A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) Score

ESA PAPER

A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation

Time–Radiological-response–Alpha-fetoprotein–INflammation (TRAIN) Score

Quirino Lai, MD, PhD,† Daniele Nicolini, MD, PhD,‡ Milton Inostroza Nunez, MD,* Samuele Iesari, MD,*†‡ Pierre Goffette, MD, PhD,* Andrea Agostini, MD,§ Andrea Giovagnoni, MD,§ Marco Vivarelli, MD,‡ and Jan Lerut, MD, PhD, FACS**

Objective: A novel and easy prognostic score based on the combination of pre-operatively available variables in patients with hepatocellular cancer (HCC) waiting for liver transplantation (LT) has been developed from a long waiting time (WT) training set and then validated in a short-WT set.

Summary of Background Data: The role of radiological response to loco-regional therapies, alpha-fetoprotein modification, inflammatory markers, and length of WT has been recently shown to be important selection criteria for the risk of intention-to-treat (ITT)-death and recurrence.

Methods: The training set consisted of 179 HCC patients listed for LT during the period January 2000 to December 2012 from the UCL Brussels Transplant Centre; the validation set consisted of 110 patients listed during the period January 2005 to December 2014 from the Ancona Liver Centre.

Results: The proposed Time–Radiological-response–Alpha-fetoprotein–INflammation (TRAIN) score was the best predictor of microvascular invasion. A TRAIN score $\geq$1.0 excellently stratified both the investigated populations in terms of ITT and recurrence survivals. When compared with Milan criteria, the proposed score allowed obtaining an increase of potentially transplantable patients (+8.9% in training set and 24.6% in validation set) without additive recurrence risks.

Conclusions: The proposed TRAIN score is an easy selection tool based on variables available before LT. This score enables the selection process to be refined in the 2 different scenarios of long and short WT. In case of longer WT, the score is better at predicting risk of death during the WT; in case of short WT, the score is better at identifying risk of post-LT recurrence.

Keywords: alpha-fetoprotein slope, drop out, fast-track approach, hepatocellular cancer, liver transplantation, mRECIST, neutrophil-to-lymphocyte ratio, recurrence, waiting time

(*Ann Surg* 2016;xx:xxx–xxx)

From the *Starzl Unit of Abdominal Transplantation, University Hospitals Saint Luc, Université catholique Louvain, Brussels, Belgium; †Department of General Surgery and Organ Transplantation, L'Aquila University, L'Aquila; ‡Unit of Hepatobiliary surgery and Transplantation, Azienda Ospedaliero-Universitaria "Ospedali Riuniti," Torrette Ancona, Italy; and §Department of Radiology, Azienda Ospedaliero-Universitaria "Ospedali Riuniti," Torrette Ancona, Italy.

Reprints: Dr Quirino Lai, MD, PhD, and Prof Jan Lerut, MD, PhD, FACS, Starzl Unit of Abdominal Transplantation, St. Luc University Hospital, Université Catholique Louvain - UCL, Avenue Hippocrate 10, 1200 Brussels, Belgium. E-mails: lai.quirino@libero.it; jan.lerut@uclouvain.be.

Authorship: QL and JL designed the research; QL, DN, MIN, SI, and AA collected the data; QL, MV, and JL wrote the paper; QL and DN performed the statistical analysis; PG and AG gave final approval of the article to be published.

Specific funding was not used to do this study.

None of the authors has any conflict of interest to declare in relation to this paper.

ISSN: 0003-4932/14/26105-0821

DOI: 10.1097/SLA.0000000000001881

Milan criteria (MC) remain the cornerstone in the selection process of hepatocellular cancer (HCC) patients waiting for liver transplantation (LT). Despite implementation in 1996, these criteria still dominate the process of wait-list registration worldwide.[1] Many attempts to widen MC have been reported, none of them being recognized as a valid substitute.[2,3] This is explained by the fact that most extended criteria are based on morphologic (number-diameter) tumor characteristics. Recently, biological parameters have been added, the most important being alpha-fetoprotein (AFP) levels,[4–6] radiological response to loco-regional treatments (LRT),[6,7] inflammatory markers,[8] and length of waiting-time (WT).[9,10]

This study aims to develop a prognostic score combining these variables able to predict intention-to-treat (ITT)-survival and recurrence rates. This "Time–Radiological-response–Alpha-fetoprotein–INflammation" (TRAIN) score has been developed using data of a training set (TS) with a long WT and an independent validation set (VS) with a short WT.

METHODS

A retrospective analysis of prospectively collected data from 2 European centers was performed after approval by their local Ethics Committees. One hundred seventy-nine HCC patients listed for LT during the period January 2000 to December 2012 in the UCL-Brussels center were analyzed to produce a TS of data. Their median follow-up was 3.4 years [interquartile range (IQR) = 1.0 to 7.1]. We evaluated the significance of the obtained TS results using an independent VS consisting of 110 patients listed for LT during the period January 2005 to December 2014 in the Ancona center. Their median follow-up was 2.3 years (IQR = 0.7 to 4.9). According to the median WT of the 2 centers, TS was defined as having a long WT, VS as having a short WT. No specific WT cut-offs were used for this discrimination.

HCC diagnosis was made before LT according to proposed international guidelines; a biopsy was performed only in case of uncertain diagnosis.[11,12]

Both centers handled similar listing criteria: MC-IN patients were immediately considered for wait-list registration; MC-OUT patients first underwent downstaging. After downstaging, the accepted upper limit of tumor burden for registration corresponded to the University of California San Francisco criteria.

In both populations, all patients had at least 1 LRT before LT or drop-out (DO). Multimodal LRT was done in 60 of 179 (33.5%) TS patients and in 18 of 110 (16.4%) VS patients. Details on LRTs are reported in Table 1.

Demographic data, radiological and pathological tumor characteristics, clinical course, and blood tests, including albumin, C-reactive protein (CRP), AFP, neutrophil-to-lymphocyte ratio (NLR), and platelet-to-lymphocyte ratio (PLR) were collected.

TABLE 1. Demographic and Clinical Characteristics (Training Set, n = 179 and Validation Set, n = 110)

	Training Set (n = 179)	Validation Set (n = 110)	
Variables	**Median (IQR) or Number (%)**		***P***
Demographics			
Recipient age at WT registration, yrs	59.4 (54.7–63.6)	56.0 (50.7–62.2)	0.006
Recipient gender (M/F)	38/179 (21.2/78.8)	12/98 (10.9/89.1)	0.03
WT, mo	5.1 (2.0–10.0)	2.7 (1.2–5.0)	<0.001
≤120 d	75 (41.9)	80 (72.7)	<0.001
HCV positivity	69 (38.5)	67 (60.9)	<0.001
HBV positivity	27 (15.1)	25 (22.7)	0.12
Lab MELD	10 (7–13)	12 (9–16)	0.005
Child-Turcotte-Pugh C status	33 (18.4)	22 (20.0)	0.73
AFP variables			
AFP at diagnosis, ng/mL	10.2 (4.5–43.6)	11.0 (4.9–52.8)	0.55
≥1000 ng/mL	2 (1.1)	5 (4.5)	0.11
AFP at LT or DO, ng/mL	8.6 (4.0–58.2)	11.5 (4.4–63.5)	0.38
≥1000 ng/mL	8 (4.5)	9 (8.2)	0.21
AFP-slope, ng/mL/mo	0.0 (−3.6 to 2.1)	0.0 (−0.6–0.4)	0.68
≥15 ng/mL/mo	29 (16.2)	16 (14.5)	0.74
Inflammatory markers variables			
C-reactive protein (mg/dL) at LT or DO	0.7 (0.3–1.9)	0.7 (0.3–2.1)	0.78
≥10.0 mg/dL	9 (5.0)	3 (2.7)	0.55
Albumin (g/dL) at LT or DO	3.3 (2.8–3.8)	3.5 (3.0–3.9)	0.03
<3.5 g/dL	103 (57.5)	54 (49.1)	0.18
NLR at LT or DO	3.4 (1.9–5.7)	2.9 (2.0–4.2)	0.31
≥5.0	53 (29.6)	21 (19.1)	0.05
PLR at LT or DO	93.3 (64.3–139.8)	69.6 (51.9–94.1)	<0.001
≥150.0	38 (21.2)	10 (9.1)	0.009
Radiological features of the tumor			
Diameter largest lesion (cm) at diagnosis	2.0 (2.5–3.9)	2.8 (2.0–3.8)	0.28
Number of lesions at diagnosis	1 (1–2)	1 (1–2)	0.16
MC-OUT status at diagnosis	43 (24.0)	40 (36.4)	0.03
Diameter largest lesion (cm) at LT or DO	1.0 (0.0–2.0)	1.6 (0.0–2.7)	0.02
Number of lesions at LT or DO	1 (0–1)	1 (0–2)	0.007
MC-OUT status at LT or DO	16 (8.9)	19 (17.3)	0.04
mRECIST			
CR	79 (44.1)	35 (31.8)	0.05
PR	36 (20.1)	34 (30.9)	0.05
SD	27 (15.1)	15 (13.6)	0.86
PD	37 (20.7)	26 (23.6)	0.56
LRT			
LRT number of sessions	3 (2–4)	2 (1–2)	<0.001
Multimodal LRT	60 (33.5)	18 (16.4)	0.002
TACE	148 (82.7)	98 (89.1)	0.17
RFA	23 (12.8)	13 (11.8)	0.86
PEI	61 (34.1)	6 (5.5)	<0.001
Hepatic resection	12 (6.7)	17 (15.5)	0.03
Drop out			
DO patients	34 (19.0)	15 (13.6)	0.26
HCC-related DO	17 (9.5)	10 (9.1)	0.99
Death during WT	12 (6.7)	7 (6.4)	0.99
Pathological features of the tumor*			
Diameter largest lesion (cm) at pathology	1.0 (0.0–2.0)	3.0 (2.0–4.0)	<0.001
Number of lesions at pathology	1 (0–2)	2 (1–2)	0.03
MC-OUT status at pathology	30 (20.7)	34 (35.8)	0.11
Necrosis			
100%	44 (30.3)	16 (16.8)	0.05
99–31%	74 (51.0)	48 (50.5)	0.70
30%	27 (18.6)	31 (32.6)	0.01
Poor grading	18 (12.4)	14 (14.7)	0.85
Multifocal tumor	53 (36.6)	53 (55.8)	0.10
Bi-lobar tumor	34 (23.4)	1 (1.1)	<0.001
Micro-vascular invasion	24 (16.6)	11 (11.6)	0.14
Macro-vascular invasion	1 (0.7)	1 (1.1)	0.99
Recurrence	12 (8.3)	12 (12.6)	0.52

F indicates female; HBV, hepatitis B virus; HCC, hepatocellular cancer; HCV, hepatitis C virus; MELD, model for end-stage liver disease; PEI, percutaneous ethanol injection; RFA, radio-frequency ablation.

*Medians (IQR) and % calculated on 145 and 95 LT patients in training and validation set, respectively.

Response to LRT was determined according to the modified Response Evaluation Criteria in Solid Tumors (mRECIST): complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD) were defined accordingly.[13]

The AFP cut-off value of 1000 ng/mL was chosen in accordance with previously published literature.[4] AFP slope was calculated according to the formula:

[(final-AFP)-(initial-AFP)]/time lapse between the two measurments

Initial AFP corresponded to the first available measurement at moment of diagnosis; the last one corresponded to the measurement closest to LT or DO. The AFP-slope cut-off value of 15.0 ng/mL/month was chosen as previously proposed.[5]

Last available values of inflammatory markers immediately before LT or DO were used: NLR and PLR were calculated by dividing absolute neutrophil and platelet counts by absolute lymphocyte count. Cut-offs of 5.0 and 150.0 were considered as previously described.[8] Similarly, CRP and albumin cut-offs were set at 10.0 mg/mL and 3.5 g/dL.[14]

Statistical Analysis

Continuous variables were reported as medians and IQR: Mann-Whitney *U* test was used for comparisons between groups. Dummy variables were reported as numbers and percentages: comparisons were done using Fisher exact text. In the TS, 11 pre-operatively available variables able to predict the risk of ITT-death were analyzed using a multivariate Cox proportional hazard model (stepwise backward conditional procedure). Hazard ratio (HR) and 95% confidence intervals (CIs) were reported for significant variables. A prognostic score was developed on the basis of variables with a P value >0.1 in the Cox regression analysis. The best cut-off to use for the TRAIN score was established according to the best obtained diagnostic odds ratio (DOR) for the risk of ITT-death. DOR was calculated according to the formula:

$$(\text{sensitivity}^{*}\text{specificity})/[(1-\text{sensitivity})^{*}(\text{1-specificity})]$$

A receiver operator curve (ROC) analysis was performed with the intent to evaluate the prognostic ability of different variables in the diagnosis of microvascular invasion (mvi). Area under the curve (AUC) and 95% CI were reported. ITT-survival and recurrence rate were carried out using Kaplan-Meier statistics and Log-rank test. In the ITT-survival analysis, both dropped-out and transplanted patients were censored according to their date of death. When analyzing tumor recurrence, transplanted patients only were considered. Variables with a P value <0.05 were considered statistically significant. SPSS statistical package version 23.0 was used (SPSS Inc., Chicago, IL).

RESULTS

Several differences were observed between TS and VS, with longer WT (5.1 vs 2.7 months; $P < 0.001$) and a higher median number of LRT (3 vs 2; $P < 0.001$) in the TS, and more MC-OUT patients at first radiological assessment in the VS (36.4 vs 24.0%; $P = 0.03$). Post-LRT radiological success rate was similar (mRECIST-CR+PR: 64.2% vs 62.7%). TS had more DOs (19.0% vs 13.6%; $P = 0.26$) and VS more post-LT recurrences (12.6% vs 8.3%; $P = 0.52$).

At pathological examination, TS had higher mvi (16.6% vs 11.6%; $P = 0.14$) and complete tumor necrosis rates (30.3% vs 16.8%; $P = 0.05$), while VS had more pathological MC-OUT patients (35.8% vs 20.7%; $P = 0.11$).

TS had a higher number of patients exceeding the cut-offs of AFP slope ≥15.0 ng/mL/month (16.2% vs 14.5%; $P = 0.74$), NLR ≥5.0 (29.6% vs 19.1%; $P = 0.05$), and PLR ≥150.0 (21.2% vs 9.1%; $P = 0.009$), while median final AFP values were higher in VS (Table 1).

Risk Factors for ITT-death

Eleven established pre-operative risk factors for ITT-death were evaluated in TS using Cox regression analysis: mRECIST-PD (HR = 2.7; $P < 0.001$) and AFP slope ≥15.0 ng/mL/month (HR = 2.3; $P = 0.003$) were the unique independent risk factors. NLR ≥5.0 at LT or DO (HR = 1.6; $P = 0.08$) and WT (HR = 1.0; $P = 0.08$) almost reached statistical significance (Table 2).

On the basis of these results, a score was developed according to the following equation:

$$0.988\,(\text{if mRECIST-PD}) + 0.838\,(\text{if AFP slope} \geq 15.0\,\text{ng/mL/month}) + 0.452\,(\text{if NLR} \geq 5.0) - 0.03 * \text{WT}(\times \text{month})$$

The proposed TRAIN score was evaluated at ROC analysis; a cut-off of 1.0, corresponding to the 85th percentile, was defined. This value presented a poor sensitivity (26.1%) but an excellent specificity (91.8%), and a DOR value of 4.0, being the best one among the tested cut-offs.

Prediction of Mvi, ITT-death, and Post-LT Recurrence

Comparing to other scores (MC-status, AFP slope ≥15.0 ng/mL, mRECIST-PD), TRAIN score best predicted in both sets mvi, a well-known risk factor for post-LT recurrence: AUC was 68.2 ($P = 0.005$) in TS and 76.6 ($P = 0.004$) in VS.

When investigating ITT-death, TRAIN score performed very well in TS (AUC = 71.9; $P < 0.001$), being markedly superior to MC-status (AUC = 49.3; $P = 0.88$). TRAIN score and all the other tested variables however failed to be statistically significant as diagnostic tools of ITT-death in VS. When post-LT recurrence was investigated, none of the tested variables was significant in both scenarios (Table 3).

TABLE 2. Multivariable Cox Regression Analysis for the Risk of Intention-to-treat Death in the Training Set Population (Backward Conditional Method)

Variables	HR	95% CI	*P*
mRECIST PD (Y/N)	2.7	1.6–4.5	<0.001
AFP slope ≥15.0 ng/mL/mo (Y/N)	2.3	1.3–4.0	0.003
NLR ≥5.0 at LT or DO (Y/N)	1.6	1.0–2.6	0.08
WT mo (per month)	1.0	0.9–1.0	0.08

-2Log likelihood: 625.3.

Variables initially analyzed in the model and then excluded: year of LT (per year), recipient age at WT registration (per year), recipient male gender (Y/N), HCV positivity (Y/N), LRT number of procedures (per number), PLR ≥150 at LT or DO (Y/N), MC-OUT status (Y/N).

TABLE 3. Prognostic Ability of Different Variables to Predict ITT-death (Entire Population: Training Set, n = 179 and Validation Set, n = 110), Microvascular Invasion and Post-LT Recurrence (Only LT Patients: Training Set, n = 145 and Validation Set, n = 95)

	AUROC	95% CI	*P*	AUROC	95% CI	*P*
Variables	**Training Set (n = 179)**			**Validation Set (n = 110)**		
			ITT-death			
TRAIN score	71.9	64.1–79.7	<0.001	53.9	42.1–65.7	0.51
mRECIST PD	63.8	55.2–72.5	0.002	52.6	41.0–64.1	0.66
MC-OUT status	49.3	40.6–58.0	0.88	49.1	37.6–60.5	0.87
AFP slope ≥15.0 ng/mL	60.4	51.6–69.2	0.02	53.3	41.7–64.9	0.57
		Training set (n = 145)			Validation set (n = 95)	
			Microvascular invasion			
TRAIN score	68.2	59.9–75.8	0.005	76.6	65.2–88.0	0.004
mRECIST PD	65.3	56.9–73.1	0.02	72.2	54.9–89.6	0.02
MC-OUT status	55.4	46.8–63.8	0.8	60.6	41.5–79.7	0.3
AFP slope ≥15.0 ng/mL	54.1	45.5–62.5	0.5	57.1	38.0–76.1	0.4
			Post-LT recurrence			
TRAIN score	58.2	37.2–79.1	0.35	58.8	36.9–80.6	0.32
mRECIST PD	65.2	46.9–83.5	0.08	64.8	47.0–82.6	0.10
MC-OUT status	50.1	33.0–67.2	0.99	62.3	45.2–79.4	0.16
AFP slope ≥15.0 ng/mL	58.4	40.0–76.8	0.34	65.2	46.7–83.7	0.09

ITT-survival and Recurrence Rate

ITT-survival rates were extremely poor in TRAIN score ≥1.0 in both sets. Twenty-seven TS and 13 VS patients exceeding this cut-off had 5-year survivals of 23.5% (median value = 12.0, 95% CI = 9.6–14.4) and 20.7% (median value = 53.7, 95% CI = 0.0–141.2).

In TS, TRAIN score allowed to discriminate patients in relation to 5-year recurrence rate (30.0% vs 8.9% in patients with a score ≥ or <1.0; log-rank = 0.1). All VS patients exceeding the cut-off recurred within 3 years (5-year recurrence rates: 100.0% vs 13.8% in patients with a score ≥ or <1.0, log-rank < 0.001) (Table 4, Fig. 1A).

In TS, 5-year ITT-survival rates were 67.5% and 61.1%, when meeting TRAIN score and MC-status; when exceeding them, survival rates were 23.5% and 62.6%, respectively. Similarly, 5-year recurrence rates were 8.9% and 10.4% in patients meeting TRAIN and MC-status; when exceeding the scores, recurrence rates were 30.0% and 10.4%, respectively.

In VS, 5-year ITT-survival rates were 66.7% and 61.5% in patients meeting TRAIN score and MC-status; when exceeding them, survival rates were 20.7% and 63.4%, respectively. Similarly, 5-year recurrence rates were 13.8% and 9.9% in patients meeting TRAIN score and MC-status; when exceeding the scores, recurrence rates were 100.0% and 34.1%, respectively.

These findings indicate that, when compared with the ''gold-standard MC'', TRAIN score allowed, in both sets, to obtain higher survivals without increasing recurrence risk.

As the number of patients meeting the TRAIN score encompassed the number of MC-IN patients, these results indicate that the number of transplantable patients can be substantially raised (+8.9% in TS and 24.6% in VS) without raising the recurrence risk (Fig. 2).

ITT-survival and Recurrence rate in the Subgroups of Initially MC-IN and MC-OUT Patients

When considering initially MC-IN patients only, TRAIN score allowed to discriminate both sets in terms of ITT-survival and recurrence rate. When considering TS, TRAIN score well discriminated initially MC-IN patients, with a more than 4-fold increased risk of 5-year recurrence in patients with a score ≥1.0 (35.7% vs 8.4%; log-rank = 0.04). In VS, half of the patients having a TRAIN score ≥1.0 recurred, no patient surviving more than 21 months after LT (Table 4, Fig. 1B).

When considering initially MC-OUT patients only, TRAIN score ≥1.0 well selected TS patients in terms of 5-year ITT-survival (69.9% vs 21.4%; $P = 0.009$); no conclusions could be made in terms of recurrence due to the small sample size. In VS, a good stratification was possible in relation to 5-year recurrence rate (26.7% vs 100%: $P = 0.008$). Although better ITT-survivals were observed in patients meeting the TRAIN score, no statistical significance was observed in the specific subanalysis of MC-OUT patients (70.0% vs 41.7%; $P = 0.28$) (Table 4, Fig. 1C).

DISCUSSION

In HCC patients, WT and post-LT course may be considered as a ''continuum,'' in which HCC-related DO and post-LT recurrence represent different consequences of a higher tumor aggressiveness. Despite refined allocation strategies, up to one-fifth of patients still experience an unfavorable tumor-related course following waiting-list registration. Differences observed across centers in terms of DO or recurrence may relate to different WT lengths. In ''long-WT centers,'' tumor selection more commonly takes place during the wait-list period (>DO), leading to a lower number of post-LT recurrences. Conversely, ''short-WT centers'' report less DO but more post-LT recurrences. This is demonstrated at its best in the living-donor LT scenario, in which there is theoretically no WT.[15] A UNOS survey, including 5002 HCC recipients, revealed higher post-LT 1-year recurrences in case of short-WT (≤120 days) (3.9% vs 2.2%, $P = 0.002$); long WT reduced recurrence risk by 40%.[9] Another US study, including 6160 HCC listed patients, showed a higher incidence of death on the waiting list (8.4% vs 1.6%, $P < 0.001$) but a better ITT recurrence-free survival ($P < 0.001$) in patients belonging to long WT regions.[10]

All these observations are in line with our findings: the long WT TS had more DO but less recurrences, the short WT VS revealed the contrary. In the TRAIN score, WT was a protective parameter; the longer the WT, the less the risk of ITT-death.

NLR, another TRAIN score component, has been reported to be a predictor of death and recurrence in HCC patients.[10,16] The importance of NLR may be linked to the tumor burden. A UK study

TABLE 4. ITT-Survival and Post-LT Recurrence Rate According to the Proposed TRAIN Score Cut-Off ≤ or >1.0 in the Entire Population and in the Subgroups of Initially MC-IN and MC-OUT Patients

Variables							
Entire Population	**1 yr**	**3 yrs**	**5 yrs**		**1 yr**	**3 yrs**	**5 yrs**
ITT-survival							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 152)	83.5	73.0	67.5	≤1.0 (n = 97)	83.3	73.9	66.7
>1.0 (n = 27)	45.2	23.5	23.5	>1.0 (n = 13)	62.2	62.2	20.7
		Log-rank $P < 0.001$				Log-rank $P = 0.01$	
Post-LT recurrence*							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 133)	3.5	7.7	8.9	≤1.0 (n = 88)	2.8	11.3	13.8
>1.0 (n = 12)	12.5	30.0	30.0	>1.0 (n = 7)	0.0	100.0	100.0
		Log-rank $P = 0.1$				Log-rank $P < 0.001$	
Initially MC-IN patients							
ITT-survival							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 116)	82.7	70.6	66.8	≤1.0 (n = 64)	84.3	76.2	66.0
>1.0 (n = 20)	52.9	31.8	23.8	>1.0 (n = 6)	50.0	50.0	0.0
		Log-rank $P = 0.001$				Log-rank $P = 0.011$	
Post-LT recurrence*							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 99)	1.1	6.7	8.4	≤1.0 (n = 60)	1.9	6.9	6.9
>1.0 (n = 10)	14.3	35.7	35.7	>1.0 (n = 4)	0.0	50.0†	—
		Log-rank $P = 0.041$				Log-rank $P = 0.012$	
Initially MC-OUT patients							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 36)	86.0	80.0	69.9	≤1.0 (n = 33)	81.5	70.0	70.0
>1.0 (n = 7)	21.4	21.4	21.4	>1.0 (n = 7)	83.3	83.3	41.7
		Log-rank $P = 0.009$				Log-rank $P = 0.28$	
Post-LT recurrence*							
		Training set				Validation set	
TRAIN score				TRAIN score			
≤1.0 (n = 34)	10.9	10.9	10.9	≤1.0 (n = 33)	4.5	19.4	26.7
>1.0 (n = 2)	0.0	0.0	0.0	>1.0 (n = 7)	50.0	100.0‡	—
		Log-rank $P = 0.70$				Log-rank $P = 0.008$	

*Post-LT recurrence analysis was performed only in transplanted patients.
†All the patients censored within 21 months after LT with a 50.0% of recurrence rate.
‡All the patients censored within 26 months after LT with a 100.0% of recurrence rate.

looking only at MC-IN patients infirmed its prognostic value.[17] However, it has been shown that the impact of NLR increases with the growing percentage of initially MC-OUT patients.[18] In the present study, the initial radiological MC-OUT cases were 24% to 36%, explaining the inclusion of NLR in the equation.

In our study, the strongest predictor of ITT-death was, as already shown in previous reports, mRECIST-PD.[6,7,19–21] The Mainz group reported that any post transarterial chemo-embolization (T-ACE) progression overruled MC-status as a predictor for recurrence risk, with an AUC = 86.[7] A recent multicenter European study combining radiological and AFP progression corroborated these observations in both MC-IN and MC-OUT scenarios.[6] Similarly, the San Francisco group showed that effective downstaging allowed to obtain similar survivals independently from the initial MC-status, while LRT failure was associated with a higher number of DOs.[20] This group was also able to identify a subgroup of patients with a low risk of DO (<2% at 2 yrs) based on their post-LRT radiological and biological response.[21]

AFP increase during the WT was another significant risk factor for ITT-death in our study, confirming recent findings looking at raising AFP values.[22] Multicenter studies from France and US showed that a pre-LT ''static'' AFP value >1000 ng/mL strongly correlated with recurrence risk in MC-IN patients.[4,23] Studies investigating ''dynamic'' AFP modifications also showed an excellent ability to predict ITT-death and recurrence.[5,6,24] Recently, scores integrating AFP have been developed with the intent to predict the risk of DO.[25,26]

The combination of these 4 variables, making up the TRAIN score, has been tested in the different scenarios of ''short-'' and ''long-WT.'' The TRAIN score revealed to best predict mvi, an established risk factor for post-LT recurrence, in both scenarios.[27,28] ITT-survival was poor in both TS and VS patients exceeding TRAIN score ≥1.0, thereby confirming the efficacy of the model in predicting death during the entire period starting from first waiting-list registration. Also of note is that in the ''long-WT'' scenario, TRAIN score was partially able to discriminate patients in terms of

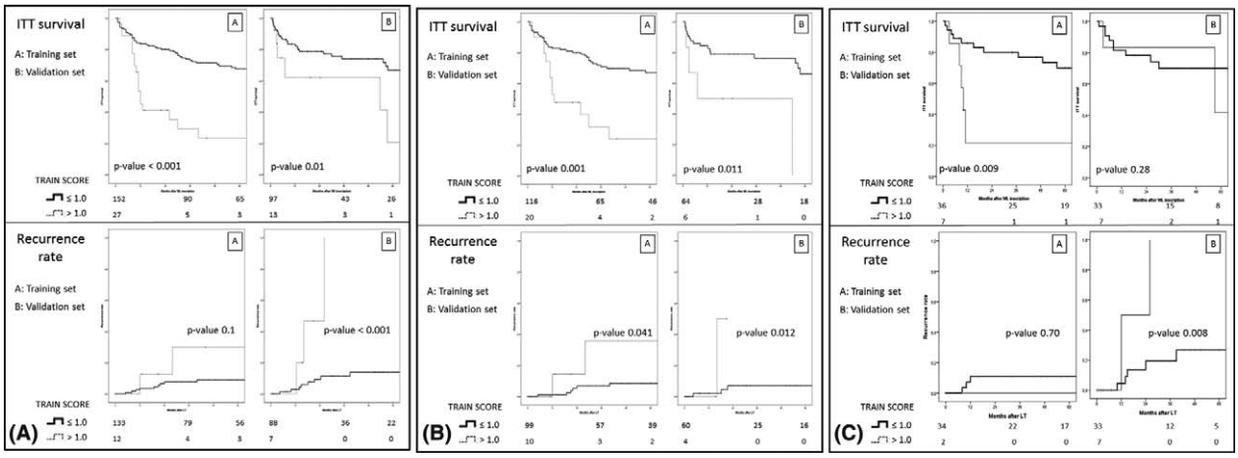

FIGURE 1. (A) Kaplan-Meier survival curves: ITT-survival and post-LT recurrence rate according to the proposed TRAIN score. (B) Kaplan-Meier survival curves: ITT-survival and post-LT recurrence rate according to the proposed TRAIN score in the subgroup of initially MC-IN patients. (C) Kaplan-Meier survival curves: ITT-survival and post-LT recurrence rate according to the proposed TRAIN score in the subgroup of initially MC-OUT patients.

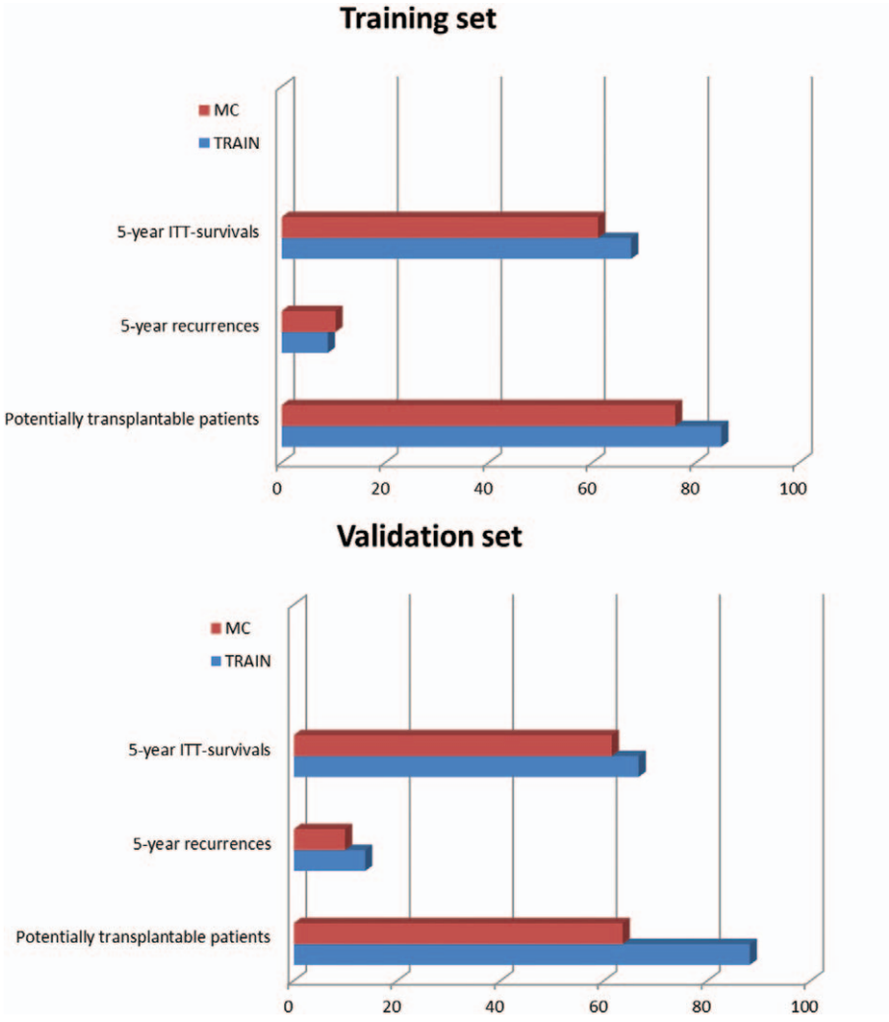

Figure 1 & Figure 2:
The art work needs to be re-done, because the text/figure legends in Figure 1 (A, B, C) are unreadable.

The resolution of Figure 1 needs to be improved and the size of Figure 1 should increase, whereas Figure 2 could be printed smaller.

Please make us a proposal.

FIGURE 2. Comparison between TRAIN and MC scores in terms of ITT-survivals, recurrence rates, and number of potentially transplantable patients.

recurrence rates. This could be explained by the fact that the patients were highly selected before LT by the ''time-factor,'' so that finally all transplanted patients were effectively at low risk for recurrence. However, when looking at MC-IN patients only, TRAIN score allowed to discriminate for recurrence risk also in the ''long-WT'' group. This can be due to the fact that a small but non-negligible group of MC-IN patients are at a high risk for post-LT recurrence. Indeed, although MC-IN patients are commonly considered as a homogeneous ''low-risk'' group, it has been recently shown that this is not fully true, showing that biologic markers are able to predict worse outcomes even in the favorable MC-IN setting.[6,21]

Several criticisms can be addressed to this study. First, TRAIN score variables are not available when entering the ITT analysis because at least 1 LRT is required before doing so. It should be reminded that such practice corresponds to the typical ''todays'' management of, even MC-IN, HCC patients waiting for LT (''bridging'' approach). On the basis of our results, we propose to consider patients with TRAIN score $\geq$1.0 as at ''high-risk'' for DO and recurrence, therefore managing them similarly to MC-OUT patients by applying more frequent pre-LT LRT and respecting a ''mandatory'' pre-LT minimal WT.[29]

Second, TRAIN score did not entirely give satisfactory AUCs in relation to diagnosis of mvi and incidence of post-LT recurrence. This can be explained by the fact that the score does not fully capture the complex reality of HCC patients waiting for LT: however, this fact is even more true for other scores such as MC. It is likely that the few number of events and small sample sizes reported in some subanalyses explain some ''unsatisfactory'' results.

Third, the presence of a ''time-dependent'' variable in the TRAIN score, namely the length of WT, might represent a problem in the specific subcohort of HCC patients presenting advanced liver disease (Child-Turcotte-Pugh-C, MELD >15). In fact, the first objective of TRAIN score, namely selecting ''high-risk'' HCC population in order to avoid organ ''wasting'' (concept of ''utility''), is overruled by the concept of ''transplant benefit'' (''the sickest first'') in case of advanced liver disease. So, the advantage of longer WT, aiming at achieving lower recurrence rates, is counteracted by the higher risk of tumor-unrelated DO. Consequently, in such (fortunately uncommon) condition, a balancing between these 2 concepts should be considered for minimizing the risk to remove too many patients from LT opportunity.

Fourth, this study includes relatively small patient cohorts belonging to 2 geographical distant centers adhering also to some, albeit small, clinical management differences. Apart from a very limited number of (heterogeneous) multicenter studies,[3,4,6] almost all studies reported in this field include smaller numbers. Moreover, a major benefit of our study was the confirmation of a mathematical model in a VS. We feel that the differences between the studied groups rather represent a benefit, allowing to test and validate a score in 2 very different scenarios.

Finally, it is important to emphasize that the arbitrary decision to select the TRAIN score cut-off at 1.0 resulted from the intention to obtain a high specificity of the test (avoiding thereby false positives). Despite the fact that such decision hampers the possibility of completely eliminating the risk of post-LT HCC recurrence, we feel that the main objective of a good selection tool in the ''oncological transplant scenario'' should be to look simultaneously at minimizing both risks of recurrence and exclusion from potentially curative treatments. For this reason, a ''high'' cut-off value representing the 85th percentile of the TS population was chosen. Such strategy has been already been adopted by other authors focusing on the detection of cut-off values for selective scores of HCC patients waiting for LT.[4,23]

CONCLUSIONS

The proposed TRAIN score combining radiological progression, increase of AFP and inflammatory marker NLR, and short WT is a promising tool allowing to predict, on an ITT base, the risks of death and post-LT recurrence both in long- and short WT scenarios. The availability of all components of this score in all HCC transplant candidates will allow its implementation and validation in daily clinical practice.

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. 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;10:35–43.
4. 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.
5. 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;10:129–137.
6. Lai Q, Avolio AW, Graziadei I, et al., European Hepatocellular Cancer Liver Transplant Study Group. Alpha-fetoprotein and modified response evaluation criteria in solid tumors progression after locoregional therapy as predictors of hepatocellular cancer recurrence and death after transplantation. *Liver Transpl*. 2013;19:1108–1118.
7. Otto G, Schuchmann M, Hoppe-Lotichius M, et al. How to decide about liver transplantation in patients with hepatocellular carcinoma: size and number of lesions or response to TACE? *J Hepatol*. 2013;59:279–284.
8. Lai Q, Castro Santa E, Rico Juri JM, et al. Neutrophil and platelet-to-lymphocyte ratio as new predictors of dropout and recurrence after liver transplantation for hepatocellular cancer. *Transpl Int*. 2014;27:32–41.
9. Samoylova ML, Dodge JL, Yao FY, et al. Time to transplantation as a predictor of hepatocellular carcinoma recurrence after liver transplantation. *Liver Transpl*. 2014;20:937–944.
10. 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.
11. Bruix J, Sherman M, American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. *Hepatology*. 2011;53:1020–1022.
12. European Association for the Study of the Liver, European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2012;56:908–943.
13. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis*. 2010;30:52–60.
14. Pinato DJ, Stebbing J, Ishizuka M, et al. A novel and validated prognostic index in hepatocellular carcinoma: the inflammation based index (IBI). *J Hepatol*. 2012;57:1013–1020.
15. Lai Q, Avolio AW, Lerut J, et al. Recurrence of hepatocellular cancer after liver transplantation: the role of primary resection and salvage transplantation in East and West. *J Hepatol*. 2012;57:974–979.
16. Harimoto N, Shirabe K, Nakagawara H, et al. Prognostic factors affecting survival at recurrence of hepatocellular carcinoma after living-donor liver transplantation: with special reference to neutrophil/lymphocyte ratio. *Transplantation*. 2013;96:1008–1012.
17. Parisi I, Tsochatzis E, Wijewantha H, et al. Inflammation-based scores do not predict post-transplant recurrence of hepatocellular carcinoma in patients within Milan criteria. *Liver Transpl*. 2014;20:1327–1335.
18. Lai Q, Lerut J. Inflammation-based scores and hepatocellular carcinoma: ''Faithful friends are hard to find''. *Liver Transpl*. 2014;20:1527–1528.
19. Lai Q, Avolio AW, Graziadei I, et al., European Hepatocellular Cancer Liver Transplant (EurHeCaLT) Study Group. Response to locoregional treatment and alpha-fetoprotein trend in liver transplant candidates for HCC: dwarfs standing on the shoulders of giants. *J Hepatol*. 2014;60:1331–1332.

20. Yao FY, Mehta N, Flemming J, et al. Downstaging of hepatocellular cancer before liver transplant: long-term outcome compared to tumors within Milan criteria. *Hepatology*. 2015;61:1968–1977.
21. Mehta N, Dodge JL, Goel A, et al. Identification of liver transplant candidates with hepatocellular carcinoma and a very low dropout risk: implications for the current organ allocation policy. *Liver Transpl*. 2013;19:1343–1353.
22. Lai Q, Melandro F, Pinheiro RS, et al. Alpha-fetoprotein and novel tumor biomarkers as predictors of hepatocellular carcinoma recurrence after surgery: a brilliant star raises again. *Int J Hepatol*. 2012;2012:893103.
23. 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.
24. Lai Q, Inostroza M, Rico Juri JM, et al. Delta-slope of alpha-fetoprotein improves the ability to select liver transplant patients with hepatocellular cancer. *HPB (Oxford)*. 2015;17:1085–1095.
25. Toso C, Dupuis-Lozeron E, Majno P, et al. A model for dropout assessment of candidates with or without hepatocellular carcinoma on a common liver transplant waiting list. *Hepatology*. 2012;56:149–156.
26. Vitale A, Volk ML, De Feo TM, et al., Liver Transplantation North Italy Transplant program (NITp) working group. A method for establishing allocation equity among patients with and without hepatocellular carcinoma on a common liver transplant waiting list. *J Hepatol*. 2014;60:290–297.
27. Cucchetti A, Piscaglia F, Grigioni AD, et al. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study. *J Hepatol*. 2010;52:880–888.
28. Ciccarelli O, Lai Q, Goffette P, et al. Liver transplantation for hepatocellular cancer: UCL experience in 137 adult cirrhotic patients. Alpha-foetoprotein level and locoregional treatment as refined selection criteria. *Transpl Int*. 2012;25:867–875.
29. Terzi E, Ray Kim W, Sanchez W, et al. Impact of multiple transarterial chemoembolization treatments on hepatocellular carcinoma for patients awaiting liver transplantation. *Liver Transpl*. 2015;21:248–257.

DISCUSSANTS

~~J. Dong (China)~~:

It is an honor for me to review and comment on this work. Dr. Lai, Prof Lerut, and colleagues from Belgium and Italy described a prognostic score in patients with HCC waiting for liver transplantation, namely the ''TRAIN'' score, which combined the predictive power of radiological response (mRECIST), AFP level/slope, index of inflammatory markers (NLR), and waiting time. It deals with the clinically relevant issue of patient selection for liver transplantation and tries to identify a more appropriate selection than the current orientation to the Milan criteria. With a training set of 179 patients and a validation set of 110, the authors proved that the score can predict the risk of ITT-death during the waiting time (WT) in case of long WT and the post-LT recurrence in case of short WT.

The ''TRAIN'' score has a potential to serve as a new criterion of selection and prognosis of transplant candidates. The Milan criteria and current expanded practice such as the UCSF and ''Up-to-seven'' criteria are mainly based on the solid tumor burden (size and number). Tumor biological behaviors remain an ''overlooked'' major issue. The TRAIN score could fill this important gap by including the mRECIST, AFP, and NLR.

I have some questions. As some colleagues may notice that the patient numbers in both training and validation sets are limited, and the 2 cohorts were selected from 2 centers, 1 in Belgium and the other in Italy. Patient heterogeneity in the 2 centers may undermine this score and rather represent a benefit, allowing to test and validate a score in different scenarios. The authors may want to add some explanation to this.

And, the percentages of hepatic resection are different in the training and validation sets (6.7% vs 15.5%), is that statistically significant? Will this have some effect on the eventual results?

Again, congratulations to Dr Lai, Prof Lerut, and colleagues for this work.

Response From J. Lerut (Brussels, Belgium):

We agree with you with your comments. We feel however that using 2 different populations absolutely represents an advantage. In fact, the maximum mathematical benefit in testing a training set with a validation set is observed when the 2 populations are very different.

In relation to resection, no specific analysis has been done due to the smallness of the samples. In the validation group, there were statistically more partial resections than in the training group. However, we did not give too much attention to this difference because as you have seen, the most powerful variable in the TRAIN score is the mRECIST criteria. In the mRECIST criteria, if you have a recurrence after resection, it is considered an aggressive tumor. So in fact, resection is included in the most powerful variable of the TRAIN score.

~~P.-A. Clavien (Switzerland)~~:

Congratulations for this innovative and relevant study. I have 3 short questions. First, it seems—if I am correct—that you are mixing in your study, patients who underwent liver transplant with others who did not. So, I am confused about which endpoints we are looking at? My second question is about the comparisons of your TRAIN score with other scores, and not only the Milan criteria. The TRAIN score is based on combination of previous scores. Therefore, we would like to know whether your new score improves the predictive value of the individual previously described scores that are aggregated in the TRAIN score. For example, is your score better than the popular AFP level/slope. As a follow-up of this question, how should we apply the TRAIN score? Should it replace the Milan criteria? What will you do in your own center with this new predictive tool?

Response From J. Lerut (Brussels, Belgium):

This is an intention-to-treat analysis. Our end-points were to select a score able to predict drop-out and recurrence. When a patient is initially considered for liver transplantation, he is treated with a loco-regional approach; then, we start to consider what happens.

In relation to other scores, you can imagine we looked for this because we are very interested by what has been done in the Paul Brousse Centre in relation to the AFP level and slope. Compared with all different scores, the TRAIN score came always out as the best one.

Third, the question on how to apply the TRAIN score; I think the message of the paper is not to use it with the intent to exclude patients from liver transplantation, but to select patients who should be considered at high risk for recurrence, therefore potentially requiring more aggressive loco-regional treatment before transplantation. Unfortunately, mostly hepatologists, gastroenterologists, and oncologists do, after the loco-regional treatment, a magnetic nuclear resonance or CT scan. If the tumor does not enhance, it is many times stated that the tumor has been ''zeroed.'' However, we know from the pathology reports of the total hepatectomy specimen that almost always 5% to 10% of vital tissue remains observed in the local area of treatment. This means that these patients should have a more aggressive treatment before the transplantation, especially when fitting within the TRAIN score. So, the TRAIN score can be useful in stratifying the patients.

~~C. Bruns (Germany)~~:

The developed score was designed for patients outside MILAN. So, could you define more detailed whether they were outside MILAN but inside up-to-7?

Response From J. Lerut (Brussels, Belgium):

We are no longer interested in Milan criteria. Milan criteria were proposed in 1996; today, we are in 2016. We should not

consider anymore morphology only as a possible inclusion limit for liver transplantation. The best proof of this comes from the Southeast Asia countries such as Korea, Japan, and Taiwan, where thousands of living donor liver transplantations were performed in patients with advanced HCC. Milan criteria are only morphology-based ones; we must from now onwards combine tumor morphology and biology. Of course if you have a Milan-OUT patient with 5000 ng/mL of AFP and the loco-regional treatment does not downstage the tumor, such patient should not be transplanted.

R. Adam (France):

In France, we do not use anymore the Milan criteria as you know, and use a model integrating the number, the size, and the AFP. This has been published in Gastroenterology recently. Have you been able to correlate TRAIN-risk score to that of the French multicenter study in a way to see if it brings something additional?

Response From J. Lerut (Brussels, Belgium):

What you say confirms what we presented. We should come off the morphological criteria, but unfortunately all the organ allocation organisms still work only with the Milan criteria. We did not test our TRAIN score in comparison with the French score, but we are confident its validity is at least similar because the 2 ''French'' parameters are included in the TRAIN score. Compared with Milan Criteria, and simulations in both the scenarios of MC-OUT and MC-IN status, the TRAIN score proved to have a better predictive power.