R3-AFP score is a new composite tool to refine prediction of hepatocellular carcinoma recurrence after liver transplantation

Charlotte Costentin, Federico Piñero, Helena Degroote, Andrea Notarpaolo, Ilka F. Boin, Karim Boudjema, Cinzia Baccaro, Luis G. Podestá, Philippe Bachellier, Giuseppe Maria Ettorre, Jaime Poniachik, Fabrice Muscari, Fabrizio Dibenedetto, Sergio Hoyos Duque, Ephrem Salame, Umberto Cillo, Sebastian Marciano, Claire Vanlemmens, Stefano Fagiuoli, Patrizia Burra, Hans Van Vlierberghe, Daniel Cherqui, Quirino Lai, Marcelo Silva, Fernando Rubinstein, and Christophe Duvoux for the French-Italian-Belgium and Latin American collaborative group for HCC and liver transplantation

Table of contents

Statistical Analysis

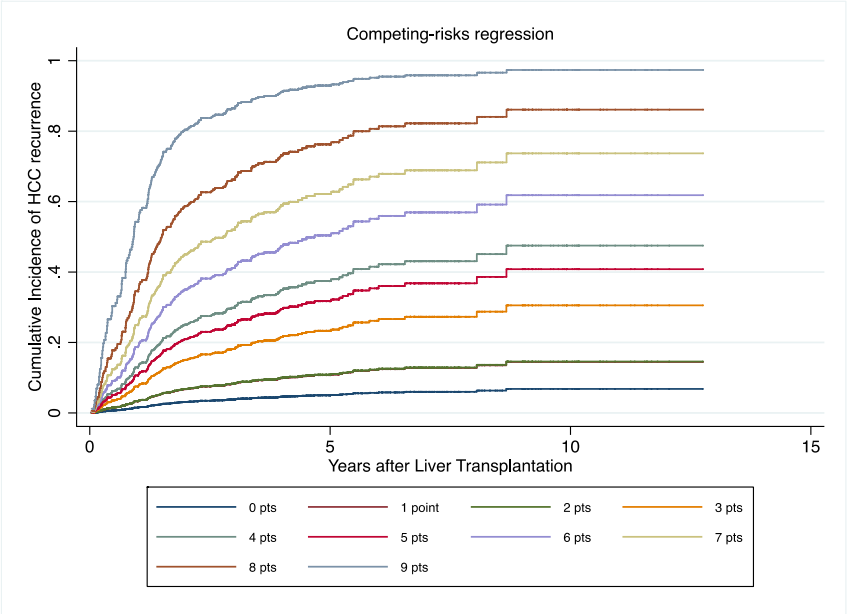
We conducted multivariable competing risk regression models for the primary outcome (HCC recurrence), estimating sub-distribution of hazard ratios (SHR) and 95% confidence intervals (95% CI) using the Fine and Gray method (1). Any cause of death prior to HCC recurrence was considered a competing event. In the TC, we assessed variables independently associated with HCC recurrence. Variables with a P value <0.05 after univariate analysis (Wald test) were included in the multivariable analysis by stepwise forward elimination considering confounding effect (>20% of change in crude HR) and following the "1 variable per 10 events" rule to avoid overfitting (1). For the construction of the final predictive score, points were assigned dividing each SHR with the lowest SHR observed from the final multivariable competing risk model. A stratification risk assessment was performed according to the observed incidence of 5-year HCC recurrence based on the cumulative scoring model in different stratum. To further validate the final model, all variables included in the final model in the TC were evaluated in the VC. SHR and their 95% CI in the VC were numerically compared to those in the VC. Additionally, the same stratification risk assessment in the TC was conducted in the VC.

Each model's performance was explored, including calibration and discrimination. Calibration was assessed comparing observed and predicted risk curves and discrimination with Wolber's adapted c-statistics (2). The new model was compared with prior explant models including tumor pathology features (Milan and the Up-to-7 criteria), and with the RETREAT score, a composite model including explant features and AFP values at last pre-transplant evaluation. Potential cut-offs for the final model were identified using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for exploring the best model of fit. Finally, the net reclassification index (NRI) considering competing events was estimated comparing selected thresholds to evaluate the re-categorization of risk (3, 4).

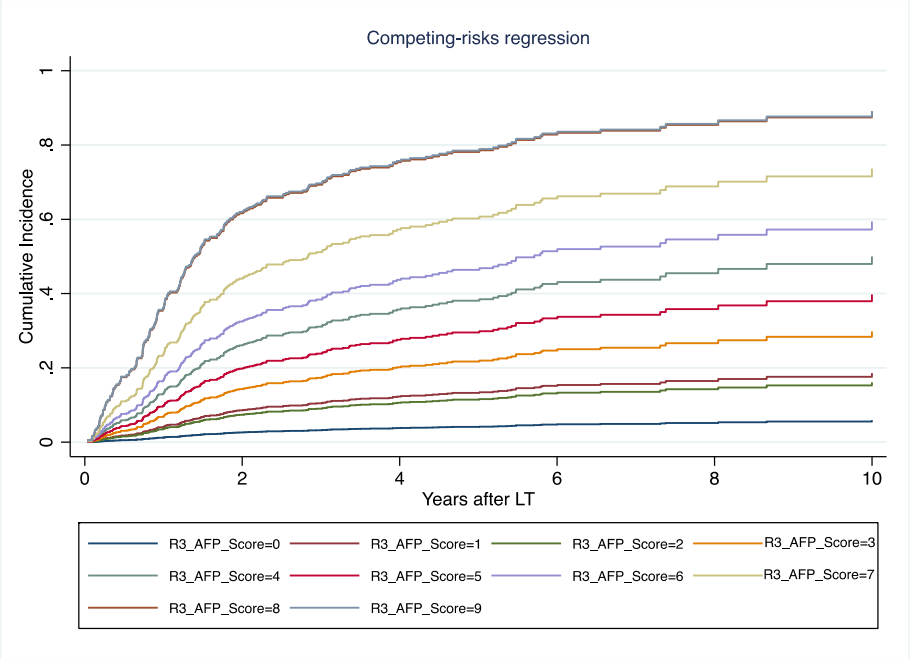
External validation of the Recurrence Risk Reassessment scores

Multivariable competing risk regression analysis including the exposure variables in the R3 score was also tested in the validation cohort (Supplementary Table 6). In the model limited to explant features, nuclear grade adjusted for the other independent variables was not significantly associated with recurrence in the VC. However, the Wolber's c-index for the R3 explant-based model was 0.73 (95%CI 0.67-0.79). When exploring the effect of last AFP value in the VC, nuclear grade was also not independently associated with recurrence, adjusted for tumor number and largest diameter, and AFP values (Supplementary Table 7). However, the R3-AFP model in the VC had a Wolber's c-index of 0.78 (95%CI=0.73-0.83), which outperformed that from the original R3 score without AFP (P=0.018).

Supplementary Figure 1: 5-year HCC cumulative recurrence according to R3 score integer value in the test cohort



Supplementary Figure 2: 5-year HCC cumulative recurrence according to R3-AFP score integer value in the test cohort



Supplementary Figure 3: Expected vs. Observed cumulative incidence curves of events according to R3-AFP score categories

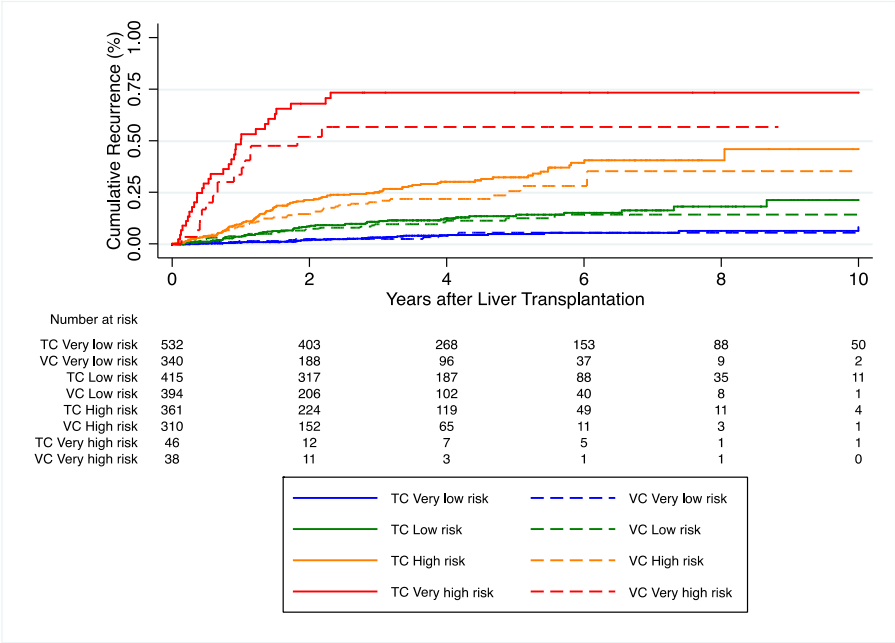


Table S1: main features of the training and validation cohorts

Variable Name	**Training cohort**	**Validation cohort**	***P***
Number of patients n	1359	1085	
Number of LT centers n	30	17	
Year, Period of LT n (%)			
2000-2005	572 (42.1)	22 (2.0	<0.0001
2006-2011	610 (44.9)	368 (33.9)	
2011-2018	177 (13.0)	695 (64.1)	

Note: Patients in the TC were included in 20 centers from France (n=352), 4 from Italy (n=481) and 6 centers from Belgium (n=526). Seventeen LT centers participated in the Latin American VC, including 5 centers from Argentina (n=325), 3 centers from Brazil (n=376), 3 from Chile (n=90), 2 from Colombia (n=157), 2 from Mexico (n=63) and 1 center from Ecuador (n=13), Peru (n=26) and Uruguay (n=35).

Table S2. Findings at explant pathology analysis.

VARIABLE	*Test cohort (n=1359)*	*Validation cohort (n=1085)*	*P*
Number of HCC nodules			
1-3 nodules	1005 (73.9)	911 (84.0)	<0.0001
≥4 nodules	354 (26.0)	174 (16.0)	
Largest nodule diameter			
≤3 cm	849 (67.1)	633 (59.2)	<0.0001
3-6 cm	361 (28.6)	398 (37.3)	
>6 cm	54 (4.3)	38 (3.5)	
Complete major nodule necrosis, n (%)	94 (6.9)	11 (1.0)	<0.0001
Presence macrovascular invasion, n (%)	52 (3.8)	0	<0.0001
Presence microvascular invasion, n (%)	369 (27.1)	249 (22.9)	0.017
Tumor differentiation, n (%)			
Nuclear grade I-II	1003 (85.3)	739 (71.2)	<0.0001
Nuclear grade >II	173 (14.7)	299 (28.8)	
Within Milan, n (%)	847 (62.3)	721 (66.4)	0.034
Within Up-to 7 without mvi	832 (61.2)	737 (67.9)	0.001
Within Up-to 7 with mvi	212 (15.6)	169 (15.6)	
Beyond Up-to 7 without mvi	158 (11.6)	99 (9.1)	
Beyond Up-to 7 with mvi	157 (11.5)	80 (7.4)	
RETREAT score, n (%)			
≤3 points	981 (72.3)^	770 (71.6)^	0.28
>3 points	376 (27.7)^	306 (28.4)^	

Note: Test cohort (European cohort). Validation cohort (LATAM cohort). ^At last pre-transplant evaluation. Each patient was classified according to Milan criteria (4) and the Up-to-7 criteria (5) based on explant pathology features, both as post-LT models for risk of recurrence, and the RETREAT score (11) based on last AFP values. Low and high risk of recurrence categories for each explant-based model were defined as within/beyond Milan, within Up-to-7 with or without microvascular invasion, and a RETREAT score equal or less than 3 points, as reported in the seminal publications.

Table S3. Development of the R3 score (without AFP values). Points assigned from the multivariable competing risk regression analysis in the test cohort.

VARIABLE	*Adjusted SHR (95% CI)*	*P*	*Points*
Number of nodules			
1-3 nodules (n=1005)			0
≥4 nodules (n=354)	1.81 (1.30-2.53)	<0.0001	1
Major nodule diameter			
≤3 cm (n=849)	-	-	0
3-6 cm (n=361)	1.91 (1.35-2.70)	<0.0001	2
>6 cm (n=54)	5.82 (3.60-9.39)	<0.0001	5
Microvascular invasion			
Absence (n=990)	2.70 (1.94-3.76)	<0.0001	0
Presence (n=369)			2
Nuclear grade >II			
Absence (n=1003)			0
Presence (n=173)	1.22 (1.02-1.46)	0.02	1

Note: Scoring model was done dividing each SHR with the lowest SHR observed (total of 9 points).
Median 1 point (IQR 0-3); SHR 1.51 (CI 1.43-1.60).

Table S4. Competing risk regression analysis evaluating AFP inclusion in exposure variables included in the original R3 score. Test cohort.

VARIABLE	*5-year recurrence rate (95% CI)*	*Unadjusted Sub-Hazard Ratio (95% CI)*	*P*	*Adjusted Sub-Hazard Ratio (95% CI)*	*P*
Number of nodules		1.03 (1.01-1.04)	<0.0001		
1-3 nodules (n=1005)	14.2 (11.7-17.1)				
≥4 nodules (n=354)	35.7 (29.4-42.9)	2.73 (2.07-3.59)	<0.0001	1.88 (1.34-2.64)	<0.0001
Major nodule diameter		1.37 (1.31-1.44)	<0.0001		
≤3 cm (n=849)	13.8 (11.1-17.1)	-	-	-	-
3-6 cm (n=361)	30.4 (24.5.37.7)	2.32 (1.72-3.14)	<0.0001	1.83 (1.29-2.59)	0.001
>6 cm (n=54)	74.5 (58.7-87.9)	9.21 (5.98-14.17)	<0.0001	5.82 (2.97-8.20)	<0.0001
Microvasc invasion					
Absence (n=369)	11.4 (9.2-14.0)				
Presence (n=990	39.6 (32.9-46.3)	4.04 (3.06-5.32)	<0.0001	2.69 (1.93-3.75)	<0.0001
Nuclear grade >II					
Absence (n=1003)	15.9 (13.3-19.0)				
Presence (n=173)	28.2 (21.2-36.9)	1.47 (1.23-1.74)	<0.0001	1.20 (1.01-1.43)	0.048
AFP (ng/ml)*		-	-	-	-
≤100 (n=1191)	16.9 (14.4-19.9)				
101-1000 (n=136)	36.9 (27.8-47.8)	2.49 (1.74-3.54)	<0.0001	1.57 (1.03-2.39)	0.035
>1000 (n=27)	45.9 (29.2-66.6)	4.64 (2.37-9.01)	<0.0001	2.83 (1.01-7.96)	0.049

Abbreviations: AFP: alpha-fetoprotein. HCC: Hepatocellular carcinoma. *Last available AFP values prior to LT (Median time from last AFP values to transplantation was 2.2 months (IQR 0.9-4.0 months). Calibration between observed/predictive was adequate and c-statistic (Wolber's c-index) was **0.76 (CI 0.72-0.80)**

Table S5. Explant pathology liver findings associated with HCC recurrence after liver transplantation in the validation cohort. Competing risk regression analysis.

VARIABLE	*5-year recurrence rate (95% CI)*	*Unadjusted Sub-Hazard Ratio (95% CI)*	*P*	*Adjusted Sub-Hazard Ratio (95% CI)*	*P*
Number of nodules		1.10 (1.06-1.13)	<0.0001		
1-3 nodules (n=911)	14.5 (11.1-18.8)				
≥4 nodules (n=174)	29.8 (21.5-40.3)	2.72 (1.81-4.08)	<0.0001	2.03 (1.32-3.12)	0.001
Major nodule diameter		1.38 (1.24-1.54)	<0.0001		
≤3 cm (n=633)	11.2 (8.0-15.6)	-	-	-	-
3-6 cm (n=398)	21.6 (15.5-29.7)	1.90 (1.25-2.89)	0.003	1.57 (1.02-2.41)	0.037
>6 cm (n=38)	61.2 (42.0-80.6)	8.61 (4.82-15.39)	<0.0001	5.27 (2.85-9.77)	<0.0001
Microvascular invasion					
Absence (n=825)	11.0 (8.0-14.9)				
Presence (n=249)	37.4 (27.7-49.2)	4.07 (2.78-5.95)	<0.0001	3.26 (2.07-5.12)	<0.0001
Nuclear grade >II					
Absence (n=739)	8.4 (6.5-10.6)				
Presence (n=299)	12.0 (8.6-16.3)	1.26 (1.03-1.54)	0.02	0.96 (0.76-1.21)	0.71

Abbreviations: HCC: Hepatocellular carcinoma. C-statistic (Wolber's index) was 0.73 (CI 0.67-0.79).
Nuclear grade was included in the final model to adjust the effect or explore confounding effect in regard

Table S6. Competing risk regression analysis evaluating AFP at last pre-LT assessment and exposure variables in the R3 score. Validation cohort.

VARIABLE	*5-year recurrence rate (95% CI)*	*Unadjusted Sub-Hazard Ratio (95% CI)*	*P*	*Adjusted Sub-Hazard Ratio (95% CI)*	*P*
Number of nodules		1.10 (1.06-1.13)	<0.0001		
1-3 nodules (n=911)	14.5 (11.1-18.8)	-	-	-	-
≥4 nodules (n=174)	29.8 (21.5-40.3)	2.72 (1.81-4.08)	<0.0001	1.89 (1.19-3.01)	0.007
Complete necrosis					
Yes (n=11)	0	-**	-		
No (n=1074)	17.1 (13.8-21.0)				
Major nodule diameter		1.38 (1.24-1.54)	<0.0001		
≤3 cm (n=633)	11.2 (8.0-15.6)	-	-	-	-
3-6 cm (n=398)	21.6 (15.5-29.7)	1.90 (1.25-2.89)	0.003	1.45 (0.93-2.26)	0.10
>6 cm (n=38)	61.2 (42.0-80.6)	8.61 (4.82-15.39)	<0.0001	5.19 (2.73-9.83)	<0.0001
Total tumor diameter		1.12 (1.08-1.16)	<0.0001		
Microvascular invasion					
Absence (n=825)	11.0 (8.0-14.9)	-	-	-	-
Presence (n=249)	37.4 (27.7-49.2)	4.07 (2.78-5.95)	<0.0001	2.66 (1.66-4.25)	<0.0001
Nuclear grade >II					
Absence (n=739)	8.4 (6.5-10.6)	-	-	-	-
Presence (n=299)	12.0 (8.6-16.3)	1.26 (1.03-1.54)	0.02	0.93 (0.73-1.18)	0.55
AFP (ng/ml)*					
≤100 (n=893)	12.7 (9.5-17.1)	-	-	-	-
101-1000 (n=147)	29.1 (20.2-40.7)	3.0 (1.93-4.66)	<0.0001	2.33 (1.43-3.79)	<0.001
>1000 (n=39)	49.3 (32.0-69.7)	6.40 (3.46-11.84)	<0.0001	4.53 (2.36-8.73)	<0.0001

Abbreviations: HCC: Hepatocellular carcinoma. Wolber's c-index was **0.78 (CI 0.73-0.83).**

***** Last available AFP values prior to LT (Median time from last AFP values to transplantation in the VC was 2.3 months (IQR 0.9-5.3 months)). ****Inviable to be included in a mathematical modelling due to absence of events.**

Table S7. Comparison regarding discrimination power of explant-based models in the test and validation cohorts.

VARIABLE	*Wolber's c-index (95% CI)*	*P values*	
	a-Test cohort		
	Comparison against explant models (not including AFP)		
Milan criteria	0.66 (0.62-0.69)	Ref	
Up-to 7 criteria without MVI	0.70 (0.67-0.73)	0.005	Ref
R3 score (without AFP)	0.75 (0.72-0.79)	<.0001	<.0001
	b-Validation cohort		
	Comparison against explant models (not including AFP)		
Milan criteria	0.66 (0.61-0.71)	Ref	
Up-to 7 criteria without MVI	0.68 (0.63-0.73)	0.56	Ref
R3 score (without AFP)	0.73 (0.67-0.79)	0.002	<.0001

Table S8: Locoregional treatment before LT. Comparison between cohorts.

VARIABLE	Test cohort (n=1359)	Validation cohort (n=1085)	*P*
Any treatment, n (%)	931 (68.5)	782 (72.1)	0.055
Type of locoregional treatment, n/patients receiving any treatment (%)	710/931 (76.3)	759/782 (97.1)	<.0001
TACE	322/931 (34.6)	29/782 (3.7)	<.0001
RFA	119/931 (12.8)	4/782 (0.5)	<.0001
PEI	103/931 (11.1)	4/782 (0.5)	<.0001
LR			
Locoregional treatment, n/whole cohort (%)			
Within Milan at listing	710/1039 (68.3)	660/939 (70.3)	0.35
Beyond Milan at listing	221/320 (69.1)	122/146 (83.6)	0.001
Locoregional treatment, n/whole cohort (%)			
AFP score ≤2 points at listing	841/1221 (68.9)	668/942 (70.9)	0.31
AFP score >2 points at listing	90/137 (65.7)	111/139 (79.9)	0.008
Locoregional treatment, n/whole cohort (%)			
AFP ≤100 ng/ml	823/1212 (67.9)	631/877 (71.9)	0.53
AFP 101-1000 ng/ml	96/129 (74.4)	117/165 (70.9)	
AFP >1000 ng/ml	12/18 (66.7)	31/39 (79.5)	

Abbreviations: TACE: transarterial chemoembolization; RFA: radiofrequency ablation; PEI: percutaneous ethanol ablation; LR: liver resection. Test cohort (European cohort). Validation cohort (LATAM cohort).

References

1. Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. Clin Cancer Res. 2012;18(8):2301-8.

2. Wolbers M, Blanche P, Koller MT, Witteman JC, Gerds TA. Concordance for prognostic models with competing risks. Biostatistics. 2014;15(3):526-39.

3. 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):114-21.

4. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30(1):11-2