection tool difficult.

Competing Risk Regression and Overall Survival

Once the NYCA score was created, we formulated a competing risk regression model to assess the utility of the score, with death acting as a competing risk to recurrence. As shown in Fig. 3, there were a significant difference in the cumulative incidence of recurrence between the 3 NYCA categories on competing risk regression ($P < 0.0001$ for all 3 categories). The utility of NYCA at determining overall survival is shown in Fig. 4. Again, the NYCA score categories correlated with overall survival, with significantly worse overall survival for patients with higher NYCA scores (5-year overall survival NYCA Low-Risk 75% vs Acceptable Risk 62%, $P < 0.0001$, 5-year overall survival Low-Risk 75% vs High Risk 40%, $P < 0.0001$, 5-year overall survival Acceptable Risk 62% vs High-Risk 40%, $P = 0.001$).

ROC Analysis and Milan Recategorization

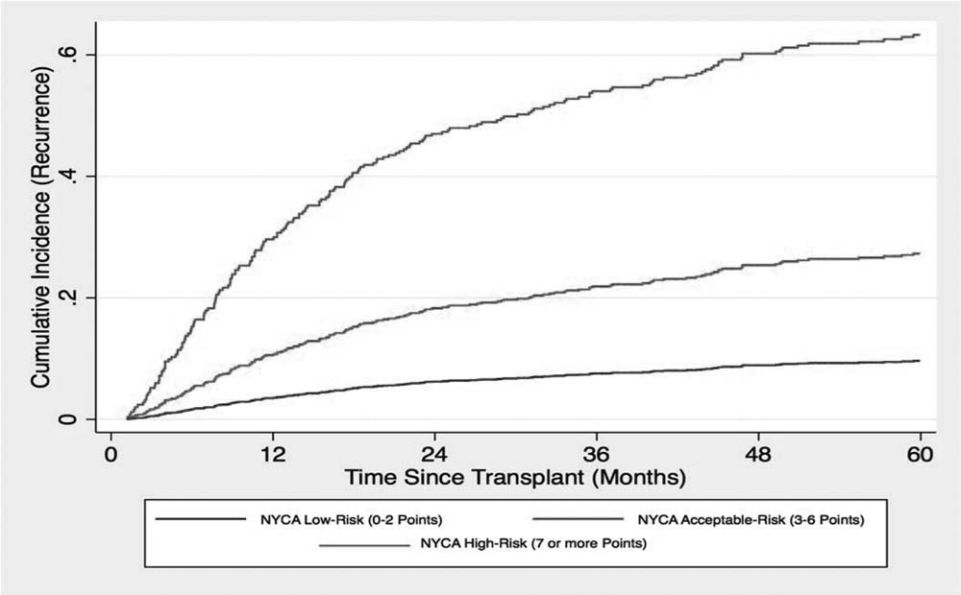
ROC analysis was then used to compare NYCA to the Milan criteria and the French AFP score in our patient groups. The C-statistics of Milan and French AFP were 0.612 and 0.658, respectively, compared with the C-statistic for NYCA, which was 0.731, which is significantly better than both Milan and French AFP at predicting 5-year RFS ($P < 0.0001$). Two hundred thirty-five patients were originally outside Milan at the time of diagnosis based on their imaging, and 201 (85.5%) would be recategorized into NYCA low or acceptable risk categories.

Correlation With Explant Pathology

The NYCA score correlates well with explant pathological tumor differentiation and vascular invasion. Almost half the patients (45.8%) in the NYCA high-risk group had poorly differentiated tumors compared with 10.7% of the NYCA Low-Risk group patients ($P < 0.0001$). Similarly, over 3 quarters (76.1%) of patients in the NYCA High-Risk group had evidence of vascular invasion on explant pathology, compared with 24.3% of patients in the NYCA Low-risk category ($P < 0.0001$). Correlations with explant pathology are summarized in Table 5.

DISCUSSION

This study introduces the first US-based prognostic score that incorporates an AFP response to predict HCC recurrence post-LT based on a multi-institutional patient cohort.



NYCA score	Cumulative Incidence of Recurrence Risk
0-2	7.0%
3-6	27.5%
7 or More points	62.5%

FIGURE 3. Competing risk regression analysis (with death as a competing risk to recurrence) by NYCA score category showing significant difference between the 3 NYC categories.

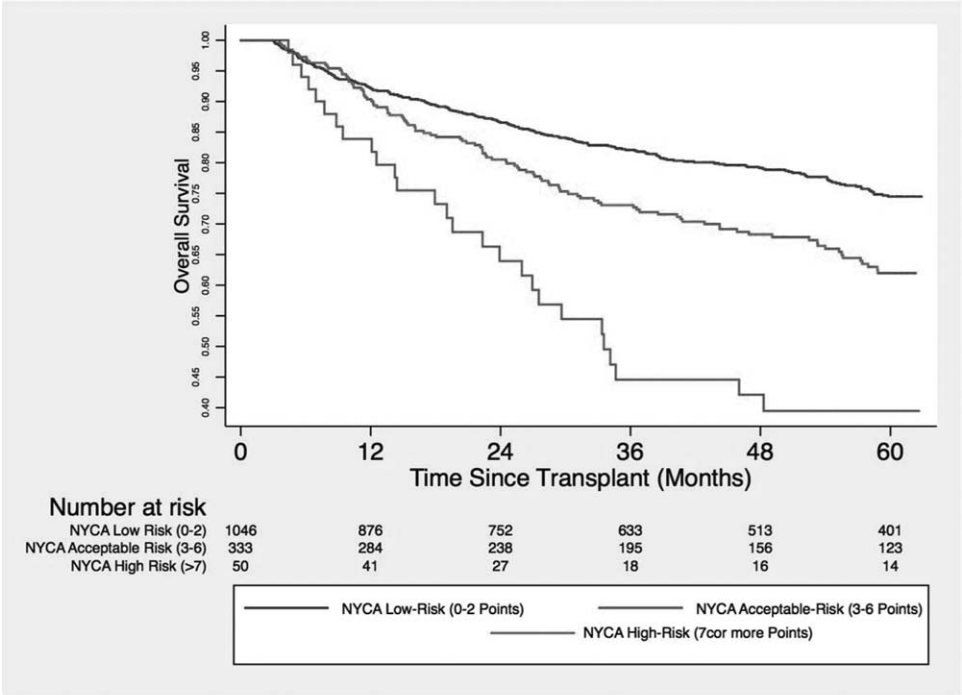


FIGURE 4. Kaplan-Meier curve and Log Rank testing according to NYCA score categories showing significant differences in overall survival between the 3 groups.

LOG-RANK TTESTING		
	NYCA Low Risk	**NYCA Acceptable Risk**
NYCA Low Risk	-	**p<0.0001**
NYCA Acceptable Risk	**p<0.0001**	-
NYCA High Risk	**p<0.0001**	**p=0.001**

Although the Milan criteria have been the standard for both clinical and administrative management of LT of HCC, they are dichotomous and lack true tumor biological indices included in patent selection. This has resulted in alternate criteria being proposed that include surrogates of tumor biology, both radiological[4,5] and biochemical.[6–8] The practicality and reliability of AFP as a useful predictor of tumor biological behavior, propensity to recurrence, as well poor outcome in HCC patients[17–19] has resulted in its inclusion in recent scoring systems, including the Metroticket 2.0 and the French AFP model.[11,20] However, both of these models fail to take into account the kinetics of AFP and the response to therapy, while patients are on the list and utilize only the baseline AFP at the time of listing.

NYCA incorporates the radiographic indices traditionally used for HCC patient selection, with the response from Max-AFP to Final-AFP, thereby allowing objective biological risk stratification, identifying the patients with the highest risk of recurrence and poor outcome (NYCA High-Risk). Although a 39% 5-year survival is good for cancer therapy, one can argue that this group should be excluded from transplantation or at least have restricted access to deceased donor livers. Most important is that NYCA allows the successful transplantation of patients who by other criteria (Milan, French AFP, Metroticket 2.0) would be wrongly considered too high risk either based on morphological tumor characteristic or high initial AFP levels. A total of 235 patients in our study group were outside of Milan at diagnosis and would have been potentially excluded based on size and/or number by some centers, and of these 201 (85.5%) would be reclassified into Low-Risk (n = 31) or Acceptable-Risk (n = 170) groups and would have 5-year RFS in excess of 69%. Two other groups that would have prohibitive 5-year RFS by other scoring systems include patients with Max-AFP levels >1000 ng/dL with a fall to a Final-AFP <1000 (with a > 50% response) who are categorized into the Low-Risk group (small <3 cm, single tumors n = 13), and Acceptable-Risk group (n = 32, no tumors >6 cm and/

TABLE 5. Correlation of NYCA Score Category With Explant Pathology, Namely Proportion of Necrotic Nodules, Poorly Differentiated Tumor and Vascular Invasion, Showing Significantly More Poorly Differentiated Tumors With Vascular Invasion in the NYCA High-Risk Group and Significantly Higher Proportion of Necrotic Nodules in the Low-risk Group

	NYCA Category			
Explant Pathology	**Low Risk**	**Acceptable Risk**	**High Risk**	***P***
% Necrotic nodules	20.6%	16.1%	10.4%	<0.0001
% Poorly differentiated	10.7%	24.8%	45.6%	<0.0001
% Vascular invasion	24.3%	38.2%	76.1%	<0.0001

or not more than 3 lesions). These patients have 5-year RFS of 74% and 91%, respectively, according to NYCA and would exclude purely due to high AFP levels at diagnosis by the other scoring systems. This highlights the fact that using an initial AFP of >1000 as an absolute contraindication to transplantation, as suggested by current UNOS criteria and other literature,[14,15] may result in the exclusion of a subset of patients that would benefit from LT and potentially be cured. It also likely explains the reason why NYCA has a significantly higher C-statistic for predicting 5-year RFS than the French AFP score (0.73 vs 0.65) in our patients. In fact, reassessing 5-year RFS based on French AFP using Final-AFP instead of Listing-AFP increases the C-statistic of the French AFP score to 0.70, again highlighting the importance of AFP response on outcome of patients.

From an oncologic standpoint, measuring an AFP response rather than using a single value at listing makes clinical sense, especially in the US where patients currently wait a minimum of 6 months before receiving MELD exception points. Time has been shown repeatedly to influence outcome in patients being transplanted for HCC, and is an excellent surrogate of tumor biology, as patients with aggressive tumors progress rapidly on the waitlist and have been shown not to benefit from early transplantation .[10,21,22] While we did not specifically assess the impact of waiting time in this study, one of the strengths the current work is that all 3 centers are located in long-waiting time regions (region 5 and 9) where the median wait time from listing to transplantation exceeds 8 months (after attaining MELD exception points) for patients with HCC,[21] and HCC patients are rarely, if ever, transplanted within 2 to 3 exception point cycles.[23] This study also provides further support for liver-directed neoadjuvant HCC therapy of patients on the waiting list and identifying subjects with good treatment responses. The association between a large reduction in AFP after treatment and the post-transplant recurrence risk is dramatic and provides powerful evidence for both the treatment strategy and the use of kinetic AFP in allocation. The data support a structured approach to pretransplant cancer care with constant reassessment of tumors during the listing period. NYCA would allow for an objective assessment of the risk of recurrence by incorporating a clinically useful AFP response algorithm that can be easily implemented immediately in any center using laboratory values already being routinely measured.

One of the issues plaguing many of the proposed HCC scoring systems for liver transplant selection is complexity resulting in difficult applicability and clinical utility. Similar to Milan, the French AFP score is both simple and clinically useful and therefore easy to adopt.[11] We believe NYCA to be equally easy to implement, with superior utility by allowing for a continual assessment of patients while on the list, adding predictive data to the radiologic information obtained on a 3-monthly basis by most centers, without complicating patient selection. As we use the Final-AFP before transplant in the response, we would advocate assessment of AFP on a monthly basis when the tumor MELD score reaches a transplantable level in respective regions in the US. Patients with rising AFP levels at this juncture (above a 200 ng/dL threshold) or high stagnant levels should have radiologic reassessment and no further exception points until levels decline, ideally to <200 ng/dL, or to <1000 with > 50% for those with levels >1000. Carefully scrutinizing these patients and preventing rapid transplantation in scenarios where AFP stops responding, starts to climb, or never responds would potentially allow centers to avoid transplantation of patients with biologically aggressive tumors, and decrease recurrence rates.

While we applaud UNOS for finally moving away from Milan and including AFP in its selection algorithm for patients undergoing LT for HCC, incorporating NYCA with an AFP response parameter into the selection algorithms is crucial in order to identify patients with poor tumor biology and more importantly, those with tumors that have responded to appropriate treatment before transplantation. There have been several changes to HCC selection algorithms in the US over the past 16 years since adopting Milan, many due to the ongoing analysis of outcomes relative to non-HCC patients or other HCC patients, the most recent change having been implemented in October 2015 to place a 6-month waiting period before receiving MELD exception points. The liver transplant community has done well to adapt to these changes. However, at this juncture, allocation policies in general in the US are undergoing drastic changes with the incorporation of the National Review Liver Board to allocate exception points for HCC. As such, it is perhaps more important than ever to get our HCC selection algorithms right. Given the data presented herein, it seems logical to incorporate a scoring system such as NYCA that takes AFP response into account into patient selection. Failure to do so will result in further short-term changes to the system that can be disruptive and difficult to incorporate.

One of the major limitations of our study is that it lacks information on patients who were delisted due to tumor progression that would be crucial to further understanding the AFP response. Assessing the NYCA score in all listed patients in an intent-to-treat analysis will assess the ability to predict waitlist drop-out and overall clinical utility. There is likely an ascertainment bias in this study, as all patients were transplanted and thought to have appropriate tumor biology by the listing center. A further limitation is that we did not incorporate a specific timeline with respect to AFP response and locoregional therapy, and we did not incorporate a marker of inflammatory response. The NLR has been shown by several groups to be important immediately before transplant and again our study showed it to be an independent predictor of RFS. However, we did not assess NLR over time and as such did not incorporate it, as we felt that the understanding of how NLR is affected by locoregional therapy, rising MELD or other recipient factors unrelated to the HCC is poor, and incorporating it into our score would make its applicability less likely.

CONCLUSION

NYCA provides an accurate objective measure of HCC outcome through incorporation of an AFP response, predicting both 5-year RFS, overall survival, as well as correlating with explant pathology. As UNOS moves to abandon the strictly dichotomous Milan criteria, adding static AFP levels, incorporation of an AFP response using NYCA into the current UNOS model would further advance our goal of offering transplantation to patients that would benefit most through a better understanding of biological HCC behavior.

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;334:693–699.
2. 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;33:1394–1403.
3. Ito T, Takada Y, Ueda M, et al. Expansion of selection criteria for patients with hepatocellular carcinoma in living donor liver transplantation. *Liver Transpl*. 2007;13:1637–1644.
4. Onaca N, Davis GL, Goldstein RM, et al. Expanded criteria for liver transplantation in patients with hepatocellular carcinoma: a report from the International Registry of Hepatic Tumors in Liver Transplantation. *Liver Transplant*. 2007;13:391–399.
5. Mazzaferro V1, 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;10:35–43.
6. 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;62:158–165.

7. Agopian VG1, Harlander-Locke M1, Zarrinpar A1, 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;220:416–427.
8. Halazun KJ1, Najjar M, Abdelmessih RM, et al. Recurrence after liver transplantation for hepatocellular carcinoma: a new MORAL to the story. *Ann Surg*. 2017;265:557–564.
9. Sapisochin G1, Goldaracena N1, Laurence JM2, et al. The extended Toronto criteria for liver transplantation in patients with hepatocellular carcinoma: a prospective validation study. *Hepatology*. 2016;64:2077–2088.
10. Lai Q1, Nicolini D, Inostroza Nunez M, et al. A novel prognostic index in patients with hepatocellular cancer waiting for liver transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) score. *Ann Surg*. 2016;264:787–796.
11. 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;143:986–994.
12. Peng SY, Chen WJ, Lai PL, et al. High alpha-fetoprotein level correlates 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;112:44–50.
13. 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;48:1118–1127.
14. Hameed B, Mehta N, Sapisochin G, et al. 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;20:945–951.
15. Available at: https://optn.transplant.hrsa.gov/media/1922/liver_hcc_criteria_-for_auto_approval_20160815.pdf Accessed March 26, 2018.
16. Bavry AA, Kumbhani DJ, Gong Y, et al. Simple integer risk score to determine prognosis of patients with hypertension and chronic stable coronary artery disease. *J Am Heart Assoc*. 2013;2:e000205.
17. DuBay D, Sandroussi C, Sandhu L, et al. Liver transplantation for advanced hepatocellular carcinoma using poor tumor differentiation on biopsy as an exclusion criterion. *Ann Surg*. 2011;253:166–172.
18. Tyson GL, Duan Z, Kramer JR, et al. Level of α-fetoprotein predicts mortality among patients with hepatitis C-related hepatocellular carcinoma. *Clin Gastroenterol Hepatol*. 2011;9:989–994.
19. Vibert E, Azoulay D, Hoti E, et al. Progression of alphafetoprotein before liver transplantation for hepatocellular carcinomain cirrhotic patients: a critical factor. *Am J Transplant*. 2010;10:129–137.
20. 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;154:128–139.
21. Halazun KJ, Patzer RE, Rana AA, et al. Standing the test of time: outcomes of a decade of prioritizing patients with hepatocellular carcinoma, results of the UNOS natural geographic experiment. *Hepatology*. 2014;60:1957–1962.
22. Mehta N, Heimbach J, Lee D, et al. Wait time of less than 6 and greater than 18 months predicts hepatocellular carcinoma recurrence after liver transplantation: proposing a wait time ''sweet spot''. *Transplantation*. 2017;101:2071–2078.
23. Roberts JP, Venook A, Kerlan R, et al. Hepatocellular carcinoma: ablate and wait versus rapid transplantation. *Liver Transpl*. 2010;16:925–929.

DISCUSSANTS

Dr Kim Olthoff (Philadelphia, PA):

Thank you for the opportunity to review your manuscript ahead of time and the opportunity to comment. I am also glad to see you return to this forum to continue to help us find a better way to allocate livers for patients with HCC using this bicoastal approach, which is great.

Because of the limited resources of deceased donor liver, some sort of system of prioritizing and allocating livers needed to be put in place. And when the MELD system was implemented, as you stated, we had to decide which patients would get exception points, and we chose the Milan criteria, granted that it was based on a very small group of patients. Although this has served an important role in deceased donor allocation with excellent results, as you stated, it put some artificial limitations in place. And with living donor transplantation, while we do not have to rely on allocation rules, we also owe donors the responsibility of making sure that our outcomes are reasonable before putting them through such a huge operation.

As you stated, there is plenty of evidence that excellent outcomes can be accomplished for patients with HCC outside of Milan, and many groups, including yours, have proposed expanded standard criteria, with our Asian centers and our colleagues in Toronto really pushing the envelope, doing patients way outside of Milan using liver donor transplants.

You are now trying to go beyond radiologic criteria and use AFP as a serum marker, and its dynamic changes, with either locoregional therapy or just on the wait list to serve as a surrogate for biologic behavior, and a simple scoring system, which hopefully can be utilized to better allocate livers. And I truly applaud your efforts because I do think we need to do something like this.

You have outlined some of the limitations, and I think it does deserve a prospective validation.

Now I have a few questions.

You chose very specific AFP cutoffs, all of which have been published to have previously correlated with outcomes, and you chose very specific tumor size cutoffs. But we also know that someone with an AFP of 199 would be very similar to someone with an AFP of 201, and someone with an AFP of 201 may be very, very different from someone with an AFP of 999. Similarly, for tumor sizes, 2.9 cm can be very similar to someone with a 3.1-cm HCC.

So, it seems, with the size of your database, you may have missed an opportunity to evaluate these in a more rigorous fashion – either looking at it as a continuous variable (increased HR for some increased increment) or performing ROC curves to predict the best cutpoints to predict recurrence, or smaller groups. I know you wished to keep it simple, but even if you add these small complexities, it could be plugged into some app like we do with MELD to make it simple for the user. I would like you to comment on why you did not do that.

A second question is, I was wondering why you did not use wait times or response to locoregional therapy. The authors in this group have published both. Both of those have significant impact on post-transplant survival. Also, as you did have quite a few patients who were outside Milan before looking at pathologic diagnosis, you must have had some way to transplant them without exception points either using living donors or expanded criteria donors. And I know, all these centers are very aggressive centers. So, you probably have a relatively broad range of wait times that could have been utilized and maybe added into this calculation, and it is a very easy data point to get. So, I was wondering why you did not put that in your model.

My final question relates to sort of a practicality of it and the implementation. You state you wish to include this in the UNOS allocation model, and I would like to understand how you would do this? Would you give extra exception points and how much to the lowest category, or would you use it to the exclude patients, excluding those who have high risk and take them out of the running for a liver transplant?

So, as you continue to try to refine how livers are allocated in a very fair, just, and prudent fashion, studies like these are extremely important, so we do not do futile transplants, but we also do not exclude patients who could do very well with transplantation.

I applaud your efforts. And thank you very much for the opportunity to comment.

Dr Karim J. Halazun:

Thank you Dr Olthoff for your thorough review, great question, and kind words. I am going to try to answer your questions in sequence.

You are right, we did not use continuous variables, and many other people, including Dr Agopian, who is on the paper, used continuous models to produce nomograms, which frankly have not been adopted because of complexity. One of the reasons Milan has been adopted is because 1 under 5 and 3 under 3 is very easy remember, and people are able to use it very readily.

We did try to incorporate AFP as a continuous variable. The problem is that the range of AFP is between 0 and upwards of 100,000. So, the incremental differences of 1-dL change of AFP does not really translate into something you can put into a score easily, and it becomes a logarithmic equation that is very, very difficult to use, and probably ends up excluding more patients than you want simply because of complexity. Using a logarithmic mathematical equation is a good thought, and I think maybe someone with more advanced mathematical and statistical knowledge than I can advance such a score. But I think keeping it simple and relatively objective allows to us look at it in a specific way.

The cutoffs that are around are tried and tested, and we tried not to steer away too much from the morphometic criteria that already exists because people have tried to steer away to try to incorporate these different morphometric criteria, have not really achieved any change, except for France where they just stuck to it. They used the logarithmic equation to calculate the AFP but then reverted to absolute numbers to make it easy to understand. So, I hope that answers that question.

In terms of time, all 3 of our centers are in Region 5 and 9. The median waiting time for transplantation for HCC for any of our patients is about 7 to 8 months. So, it was implied that those patients waited longer. Now, I agree with you wholeheartedly that when we validate the score, and a multicenter validation is underway, we need to include time, because not all centers have equal waiting times, and some centers outside of our region transplant HCCs as soon as they are listed. So, there has to be a time component to understand how better AFP is calculated and how the changes affect outcome.

I think the last question is the most important question, how to implement this. And I personally have a lot of thoughts how to implement it. And you are right, we are very big proponents of the waiting time model. But we already have a waiting time model where 6 months is allocated automatically to everybody who is waiting. I heard today that Region 5 has an all-comers policy where everybody can get listed no matter how big their HCC is. So, perhaps that is the way to do it, list everybody, except those with metastatic disease. And at the time when they are about to get their exception points, calculate the NYCA score, and if the NYCA score is high, do not give them exception points. And only give them exception points is their NYCA score is acceptable.

Now, we have to remember that we are talking about accessibility at the high-risk group where the 5-year overall survival is 40%. If we were in a pancreas cancer conference, everybody would say, why are you denying these patients?

I would think with the presentation earlier where we are seeing xenotransplantation is a real thing, and we are going to be transplanting pig livers at some point in my lifetime, at least, maybe all this does not matter, but I think that is one way to implement it.

Dr Goran Klintmalm (Dallas, TX):

This was a great presentation. And Kim covered most of my questions, actually.

I think when we look at these things, what we have to look at is the graft utility, and going forward, talking to UNOS, I am trying to make UNOS make changes.

I think we need to focus on who should we exclude from transplant? It is not about include. We have more patients than we actually can transplant. And the key is to avoid putting livers into patients who have a 75% likelihood of recurrence, as an example. And I think if we have that mind-set, we will approach your proposal in a slightly different way, but maybe more importantly to employ an operation like this only in patients who truly can benefit.

Thank you.

Dr Karim J. Halazun:

I think you are right. I think we have not figured out who those patients are exactly that are going to have a 75% recurrence rate. And we have all transplanted patients with small tumors with a low AFP that within a month have 20 tumors. So, perhaps its Dr Agopian's circulating tumor cells? I do not know what the next answer is. But we have not figured out a good model to predict who those super high-risk patients are. This is the best, I think, that we have.

Dr Goran Klintmalm (Dallas, TX):

I agree. But I think we have to have that in mind as we develop this further. Thank you.

Dr M.B. Majella Doyle (St. Louis, MO):

Thank you, Karim. This is excellent work. We too have taken a very aggressive approach to HCC, in St. Louis.

We are in Region 8, and at our center, we have no limitations for patients with HCC on who we will consider for downstaging. All patients who successfully downstage to with in Milan will get exception points, even those patients with major vascular involvement at diagnosis. We have shown that these far beyond Milan patients have similar outcomes to those within Milan. So, my question to you is how are these patients going to fare with the national review board that we know is coming later this year? Given that so many centers are conservative in their approach to HCC, are we going to have more denials of exception for patients with HCC beyond Milan criteria who many of us believe would do well with transplant?

Dr Karim J. Halazun:

I think that is a great question. So, we are moving to a system where we are no longer going to be regionally allowed to award execution and we are going to be mandated as to who we can list in by a national review board, and there are going to be people on that board who are going to say, Milan in, yes, and Milan out, no. A lot of your patients will suffer.

I think one of the problems is that patients with macrovascular invasion who have been successfully downstaged are a few, and experience with them has been limited to anecdotal outcomes in certain centers, and nobody is publishing that data. I think it behooves us to move forward and publish this quickly and talk about it more openly so that everybody understands that outcomes can be excellent.

The other thing that I think needs to happen is we have to make sure that the people on the national review board are people from centers like yours that understand that these patients can do just fine. And if we do not do that, then its very risky. We are going to lose a lot of patients who we are going to end up not transplanting who could be completely cured.

Dr Debra Sudan (Durham, NC):

Thank you for your presentation. You have done an excellent job of explaining it. I think one of the things that you missed with your discussion, however, is that you are only looking at 1 side of the coin. I think that there is no doubt that from an oncologic perspective a recipient is expected to have an extremely good outcome from LT if

within or close to Milan criteria, but even if you have a recurrence rate of up to 60%, because the alternatives are so dismal.

The problem is that the wait list mortality for patients with HCC is far less than those with chronic liver disease without HCC, as it stands currently and in fact UNOS has added 6 months waiting time before exception is instituted in order to address this. By giving exceptions, we appear to have artificially driven up MELD scores for liver transplant candidates across the nation pretty significantly such that we have actually increased our waiting time and mortality for those with chronic liver disease without HCC.

So, my question to you is, how do we determine which patients with HCC, or how many patients with HCC we prioritize over those with chronic liver disease without HCC given the limited resource that we currently have? So, until we have alternatives to the current cadaveric donor pool so that we can transplant all of these potential candidates who could benefit, how do we prioritize HCC and abandon the Milan criteria?

Dr Karim J. Halazun:

Again, I can give you my thoughts about this. I have a lot of thoughts. I think your question is perfect. We are probably transplanting too many patients with HCC. I think we have proven time and time again that these patients do very well if you put marginal livers in them, and we do not do that collectively as a community.

I personally think there should be 2 lists, a list of patients just with HCC and will accept marginal livers, where if your outcome is poor because you took an 85-year-old liver, you do not get penalized as heavily as if you take a 20-year-old liver. On the other list, you should have patients with chronic liver disease to accept the good livers.

Now, we are moving to an era where marginality is becoming commonplace, so we have to redefine marginality to figure out how to do that. But I strongly believe that all of these livers that are marginal, or that are old, that get discarded at a very, very high rate can be used in these patients successfully and we would not hurt anyone.

We have moved toward doing 10,000 liver transplants in the U.S. We are almost there. I think last year was 8700. Thanks to the opioid epidemic. Opioids, however, will go away at some point. But if we prevent discard of marginal livers and preferentially put these livers in cancer patients, I do not see that we are going to have a problem long-term. I actually think we are going to solve the problem that we have right now.