

Tumor burden with AFP improves survival prediction for TACE-treated patients with HCC: An international observational study[☆]

Dongdong Xia^{1,2,3,†}, Wei Bai^{2,3,†}, Qiuhe Wang^{2,4,†}, Jin Wook Chung^{5,†}, Xavier Adhoute^{6,†}, Roman Kloeckner^{7,8,†}, Hui Zhang^{9,†}, Yong Zeng¹⁰, Pimsiri Sripongpun¹¹, Chunhui Nie¹², Seung Up Kim¹³, Ming Huang¹⁴, Wenhao Hu¹⁵, Xiangchun Ding¹⁶, Guowen Yin¹⁷, Hailiang Li¹⁸, Hui Zhao¹⁹, Jean-Pierre Bronowicki²⁰, Jing Li²¹, Jiaping Li²², Xiaoli Zhu²³, Jianbing Wu²⁴, Chunqing Zhang²⁵, Weidong Gong²⁶, Zixiang Li²⁷, Zhengyu Lin²⁸, Tao Xu²⁹, Tao Yin³⁰, Rodolphe Anty³¹, Jinlong Song³², Haibin Shi³³, Guoliang Shao³⁴, Weixin Ren³⁵, Yongjin Zhang³⁶, Shufa Yang³⁷, Yanbo Zheng³⁸, Jian Xu³⁹, Wenhui Wang⁴⁰, Xu Zhu⁴¹, Ying Fu⁹, Chang Liu¹⁰, Apichat Kaewdech¹¹, Rong Ding¹⁴, Jie Zheng¹⁵, Shuaiwei Liu¹⁶, Hui Yu¹⁷, Lin Zheng¹⁸, Nan You²¹, Wenzhe Fan²², Shuai Zhang²³, Long Feng²⁴, Guangchuan Wang²⁵, Peng Zhang^{26,42}, Xueda Li²⁷, Jian Chen²⁸, Feng Zhang³⁰, Wenbo Shao³², Weizhong Zhou³³, Hui Zeng³⁴, Gengfei Cao³⁵, Wukui Huang³⁷, Wenjin Jiang³⁸, Wen Zhang³⁹, Lei Li⁴⁰, Aiwei Feng⁴¹, Enxin Wang^{2,43}, Zhexuan Wang², Dandan Han^{2,44}, Yong Lv^{1,2}, Jun Sun², Bincheng Ren⁴⁵, Linying Xia⁴⁶, Xiaomei Li^{2,3}, Jie Yuan^{2,3}, Zhengyu Wang^{2,3}, Bohan Luo^{2,3}, Kai Li^{1,2}, Wengang Guo^{2,3}, Zhanxin Yin^{2,3}, Yan Zhao³, Jielai Xia⁴⁷, Daiming Fan¹, Kaichun Wu¹, Dominik Bettinger⁴⁸, Arndt Vogel^{49,50}, Guohong Han^{1,2,3,*}, on behalf of China HCC-TACE study group

JHEP Reports 2025. vol. 7 | 1–13



Background & Aims: Current prognostic models for patients with hepatocellular carcinoma (HCC) undergoing transarterial chemoembolization (TACE) are not extensively validated and widely accepted. We aimed to develop and validate a continuous model incorporating tumor burden and biology for individual survival prediction and risk stratification.

Methods: Overall, 4,377 treatment-naïve candidates for whom TACE was recommended, from 39 centers in five countries, were enrolled and divided into training, internal validation, and two external validation datasets. The novel model was developed using a Cox multivariable regression analysis and compared with our original 6-and-12 model (the largest tumor size [ts, centimetres] + tumor number [tn]) and other available models in terms of predictive accuracy.

Results: The proposed model, named the ‘6-and-12 model 2.0’, was generated as ‘ts + tn + 1.5×log₁₀ alpha-fetoprotein (AFP)’, showed good discrimination (C-index 0.674) and calibration (Hosmer–Lemeshow test $p = 0.147$), and outperformed current existing models. An easy-to-use stratification was proposed according to the different AFP levels (≤ 100 , 100–400, 400–2,000, 2,000–10,000, 10,000–40,000, and $>40,000$ ng/ml) along with the corresponding tumor burden cutoffs (8/14, 7/13, 6/12, 5/11, 4/10, and any tumor burden); that is, if the AFP level was 400–2,000 ng/ml, the stratification should be low-(≤ 6)/intermediate-(6–12)/high-risk (>12) strata. Hence, it could divide the patients into three distinct risk categories with a median overall survival of 45.0 (95% CI, 40.1–49.9), 30.0 (95% CI, 26.1–33.9), and 15.4 (95% CI, 13.4–17.4) months ($p < 0.001$) from low-risk to high-risk strata, respectively. These findings were confirmed in validation and subgroup analyses.

Conclusions: The 6-and-12 model 2.0 significantly improved individual outcome predictions and better stratified the candidates recommended for TACE; thus, this model could be used in both clinical practice and trial design.

© 2024 The Authors. Published by Elsevier B.V. on behalf of European Association for the Study of the Liver (EASL). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Transarterial chemoembolization (TACE) is a recommended treatment for patients with intermediate-stage hepatocellular carcinoma (HCC)^{1–4} and a preferable choice of stage migration for patients in the early stage of the disease for whom curable options are unsuitable.^{3,4} However, TACE results in a heterogenous objective response rate (ORR) of 40–80% and a

median overall survival (OS) of 13–48 months,^{5–8} indicating that not all patients benefit equally from TACE. Thus, given the significant advances in molecular and immune treatments, it is essential to accurately predict individual outcomes and identify patients who are likely to have poor outcomes and for whom early intensive treatments would be most beneficial in terms of OS, particularly considering the improved

[☆] Given their role as Editor, Arndt Vogel had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to the Guest Editor Tim Meyer.

* Corresponding author. Address: Department of Liver Diseases and Digestive Interventional Radiology, Xi'an International Medical Center Hospital, Northwest University, No. 777 Xitai Road, High-tech Zone, Xi'an, 710100, China. Tel: +86-29-68302656; Fax: +86-29-68302656.

E-mail address: hangh@fmmu.edu.cn (G. Han).

[†] These authors contributed equally to this study.

<https://doi.org/10.1016/j.jhepr.2024.101216>



responses and survival reported in the advanced-stage setting.^{4,9}

As mentioned above, the notorious heterogeneity of intermediate-stage HCC might account for the wide range of survival outcomes. Tumor burden, liver function, tumor biology, performance status, and treatment response can affect outcomes in individual patients who undergo TACE, all of which are differently weighted when determining OS.¹⁰ To address this situation, several prognostic models for TACE have been developed for early prognostication.^{11–19} However, most of these models are dichotomous, resulting in loss of prognostic information, or are not well validated in generalizable datasets. Furthermore, although simple and convenient, modeling risk in binary terms overestimates intersubgroup variance and diminishes intrasubgroup variance, which could then compromise its performance as a prognostic tool.²⁰ In this context, we recently proposed an evidence-based 6-and-12 model for guideline-recommend TACE candidates, based on tumor burden ([the largest tumor size (ts, cm) plus tumor number (tn)], which could predict individual outcomes and stratify the population into three strata with cutoffs of 6 and 12, presenting significantly different OS.¹⁹

Nevertheless, the performance of the original 6-and-12 model could be improved further. Given the biological behavior of tumors, alpha-fetoprotein (AFP) could be used as a prognostic factor, providing a potential solution to the issue of model underperformance. However, the clinical utility of AFP remains a subject of heated debate.¹⁰ First, to the best of our knowledge, AFP has widely been adopted arbitrarily with unrecognized cutoffs in the prognostic models mentioned above, possibly leading to loss of information. Second, the linearity or nonlinearity between AFP, as a continuous variable, and post-TACE survival outcomes needs to be further investigated. Furthermore, the interactions between AFP and other predictors remain unclear. Finally, although numerous prognostic models incorporating both morphological and biological factors have been developed, these are either limited by debatable reproducibility or include postoperative factors that hinder their clinical applicability.

Therefore, in this study, we developed and validated a novel, easily accessible prognostic model to accurately predict outcomes in candidates for whom TACE is recommended, which could improve the current prognosis and refine the stratification for intermediate-stage HCC in clinical practice and trial design.

Materials and methods

Study design

This study was conducted as per the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines.²¹ A two-step method was used to develop and validate a novel prognostic model for candidates for whom TACE is recommended. First, dataset of the original 6-and-12 model study was used to identify predictive factors associated with OS and to develop a new prognostic model.¹⁹ Second, three datasets (Chinese, European, and Asian) were used to internally and externally validate the performance of the models in terms of both discrimination and calibration. The study protocol followed the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Ethics Review Committee of Xi'an International Medical Center Hospital.

Written informed consent was obtained from all patients before treatment initiation.

Study populations

As mentioned in our previous study,¹⁹ the target population included 'recommended' or 'ideal' candidates for TACE, defined as those with unresectable Barcelona Clinic Liver Cancer (BCLC) stage A (BCLC-A) identified using comprehensive assessment and BCLC-B. The inclusion criteria were as follows: (1) treatment-naïve patients with unresectable HCC receiving TACE; (2) Child–Pugh score of A5–B7; and (3) at least one measurable lesion sized >1 cm. Patients were excluded according to the following criteria: (1) vascular invasion or extrahepatic spread; (2) spontaneous tumor rupture; (3) comorbidity with other malignancies; (4) decompensated liver cirrhosis (gastrointestinal bleeding, ascites, jaundice, or encephalopathy); (5) Eastern Cooperative Oncology Group performance status score >0; (6) treatment with any systemic or locoregional therapy; and (7) absence of baseline imaging information. According to treatment-stage migration, patients with potentially resectable or ablative lesions but at high risk for surgery, transplantation, and ablation therapy, for reasons such as old age, tumor location, technical feasibility, organ shortage, or comorbidities, were also enrolled.^{3,4} The detailed information of these datasets is provided in the [supplementary materials](#) and [Table S1](#)

Medical care

All participating centers had specific expertise in the management of HCC and the use of TACE. TACE procedures were conducted selectively or superselectively. The types and doses of embolization material [including lipiodol and drug-eluting beads (DEBs)] and chemotherapeutic agents (including doxorubicin, cisplatin, epirubicin, oxaliplatin, or a combination regimen) were selected and injected according to the practice of each institution. The embolization was monitored until the tumor arterial flow was reduced or achieved stasis, as observed on angiography or cone beam computed tomography (CT). Additional embolization material, including gelatine sponges or polyvinyl alcohol foam particles, was introduced after the drug/lipiodol emulsion injection. Repeat TACE was conducted following an 'on demand' schedule at an interval of 6–12 weeks if there was viable tumor or intrahepatic recurrence at follow-up imaging (contrast enhanced CT/magnetic resonance imaging), depending on the tolerance of the first treatment and patient condition, including liver function and general health.

Statistical analyses

Multiple imputation by chained equation was used to impute missing data, as described in the [supplementary materials](#) and [Table S2–4](#). The measurements are presented as the median (IQR or mean ± SD) or number (percentage, %) unless otherwise noted. OS was the primary endpoint and was defined as the time interval between the date of the first TACE and the date of all-cause death. The patients who underwent different types of TACE (conventional TACE and DEB-TACE) were integrated for analysis, because there was no significant difference in OS between these treatment types.^{22,23} Patients who survived until the last follow-up date or who were lost to follow-

up were censored. Student's t-test or the Mann–Whitney U test was used to compare continuous variables, and the chi-squared test or Fisher's exact test was used to compare categorical variables. Survival curves were estimated using the Kaplan–Meier method and compared using the log-rank test. AFP was normalized using a natural logarithm transformation to reduce the effect of variations, and high values were truncated at the upper limit of detection.

Tumor and laboratory test parameters were regarded as continuous variables. Intention-to-treat analysis was performed with the use of Cox univariable and multivariable regression analyses to identify the prognostic factors for OS in the training cohort. Variables with $p < 0.10$ in univariable analyses were selected for multivariable analysis. The nonlinearity of the continuous predictors and the interactions between the predictors were tested. A contour plot was used to depict the survival estimates based on these variables. A nomogram was generated by fitting a Cox regression model, and the performance of the novel model was assessed and compared with that of the currently available models, namely hepatoma arterial-embolization prognostic (HAP) score,¹³ mHAP score,¹⁵ mHAP-II score,¹⁷ mHAP-III score,¹⁶ BCLC subclassification,¹² and albumin–bilirubin (ALBI) score, as well as tumor burden criteria, including the 6-and-12 model,¹⁹ up-to-seven criteria,¹¹ four and seven criteria,¹⁴ and seven and 11 criteria.²⁴ Discrimination was measured using the concordance index (C-index) and area under the receiver operating characteristics curve (AUROC) with a 10-fold–100-times cross-validation approach at the timepoint of 3 and 5 years, as well as the time-dependent C-index. Calibration was tested by plotting the predicted and observed mortalities, the Hosmer–Lemeshow goodness-of-fit test, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The overall improvement in predictive accuracy was assessed by calculating the continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI).

The optimal cutoff value of the new prognostic model was identified using X-tile software (version 3.6.1; Yale University, New Haven, CT, USA) by selecting the largest χ^2 value to separate patients into groups with a low-risk, an intermediate-risk, and a high risk of death. To assess whether there was heterogeneity in the predictive value of the final models, we assessed the performance of prognostic models separately in internal and external validation cohorts as well as in different subgroups.

Differences were considered statistically significant when the corresponding p value was < 0.05 . All statistical analyses were conducted using SPSS software version 25 (SPSS, Chicago, IL, USA) and R software, version 4.05 (www.r-project.org), with the aid of the Hmisc, rms, riskRegression, pec, prodlm, SurvIDINRI, and survival packages.

Results

Fig. 1 shows the flowchart of patient inclusion in this study. The baseline characteristics are summarized in Table 1. The main etiology was HBV infection in the training, internal validation, and Asian validation cohorts (85.2%, 87.7%, and 63%, respectively). HCV infection (31.4%), alcohol consumption (25.9%), and HBV infection (30.1%) were documented reasons for HCC in the European validation cohort. There were 170 (21.2%), 50 (4.4%), and 83 (9.8%) patients who underwent

DEB-TACE in the internal validation (Table S5 shows the baseline characteristics of the Chinese DEB-TACE cohort), European, and Asian validation cohorts, respectively. Fig. S1 shows the pattern and percentage of missing values.

Overall survival

The median (IQR) follow-up time was 22 (11.9–34.1), 31.5 (18.4–44.5), 22.8 (12.3–37.4), and 39.0 (19.3–92.1) months for the training, internal validation, European validation, and Asian validation cohorts, with 811 (50.6%), 522 (65%), 874 (77.3%), and 463 (55.1%) events, respectively. The median OS was 32.9 (95% CI, 30.4–35.4), 35.1 (95% CI, 32.9–37.3), 24.9 (95% CI, 22.0–27.9), and 57.9 (95% CI, 48.7–67.1) months in the abovementioned cohorts, respectively (Fig. S2).

Development of novel prognostic model

Univariate Cox regression analysis indicated that baseline t_s , t_n , \log_{10} AFP, aspartate aminotransferase (AST), and ALBI scores were significantly associated with OS (Table 2). Discriminant analysis by multivariate Cox regression indicated that the following three independent risk factors could be used as final prognostic scores: t_s [hazard ratio (HR), 1.107; 95% CI, 1.088–1.126; $p < 0.001$], t_n (HR, 1.101; 95% CI, 1.066–1.138; $p < 0.001$), and \log_{10} AFP (HR, 1.162; 95% CI, 1.102–1.225; $p < 0.001$). Restrict cubic spline functions suggested that t_s , t_n , and \log_{10} AFP presented a linear relationship with the HR (nonlinear p values were 0.11, 0.05, and 0.40, respectively; Fig. S3). Interaction tests between variables suggested no interaction terms (interaction $p = 0.886$ for t_s and \log_{10} AFP and 0.089 for t_n and \log_{10} AFP). The coefficients of the variables derived from the Cox regression analyses in each model were multiplied by 10 and rounded to one decimal place for clinical use, which were used to generate an easily computed continuous risk equation:

$$\text{Linear predictor} = t_n + t_s + 1.5 * \log_{10} \text{ AFP}$$

Based on these findings, a nomogram for individual prognostic algorithms, named the 6-and-12 model 2.0, was developed (Fig. 2A) and is available at https://sixandtwelve-version2.shinyapps.io/Six-and-twelve_2/. The estimated 1-, 2-, and 3-year survival probabilities and median survival time of individual patients could be predicted before the TACE procedure using the sum of tumor size and number and \log_{10} AFP. The relationship between tumor burden (referred to as the sum of t_n and t_s), \log_{10} AFP, and 3-year survival probability is depicted in a contour plot (Fig. 2B).

Discrimination and calibration of the 6-and-12 model 2.0 in the training and validation cohorts

Referring to 10-fold–100-times cross-validation, the mean 3-year AUROC, 5-year AUROC, 3-year C-index, and 5-year C-index in the training cohort were 0.679 (SD, 0.079), 0.680 (SD, 0.127), 0.636 (SD, 0.049), and 0.626 (SD, 0.044), respectively (Fig. S4A). The corresponding results were 0.679 (SD, 0.062), 0.681 (SD, 0.090), 0.643 (SD, 0.044), and 0.636 (SD, 0.039) in the internal validation cohort; 0.662 (SD, 0.055), 0.641 (SD, 0.069), 0.630 (SD, 0.031), and 0.623 (SD, 0.029) in the European validation cohort; and 0.736 (SD, 0.061), 0.740 (SD, 0.060),

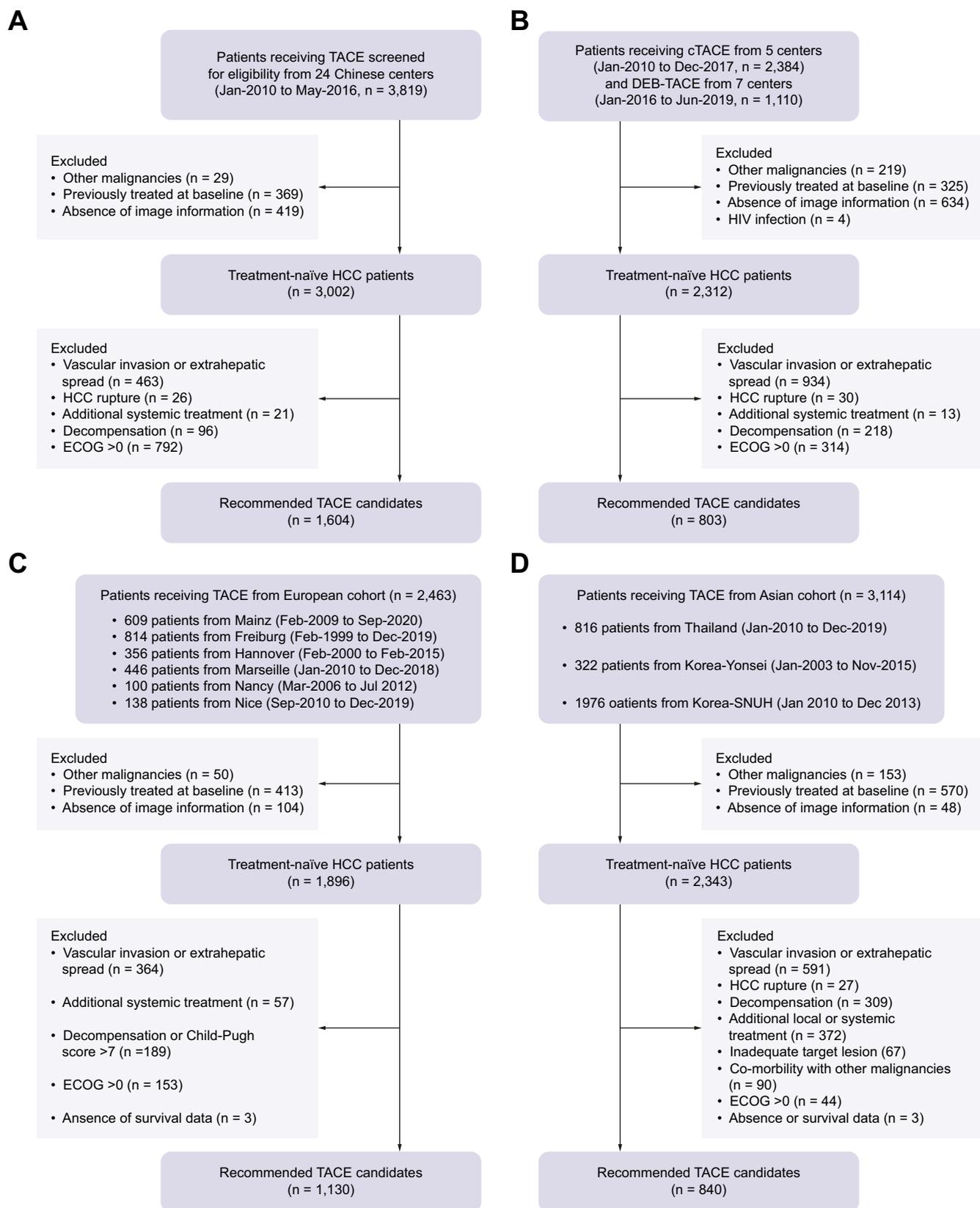


Fig. 1. Flow chart of the current study. (A) Training cohort; (B) Internal validation cohort; (C) European validation cohort; and (D) Asian validation cohort. cTACE, conventional transarterial chemoembolization; DEB, drug-eluting beads; HCC, hepatocellular carcinoma; HIV, human immunodeficiency virus; ECOG, Eastern Cooperative Oncology Group.

Table 1. Baseline demographics and clinical characteristics.

Variables	Training cohort (n = 1,604)	Internal validation cohort (n = 803)	European validation cohort (n = 1,130)	Asian validation cohort (n = 840)
Sex				
Male	1,390 (86.7%)	686 (85.3%)	944 (83.5%)	637 (75.8%)
Female	214 (13.3%)	117 (14.7%)	186 (16.5%)	203 (24.2%)
Age, years	57 (48–65)	59 (50–67)	68 (61–74)	60 (54–68)
Etiology				
HBV	1,366 (85.2%)	704 (87.7%)	170 (30.1%)*	529 (63%)
Others	238 (14.8%)	99 (12.3%)	394 (69.9%)*	311 (37%)
The largest tumor diameter (cm)	6.1 (3.8–9.8)	4.9 (3.0–7.9)	4 (2.9–6.2)	3.0 (1.9–4.9)
≤3 cm	262 (16.3%)	214 (26.6%)	357 (31.6%)	431 (51.3%)
>3, ≤7 cm	674 (42.0%)	337 (42.0%)	558 (49.4%)	297 (35.3%)
>7, ≤10 cm	302 (18.8%)	151 (18.8%)	141 (12.5%)	61 (7.3%)
>10 cm	366 (22.8%)	101 (12.6%)	74 (6.5%)	51 (6.1%)
Number of tumors				
1	919 (57.3%)	440 (54.8%)	324 (28.7%)	393 (46.8%)
2	346 (21.6%)	191 (23.8%)	242 (21.4%)	186 (22.1%)
≥3	339 (21.1%)	172 (21.4%)	564 (49.9%)	261 (31.1%)
Current BCLC stage				
A	982 (61.2%)	522 (65%)	488 (43.2%)	546 (65%)
B	622 (38.8%)	281 (35%)	642 (56.8%)	294 (35%)
AFP, ng/ml	112.4 (9–1,210)	78.3 (8–1,000)	20.9 (5.7–207.05)	26.8 (7.91–253.78)
Child-Pugh score				
5	1,239 (77.2%)	432 (68.2%)	652 (57.7%)	551 (65.6%)
6	289 (18%)	151 (23.9%)	310 (27.4%)	221 (26.3%)
7	76 (4.7%)	50 (7.9%)	168 (14.9%)	68 (8.1%)
ALBI grade				
1	799 (49.8%)	395 (49.2%)	376 (33.3%)	327 (38.9%)
2	782 (48.8%)	405 (50.4%)	702 (62.1%)	494 (58.8%)
3	22 (1.4%)	3 (0.4%)	52 (4.6%)	19 (2.3%)
ALT, U/L	39 (26–60)	37 (24–55)	54.43 (27–92.85)	37 (25–56.75)
AST, U/L	44 (21–65)	41.7 (30–62.1)	64 (39–101.74)	46 (32–67.75)
ALB, g/L	39.7 (36–43.4)	39.4 (35.9–42.8)	37 (33–41)	38 (34–41)
TBIL, μmol/L	15.6 (11.4–21.7)	15.4 (11.3–21.7)	15.4 (10.26–22)	13.68 (10.26–18.81)
INR	1.06 (1.00–1.13)	1.09 (1.01–1.17)	1.1 (1.0–1.22)	1.09 (1.03–1.21)
WBC, ×10 ⁹ /L	5.30 (4.03–6.61)	4.83 (3.80–6.31)	6.21 (4.56–7.84)	5.00 (3.88–6.20)
PLT, ×10 ⁹ /L	134 (87–186)	118 (77–174)	155 (104–218)	116 (80–162.8)
Cr, μmol/L	71 (61.2–82)	71 (62–81)	77.4 (61.6–89.1)	79.2 (68.2–88.4)
TACE procedures				
DEB-TACE	0	170 (21.2%)	50 (4.4%)	83 (9.9%)
cTACE	1,604 (100%)	633 (78.8%)	1,080 (95.6%)	757 (90.1%)
Sessions of TACE	3 (2–4)	3 (2–4)	2 (2–4)	3 (2–4)
Follow-up time, months	22 (11.9–34.1)	31.5 (18.4–44.5)	22.75 (12.3–37.4)	39 (19.3–92.2)

*Documented etiology was available in 564 (49.9%) patients; of these, 30.1%, 31.4%, 25.9%, 4.1%, and 8.5% were classified as HBV, HCV, alcohol, non-alcoholic steatohepatitis, and others, respectively. AFP, alpha-fetoprotein; ALB, albumin; ALBI, albumin-bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BCLC, Barcelona Clinic Liver Cancer; Cr, creatinine; cTACE, conventional transarterial chemoembolization; DEB-TACE, drug-eluting beads transarterial chemoembolization; INR, international normalized ratio; PLT, platelet; TACE, transarterial chemoembolization; TBIL, total bilirubin; WBC, white blood cell.

Table 2. Univariate and multivariate Cox regression analyses for overall survival in training cohort.

Risk factors	Univariable Cox regression		Multivariable Cox regression		Beta coefficient
	HR (95% CI)	p value	HR (95% CI)	p value	
Gender, refer to male	1.067 (0.874–1.303)	0.523			
Age, per 1 year increase	0.997 (0.991–1.003)	0.301			
Etiology, refer to HBV	0.997 (0.815–1.212)	0.980			
Tumor size, per 1 cm increase	1.116 (1.098–1.135)	<0.001	1.107 (1.088–1.126)	<0.001	0.102
Tumor number, refer to single	1.111 (1.074–1.149)	<0.001	1.101 (1.066–1.138)	<0.001	0.096
Log ₁₀ AFP, per 1 increase	1.222 (1.158–1.289)	<0.001	1.162 (1.102–1.225)	<0.001	0.150
ALBI score, per 1 score increase	1.158 (1.011–1.325)	0.034	1.145 (0.998–1.313)	0.053	0.135
ALT, per 1 U/L increase	1.001 (1.000–1.002)	0.057			
AST, per 1 U/L increase	1.002 (1.001–1.002)	<0.001	1.000 (0.998–1.001)	0.526	0.000
BUN, per 1 mmol/L increase	0.984 (0.948–1.021)	0.385			
Cr, per 1 μmol/L increase	0.997 (0.993–1.001)	0.198			
INR, per 1% increase	1.002 (0.815–1.233)	0.984			

AFP, alpha-fetoprotein; ALBI, albumin–bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; Cr, creatinine; HR, hazard ratio; INR, international normalized ratio; TBIL, total bilirubin.

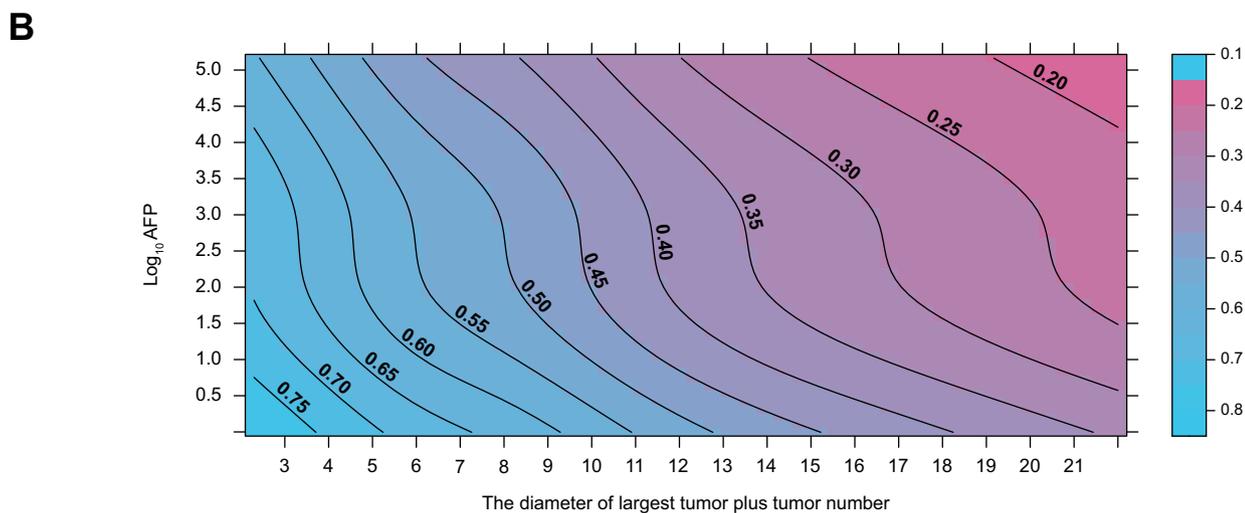
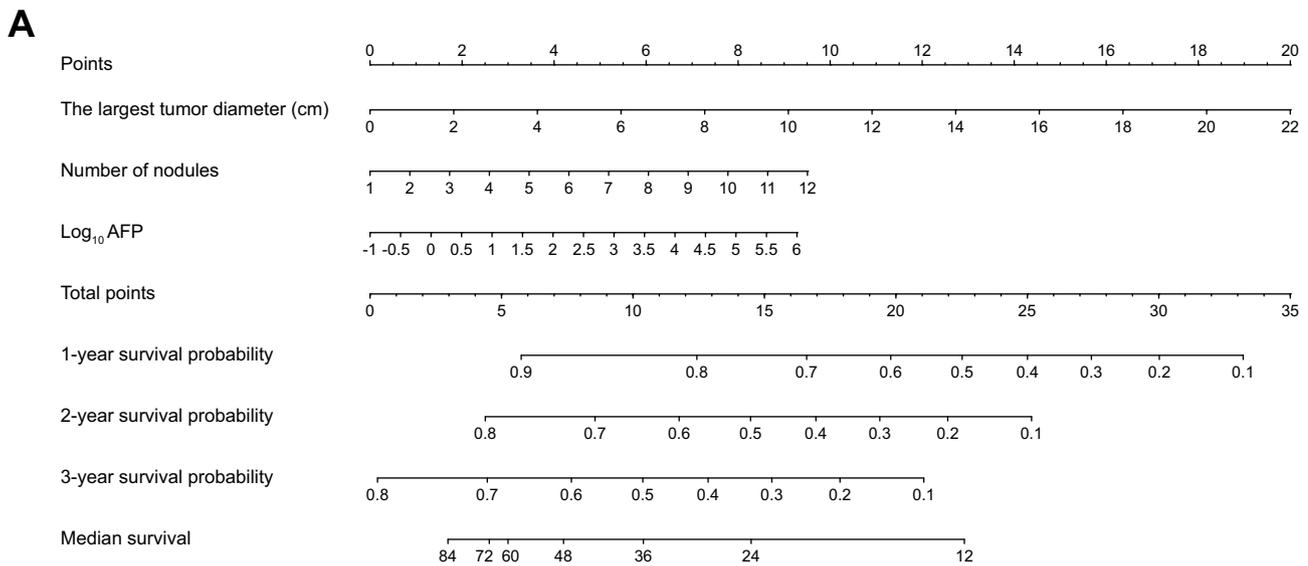


Fig. 2. Development of the prognostic model. (A) Nomogram of 6-and-12 model 2.0 for individual survival prediction; (B) contour plot of 3-year survival probability according to tumor burden (largest tumor diameter plus tumor number) and Log₁₀AFP. AFP, alpha-fetoprotein.

Table 3. Comparison of the performance and discrimination among currently available prognostic metrics.

Prognostic metric	Training cohort			Internal validation cohort			European validation cohort			Asian validation cohort		
	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC
6-and-12 model 2.0	0.674	0.001	10,621.96	10,636.06	0.643	0.013	6,158.23	6,171.00	0.630	0.010	10,712.64	10,726.95
6-and-12 model	0.664	0.010	10,648.69	10,653.31	0.628	0.014	6,173.16	6,177.42	0.608	0.010	10,712.64	10,754.45
Up to seven criteria	0.612	0.008	10,717.80	10,722.50	0.583	0.012	6,213.14	6,217.40	0.576	0.009	10,776.82	10,781.59
Four and seven criteria	0.611	0.009	10,729.45	10,734.15	0.590	0.011	6,204.62	6,208.87	0.565	0.009	10,780.24	10,785.01
Seven and eleven criteria	0.644	0.009	10,666.55	10,671.25	0.601	0.013	6,199.66	6,203.92	0.584	0.010	10,763.27	10,768.04
BCLC subclassification	0.587	0.008	10,744.42	10,749.12	0.588	0.012	6,191.56	6,195.81	0.579	0.009	10,775.98	10,780.75
HAP score	0.605	0.010	10,758.54	10,763.24	0.591	0.014	6,217.88	6,222.14	0.606	0.010	10,732.68	10,737.45
mHAP III score	0.653	0.010	10,683.01	10,687.70	0.637	0.013	6,174.48	6,178.74	0.622	0.010	10,729.18	10,733.95
mHAP II score	0.615	0.010	10,736.16	10,740.86	0.600	0.014	6,202.05	6,206.31	0.607	0.010	10,732.63	10,737.40
mHAP score	0.615	0.010	10,738.59	10,743.29	0.605	0.013	6,201.96	6,206.22	0.599	0.010	10,748.85	10,753.62
ALBI score	0.531	0.011	10,822.16	10,826.86	0.532	0.014	6,257.32	6,261.58	0.579	0.011	10,776.70	10,781.48

AIC, Akaike Information Criterion; ALBI, albumin–bilirubin; BIC, Bayesian Information Criterion; BCLC, Barcelona Clinic Liver Cancer; C-index, concordance index; HAP, Hepatoma arterial-embolization prognostic; SD, Standard deviation.

0.691 (SD, 0.050), and 0.684 (SD, 0.043) in the Asian validation cohort, respectively (Fig. S4B–D). The calibration performance across the full range of the new score showed the observed mortality and predicted probability of death in 3 years. In the training cohort, the observed mortality and predicted probabilities of death were similar, suggesting a good calibration (Fig. 3A), and the Hosmer–Lemeshow test did not indicate evidence of poor fit ($p = 0.147$). These results were consistent in the validation cohorts (Fig. 3B–D).

Comparison of the performances between the 6-and-12 model 2.0 and other metrics

The performance of the 6-and-12 model 2.0 and other models is summarized in Table 3. The current model had the highest C-index and lowest AIC and BIC values across these four datasets. The current model had the highest time-dependent C-index, suggesting a better discrimination and calibration ability (Fig. S5). Notably, these results remained consistent in different subgroups with different ages, sex, liver function, etiology, and Chinese DEB-TACE (Fig. S6 and Tables S6–S8). Additionally, compared with other metrics in the training cohort, the 6-and-12 model 2.0 showed an improvement in NRI with a range of 14.0% (95% CI, 2.80–21.4) to 37.2% (95% CI, 29.6–42.4), IDI with a range of 1.6% (95% CI, 0.7–2.9) to 8.3% (95% CI, 6.1–11.1) at the 1-year time point, and NRI with a range of 7.0% (95% CI, –3.4 to 15.8) to 26.2% (95% CI, 19.7–31.1) and IDI with a range of 1.4% (95% CI, 0.5–2.8) to 10.3% (95% CI, 7.4–13.5) at the 3-year time point (Table 4). Almost all values of NRI and IDI were significantly different, which indicated superior performance of the current model among these metrics. Improvement in NRI and IDI values were consistently observed in three validation cohorts and different subgroups, most of which were statistically significant (Tables S9–S19).

Risk stratification of the new model

The overall distribution of cases is shown in Fig. S7 (linear predictor) and Fig. S8 (baseline values of baseline log₁₀ AFP and tumor burden). The risk stratification of the current model with an X-tile plot showed that patients were separated into three risk strata of OS based on two optimal cutoff values (11.1 and 17 [11.1 rounded to the integer for easy application]): low risk (≤ 11), intermediate risk (11–17), and high risk (> 17), which showed a significantly different OS of 44.5 (95% CI, 40.4–48.6), 27.8 (95% CI, 24.2–31.4), and 15.3 months (95% CI, 13.2–17.4), respectively, in the training cohort, ($p < 0.001$) (Fig. S9).

For simplicity in clinical practice, a simplified, rounded version of the stratification is presented in Fig. 4A; this version was derived from the equation of linear predictors. AFP at baseline was expressed as an absolute value and divided into six levels (≤ 100 , 100–400, 400–2,000, 2,000–10,000, 10,000–40,000, and $> 40,000$ ng/ml). For each level of AFP, patients would be further stratified into low, intermediate, and high-risk strata according to the corresponding cutoffs of tumor burden ($t_s + t_n$), as follows:

- (1) If AFP was ≤ 100 ng/ml, the stratified criteria of tumor burden were 8 and 14, the division into risk groups should be low risk (≤ 8), intermediate risk (8–14), and high risk (> 14);

Table 4. Comparison of NRI and IDI between 6-and-12 model 2.0 and other currently available prognostic metrics (standard model) at 1-year and 3-year time point in the training cohort.

Prognostic metric	1-year survival time point			3-year survival time point		
	NRI (95% CI)	p value	IDI (95% CI)	NRI (95% CI)	p value	IDI (95% CI)
6-and-12 model	17.2% (8.60–22.8%)	<0.001	1.6% (0.7–2.9%)	11.0% (4.60–17.2%)	<0.001	1.4% (0.5–2.8%)
Up to seven criteria	18.0% (7.90–26.0%)	0.006	4.5% (2.7–6.8%)	15.4% (3.7–22.3%)	<0.001	4.4% (2.1–6.7%)
Four and seven criteria	20.8% (10.0–29.7%)	<0.001	4.5% (2.6–6.7%)	19.6% (10.1–27.5%)	<0.001	5.1% (2.9–7.5%)
Seven and eleven criteria	14.0% (2.80–21.4%)	0.014	2.4% (0.9–4.1%)	7.0% (-3.4–15.8%)	0.264	1.6% (0.1–3.4%)
BCLC subclassification	28.4% (17.1–35.1%)	<0.001	6.0% (3.9–8.5%)	15.3% (7.0–23.4%)	0.002	5.6% (3.0–8.4%)
HAP score	24.0% (16.2–32.4%)	<0.001	5.3% (3.5–7.5%)	26.0% (16.3–33%)	<0.001	7.4% (5.0–9.9%)
mHAP III score	24.3% (15.4–32.0%)	<0.001	2.5% (1.2–4.2%)	17.6% (9.7–24.9%)	<0.001	3.6% (2.2–5.4%)
mHAP II score	25.0% (15.1–32.4%)	<0.001	4.7% (2.9–6.9%)	20.5% (11.8–27.9%)	<0.001	6.0% (3.6–8.5%)
mHAP score	21.4% (13.4–31.1%)	<0.001	4.6% (2.9–6.8%)	21.6% (11.9–30%)	<0.001	5.9% (3.6–8.3%)
ALBI score	37.2% (29.6–42.4%)	<0.001	8.3% (6.1–11.1%)	26.2% (19.7–31.1%)	<0.001	10.3% (7.4–13.5%)

ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, Confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

- (2) If AFP was 100–400 ng/ml, the stratified criteria of tumor burden were 7 and 13 (low risk [≤ 7], intermediate risk [7–13], and high risk [> 13]);
- (3) If AFP was 400–2,000 ng/ml, the stratified criteria of tumor burden were 6 and 12 (low-risk [≤ 6], intermediate risk [6–12], and high risk [> 12]);
- (4) If AFP was 2,000–10,000 ng/ml, the stratified criteria of tumor burden were 5 and 11 (low risk [≤ 5], intermediate risk [5–11], and high risk [> 11]);
- (5) If AFP was 10,000–40,000 ng/ml, the stratified criteria of tumor burden were 4 and 10 (low risk [≤ 4], intermediate risk [4–10], and high risk [> 10]);
- (6) If AFP was $> 40,000$ ng/ml with any tumor burden, these patients were identified as high risk, because a low percentage of patients with an AFP $> 40,000$ ng/ml had a tumor burden of < 4 , and the median OS was similar to that of high-risk patients in other spectrums (Figs S8 and S10).

Given that no difference in median OS was observed among the same risk categories, we divided these patients into low-, intermediate-, and high-risk strata (Fig. S10). The median OS from low-risk to high-risk strata was 45.0 (95% CI, 40.1–49.9), 30.0 (95% CI, 26.1–33.9), and 15.4 months (95% CI, 13.4–17.4), respectively, in the training cohort ($p < 0.001$); 49.0 (95% CI, 43.0–55.1), 31.1 (95% CI, 28.9–33.4), and 18.8 months (95% CI, 15.5–22.1), respectively ($p < 0.001$) in the internal validation cohort; 31.4 (95% CI, 29.0–33.8), 19.8 (95% CI, 17.4–22.3), and 14.4 months (95% CI, 10.3–18.5), respectively ($p < 0.001$) in the European validation cohort; and 84.5 (95% CI, 70.9–98.0), 34.7 (95% CI, 26.0–43.4), and 20.1 months (95% CI, 11.8–28.4%), respectively ($p < 0.001$) in the Asian validation cohort (Fig. 4B–E).

Survival analysis in subgroups

The current model was able to stratify patients into the three strata mentioned above with significantly different OS values across subgroups, including patients of different sexes (male and female), age (≤ 60 years and > 60 years), ALBI grades (1 and 2), and etiologies (HBV and others), suggesting consistent performance in these populations (all $p < 0.001$; Fig. S11). The median survival and HRs with 95% CIs of the three strata in the different subgroups are detailed in Table S20. Additionally, significantly different OS in these three risk strata was consistently observed in patients with BCLC-A and BCLC-B HCC among these four cohorts and in a subgroup of Chinese DEB-TACE (all $p < 0.001$, Table S21 and Figs S12 and S13).

Discussion

In this international multicentre study, we present a 6-and-12 model 2.0 with extensions based on the original 6-and-12 model, which not only provided a preoperatively assessable, continuous method to refine outcome prediction and identify individual prognoses, but also divided the patients into low-, intermediate-, and high-risk strata with a significantly different OS. To the best of our knowledge, the current model has been derived and validated using the largest cohort of 4,377 candidates recommended for TACE. The strengths of this study are that: (1) it was conducted on the basis of the largest sample size to date, which allowed for estimates with narrow CIs; (2) it

adopted a continuous model instead of a categorized model presentation; and (3) the generalizability of the results was due to internal validation, Asian and European external validations, and confirmation in the different subgroups.

The accuracy and sensitivity of prognostic models are important for predicting outcomes in patients undergoing TACE. Although the C-index might not exhibit a satisfactory performance, it was consistently superior to other available tools using more variables. The final results of NRI suggest that the end-result of the adoption of AFP to the 6-and-12 model was that an additional 17.2% and 11% of patients who died within 1 and 3 years from TACE, respectively, would be identified, compared with the original 6-and-12 model (Table 4, similar results compared with other methods). Additionally, IDI could reflect the advantages or disadvantages of the model from the perspective of predictive probability increasing, and provide more accurate predicted outcomes based on probabilistic calculations rather than on subjective judgments. Significant improvement of IDI values indicated that the new model had positive effects and represented a meaningful improvement (Table 4). This increased sensitivity indicated that the benefits outweighed the harms of such an approach, providing clinical utility. More importantly, when greater multidimensional heterogeneity was introduced by different HCC cohorts, the performance of current model was consistently favorable among these available metrics. A model with poor specificity would not perform well across all the reported metrics, especially the calibration plots; estimates of calibration in the four cohorts showed that the model predicted mortality with

reasonable accuracy, which supports its generalization and application in clinical practice (Fig. 3).

Theoretically, adding other variables might improve the overall performance of the model, such as liver function and objective response.²⁵ In the current study, ALBI score was not identified as a predictor of OS for these populations, in whom liver function was at a relatively well-preserved level. Of note, objective response was beyond the scope of the present analysis to avoid misinterpretation of necrotic or enhancing residual areas across centers and to reduce observer bias, because modified Response evaluation criteria in solid tumors (mRECIST) reproducibility might be lower for HCC lesions with heterogeneous distribution of the viable tumor tissue.²⁶ Furthermore, assessing response is a dynamic process, and the optimal time point of response assessment is under debate.^{27,28} It is unlikely to ascertain time to best response in a given patient, and it is difficult to stick to follow-up on schedule for individual patients. Moreover, objective response lacked clinical applicability from an intention-to-treat perspective, models including this parameter can only be applied when response can be evaluated after the procedure, leading to a delay in the timing of stratification. Therefore, we elected to sacrifice absolute theoretical perfection in favor of simplicity and practical applicability for widespread clinical use. Indeed, the marginal C-index presented here underlines the need for further refinement of selection models. Future research to improve prediction performance is mandatory, and new predictive models integrating functional imaging and/or artificial intelligence could overcome this issue in the future.²⁹ While we

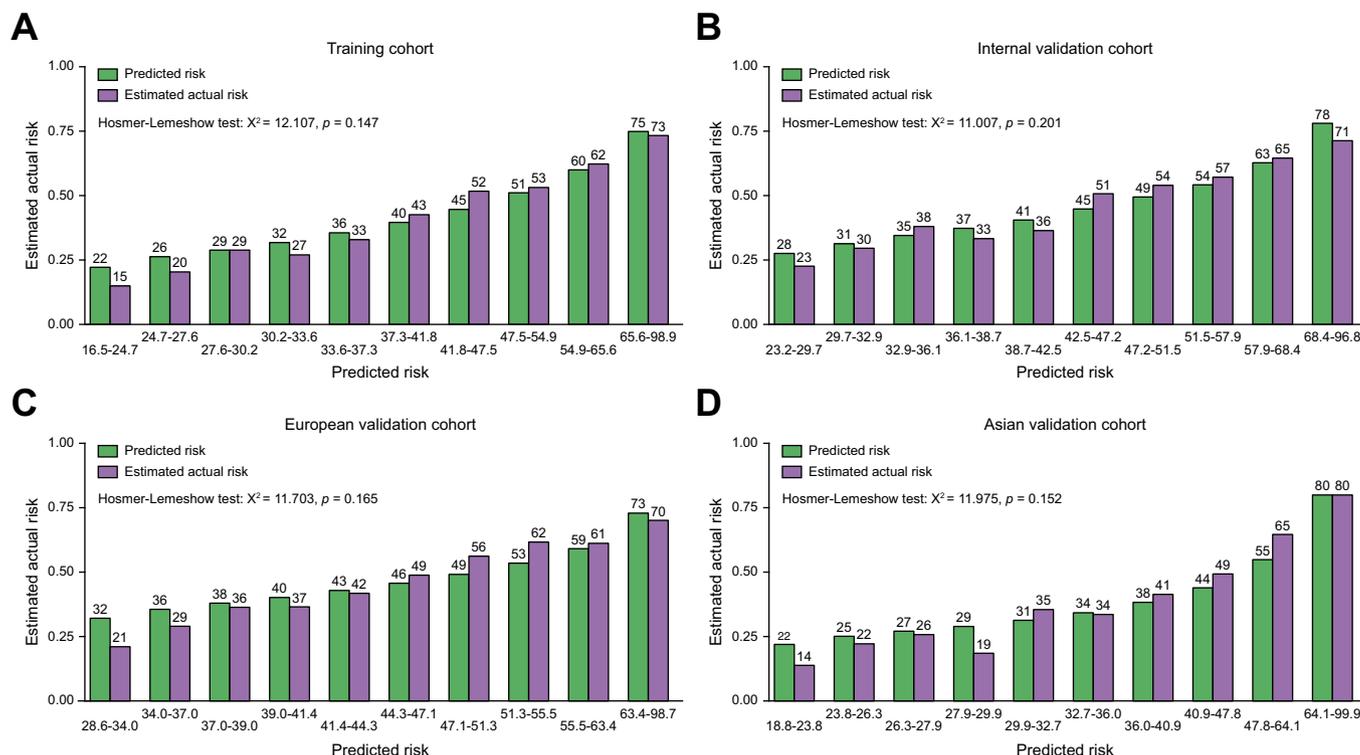


Fig. 3. Calibration analyses of 6-and-12 model 2.0 in candidates ideal for TACE. (A) Training cohort (Hosmer–Lemeshow test $\chi^2 = 12.107, p = 0.147$); (B) Internal validation cohort (Hosmer–Lemeshow test $\chi^2 = 11.007, p = 0.201$); (C) European validation cohort (Hosmer–Lemeshow test $\chi^2 = 11.703, p = 0.165$); and (D) Asian validation cohort (Hosmer–Lemeshow test $\chi^2 = 11.975, p = 0.152$). TACE, transarterial chemoembolization.

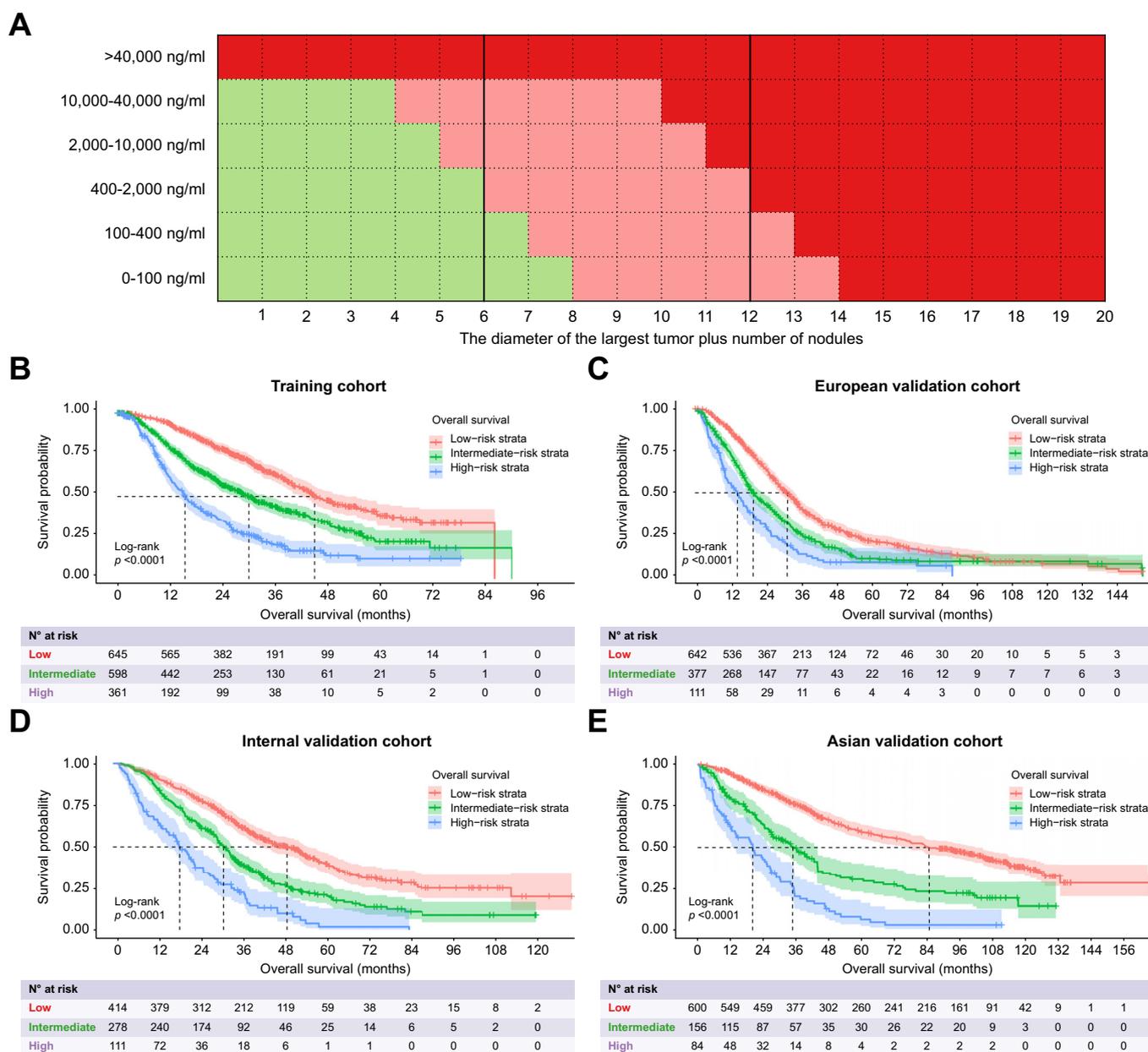


Fig. 4. Risk stratification of 6-and-12 model 2.0 and survival analyses by Kaplan–Meier method according to the stratification. (A) Risk stratification of 6-and-12 model 2.0. (B–E) Kaplan–Meier survival analysis of (B) Training cohort; (C) European validation cohort; (D) Internal validation cohort; and (E) Asian validation cohort (All p values < 0.001 for survival analyses by log-rank test among four cohorts).

wait for the development of next-generation predictive tools, the current model offers a reasonable, reproducible, and user-friendly alternative, although any expansion of targeted populations should be done with caution.

Accurate outcome prediction in patients is crucial to identify candidates who will benefit the most from treatment, while distinguishing those who are more likely to have an unfavorable prognosis. On the one hand, a prognostic score should preferably use objective and easily accessible clinical indicators to predict disease outcomes for simple and accurate clinical applications. The current model is a continuous model incorporating three commonly used parameters: tumor size, tumor number, and AFP value. Using a nomogram or a web-based

calculator, estimated survival probability prediction at different time points and median survival time in any individual patient (ideal candidates for TACE) can be objectified, which is more informative, compared with binary scores or risk-point systems. The key point is that this is the outcome prediction to be considered when recommending a specific treatment for a given patient as well as when informing about life expectancy. On the other hand, the current model was developed to further stratify optimal TACE candidates with a significant tumor burden and AFP heterogeneity that are unclassified by the BCLC system. When adopting the current stratification, patients with BCLC-A or BCLC-B HCC could be further stratified into three risk strata with significantly different OS (Fig. S12),

which indicated further clinical use on the basis of the BCLC staging system. This could also provide a referential framework to control heterogeneity and define the target population for improved clinical decision-making in future trial design. The original 6-and-12 model has been incorporated into the international randomized control trials LEAP-012 (NCT04246177) and TALENTACE (NCT04712643) as a stratification factor or inclusion criterion to balance the heterogeneity of tumor burden. If the inclusion criteria for clinical trials with TACE as a comparator do not carefully define target populations, any positive or negative outcome might be the result of an under-estimation of the baseline assumptions and events to register during follow-up.⁵ Moreover, the heterogeneous median OS, ranging from 15.4 months in the high-risk group to 30 and 45 months in the intermediate-, and low-risk group, respectively, indicated that some patients might not benefit from TACE, calling for further investigations on novel treatment strategies for these patients with poor outcomes. However, this conclusion should be interpreted with caution because of the studied populations and retrospective survival estimates. Interestingly, we also identified a subgroup of patients with baseline AFP values >40,000 ng/ml as high risk, because of poor prognosis, suggesting that the AFP value could serve as a tool for the initial selection of candidates appropriate for TACE. For other metrics containing a dichotomous AFP with a cutoff of 200 ng/ml or 400 ng/ml, the present stratification can be increased to six levels of AFP with progressive burden restrictions, which would improve its flexibility for clinical application.

Our study had some limitations. First, because of the retrospective study design, selection bias was unavoidable, and the relatively wide accrual time characterizing some of the

recruited cohorts should be acknowledged. However, our survival analysis was robustly built on the process of independent cross-validation in large multi-institutional cohorts, which might limit the potential for selection bias. Second, despite important advances, the types and doses of embolization material, chemotherapeutic agent, degree of selectivity, and endpoint varied from center to center, from East to West.^{30,31} However, there is no robust evidence supporting the superiority of any embolization material or chemotherapeutic agent over others.^{30,32} All participating centers were well experienced in performing TACE, and the principles of on-demand selective/supersselective TACE with an embolization endpoint of reduced or disappeared tumor arterial flow were strictly followed. Third, given the relatively long time over which TACE approaches and patient selection have evolved, and with advances in systemic treatment, the prognosis of patients with unresectable HCC has improved; however, despite combinations with molecular-targeted drugs or the introduction of new embolization materials, no significant improvement in OS have been observed over the past 20 years (Table S22 and Fig. S14). Thus, prospective studies to validate the performance of the model are warranted in the era of immune-targeted therapy.

In conclusion, in this international multicentre study, we developed and validated a novel, accessible, continuous model (6-and-12 model 2.0) for candidates recommended for TACE; the model showed superior discriminatory ability and goodness of fit in predicting outcomes, and could stratify patients into three risk strata with significantly different survival prognoses. Thus, it could be used as a tool for individual survival prediction and as a referential framework to control study heterogeneity and define the target population in future trial designs.

Affiliations

¹State Key Laboratory of Holistic Integrative Management of Gastrointestinal Cancers and National Clinical Research Center for Digestive Diseases, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, China; ²Department of Liver Diseases and Interventional Radiology, National Clinical Research Centre for Digestive Diseases, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, China; ³Department of Liver Diseases and Interventional Radiology, Digestive Diseases Hospital, Xi'an International Medical Center Hospital, Northwest University, Xi'an, China; ⁴Department of Cardiology, Tangdu Hospital, Fourth Military Medical University, Xi'an, China; ⁵Department of Radiology, Seoul National University Hospital, Seoul, Korea; ⁶Department of Gastroenterology and Hepatology, Hôpital Saint-Joseph, Marseille, France; ⁷Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckst, Mainz, Germany; ⁸Department for Interventional Radiology, University Hospital of Lübeck, Ratzeburger Allee 160, Lübeck, Germany; ⁹Department of Hepatobiliary Surgery, Southwest Hospital, Third Military Medical University, Chongqing, China; ¹⁰Department of Liver Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ¹¹Gastroenterology and Hepatology Unit, Division of Internal Medicine, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand; ¹²Department of Hepatobiliary and Pancreatic Interventional Cancer, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China; ¹³Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea; ¹⁴Department of Minimally Invasive International Therapy, The Third Affiliated Hospital of Kunming University, Tumor Hospital of Yunnan Province, Kunming, China; ¹⁵Department of Interventional Radiology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China; ¹⁶Department of Infectious Disease, General Hospital of Ningxia Medical University, Yinchuan, China; ¹⁷Department of Interventional Radiology, Jiangsu Provincial Cancer Hospital, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China; ¹⁸Department of Interventional Radiology, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China; ¹⁹Department of Interventional Radiology, The Affiliated Hospital of Nantong University, Nantong, China; ²⁰Department of Gastroenterology and Hepatology, Centre Hospitalo-Universitaire de Nancy, France; ²¹Department of Hepatobiliary Surgery, Xinqiao Hospital, Third Military Medical University, Chongqing, China; ²²Department of Interventional Radiology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; ²³Department of Interventional Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China; ²⁴Department of Oncology, The Second Affiliated Hospital of Nanchang University, Nanchang, China; ²⁵Department of Gastroenterology and Hepatology, Shandong Province Hospital Affiliated to Shandong University, Jinan, China; ²⁶Department of Interventional Radiology, Tangdu Hospital, Fourth Military Medical University, Xi'an, China; ²⁷Interventional Medical Centre of the Affiliated Hospital of Qingdao University, Qingdao, China; ²⁸Department of Interventional Radiology, the First Affiliated Hospital of Fujian Medical University, Fuzhou, China; ²⁹Department of Infectious Diseases, the 910 Hospital of the Chinese People's Liberation Army Joint Logistic Support Force, Quanzhou, China; ³⁰Department of Hepatic & Biliary & Pancreatic Surgery, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; ³¹Department of Gastroenterology and Hepatology, Hôpital Universitaire de l'Archet Nice, France; ³²Department of Interventional Therapy, Shandong Tumor Hospital, Jinan, China; ³³Department of Interventional Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; ³⁴Department of Radiology, Zhejiang Cancer Hospital, Hangzhou, China; ³⁵Department of Interventional Radiology, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China; ³⁶Department of Interventional Radiology and Vascular Surgery, Hunan Provincial People's Hospital, Changsha, China; ³⁷Department of Interventional Radiology, The Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, China; ³⁸Department of Interventional Radiology, Yantai Yuhuangding Hospital, Yantai, China; ³⁹Department of Medical Imaging, Nanjing General Hospital of the Nanjing Military Command, Nanjing, China; ⁴⁰Department of Interventional Medicine, The First Affiliated Hospital of Lanzhou University, Lanzhou, China; ⁴¹Department of Interventional Radiology, Peking University Cancer Hospital, Beijing, China; ⁴²Department of Digestive Medicine, Xi'an No.3 Hospital, Xi'an, China; ⁴³Department of Medical Affairs, Air Force Hospital of Western Theater Command, Chengdu, China; ⁴⁴Department of General Surgery, The Southern Theater Air Force Hospital, Guangzhou, China; ⁴⁵Department of Rheumatology and Immunology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; ⁴⁶Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China; ⁴⁷Department of Health

Statistics, Fourth Military Medical University, Xi'an, China; ⁴⁸Department of Medicine II, University Medical Center Freiburg, Freiburg, Germany; ⁴⁹Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; ⁵⁰Division of Gastroenterology and Hepatology, Toronto General Hospital Medical Oncology, Princess Margaret Cancer Centre, Schwartz Reisman Liver Research Centre, Toronto, Canada

Abbreviations

AFP, alpha-fetoprotein; AIC, Akaike Information Criterion; ALB, albumin; ALBI, albumin–bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under the receiver operating characteristics curve; BCLC, Barcelona Clinic Liver Cancer; BIC, Bayesian Information Criterion; BUN, blood urea nitrogen; C-index, concordance index; Cr, creatinine; CT, computed tomography; DEB, drug-eluting bead; HAP, hepatoma arterial-embolization prognostic; HCC, hepatocellular carcinoma; HR, hazard ratio; IDI, integrated discrimination improvement; mRECIST, modified Response evaluation criteria in solid tumors; NRI, net reclassification improvement; OR, odds ratio; OS, overall survival; PLT, platelet; TACE, transarterial chemoembolization; TLBL, total bilirubin; WBC, white blood cell.

Financial support

This study was supported by grants from National Key Technology R&D Program (2015BAI13B07) to D.F. and from Bethune Liver Cancer Research Program (X-J-2020-029) to G.H.

Conflicts of interest

AV: consulting fees (AstraZeneca, Amgen, BeiGene, Böhringer Mannheim, BMS, BTG, Daichi-Sankyo, Eisai, Incyte, Ipsen, MSD, PierreFabre, Roche, Servier, Sirtex, Tahio, and Terumo); payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing, or educational events [AstraZeneca, Amgen, BeiGene, Böhringer Mannheim, Bristol Myers Squibb, BTG, Daichi-Sankyo, Eisai, GSK, Imaging Equipment Ltd (AAA), Incyte, Ipsen, Jiangsu Hengrui Medicines; MSD, PierreFabre, Roche, Servier, Sirtex, Tahio, and Terumo]; support for attending meetings and/or travel (Roche, MSD, and Astellas); participation on a data safety monitoring board or advisory board (AstraZeneca, Amgen, BeiGene, Böhringer Mannheim, BMS, BTG, Daichi-Sankyo, Eisai, Incyte, Ipsen, MSD, PierreFabre, Roche, Servier, Sirtex, Tahio, and Terumo). RK: grants or contracts from any entity (German Federal Ministry of Research & Education and DFG – German Research Foundation); consulting fees (Boston Scientific, Bristol Myers Squibb, Guerbet, Roche, and Sirtex); payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing, or educational events (Astra Zeneca, BTG, Eisai, Guerbet, Ipsen, Roche, Siemens, Sirtex, MSD Sharp & Dohme); participation on a data safety monitoring board or advisory board (ABC HCC Trial); leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid (European Society of Radiology, Chair of the Audit and Standards Subcommittee – unpaid; and Eurosafe Imaging, Steering Committee - unpaid); JWC: grants or contracts from any entity (educational grant from Guerbet); DB: grants or contracts from any entity (German Research Foundation and Dr Rolf M. Schwiete Stiftung); payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events (Falk Foundation and Gore); support for attending meetings and/or travel (Gilead Science and Abbvie); X.A.: grants or contracts from any entity (Servier and Ipsen); consulting fees (Bayer and Ipsen); payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events (Gilead, Servier); support for attending meetings and/or travel (Roche and Gilead); the other authors have nothing to disclose.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Study design: DX, GH. Acquisition of data: DX, WB, QW, JWC, XA, RK, HZ, YZ, PS, CN, SK, MH, WH, XD, GY, HL, HZ, J-PB, JinL, JiaL, XZ, JW, CZ, WG, ZLi, ZLin, TX, TY, RA, JS, HS, GS, WR, YZ, SY, YZ, JX, WW, XZ, YF, CL, AK, RD, JZ, SL, HY, LZ, NY, WF, SZ, LF, GW, PZ, XL, JC, FZ, WS, WZ, HZ, GC, WH, WJ, WZ, LL, AF, EW, ZW, DH, YL, JS, XL, JY, ZW, BL, KL, WG, ZY, DB, AV, GH. Analysis and interpretation of data: DX, GH. Drafting and revising the manuscript: all authors. Critical revision of the manuscript for important intellectual content: DX, QW, YZ, JX, KW, GH. Statistical analysis: DX, JS, JX. Administrative and material support: DF, GH.

Data availability statement

All data, materials, and methods in this study can be made available from the corresponding author upon request for non-commercial purposes and after approval of a study proposal through a signed data access agreement.

Acknowledgements

The authors thank the patients and their families for their contributions, the subinvestigators from each participant center for their efforts, and Philip J. Johnson for his assistance in contacting European investigators.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhepr.2024.101216>.

References

Author names in bold designate shared co-first authorship

- [1] European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2018;69:182–236.
- [2] Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 2018;67:358–380.
- [3] Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol* 2022;76:681–693.
- [4] Singal AG, Llovet JM, Yarchoan M, et al. AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. *Hepatology* 2023;78:1922–1965.
- [5] **Burrel M, Reig M**, Forner A, et al. Survival of patients with hepatocellular carcinoma treated by transarterial chemoembolisation (TACE) using Drug Eluting Beads. Implications for clinical practice and trial design. *J Hepatol* 2012;56:1330–1335.
- [6] Sangro B, Salem R. Transarterial chemoembolization and radioembolization. *Semin Liver Dis* 2014;34:435–443.
- [7] Galle PR, Tovoli F, Foerster F, et al. The treatment of intermediate stage tumours beyond TACE: from surgery to systemic therapy. *J Hepatol* 2017;67:173–183.
- [8] Llovet JM, Lencioni R. mRECIST for HCC: performance and novel refinements. *J Hepatol* 2020;72:288–306.
- [9] **Kudo M, Ueshima K**, Chan S, et al. Lenvatinib as an initial treatment in patients with intermediate-stage hepatocellular carcinoma beyond Up-To-Seven Criteria and Child-Pugh A Liver Function: a proof-of-concept study. *Cancers (Basel)* 2019;11:1084.
- [10] Llovet JM, Villanueva A, Marrero JA, et al. Trial design and endpoints in hepatocellular carcinoma: AASLD Consensus Conference. *Hepatology* 2021;73(Suppl 1):158–191.
- [11] Mazzaferro V, Llovet JM, Miceli R, et al. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol* 2009;10:35–43.
- [12] Bolondi L, Burroughs A, Dufour JF, et al. Heterogeneity of patients with intermediate (BCLC B) hepatocellular carcinoma: proposal for a subclassification to facilitate treatment decisions. *Semin Liver Dis* 2012;32:348–359.
- [13] Kadalayil L, Benini R, Pallan L, et al. A simple prognostic scoring system for patients receiving transarterial embolisation for hepatocellular cancer. *Ann Oncol* 2013;24:2565–2570.
- [14] Yamakado K, Miyayama S, Hirota S, et al. Subgrouping of intermediate-stage (BCLC stage B) hepatocellular carcinoma based on tumor number and size and Child-Pugh grade correlated with prognosis after transarterial chemoembolization. *Jpn J Radiol* 2014;32:260–265.
- [15] Pinato DJ, Arizumi T, Allara E, et al. Validation of the hepatoma arterial embolization prognostic score in European and Asian populations and proposed modification. *Clin Gastroenterol Hepatol* 2015;13:1204–1208.
- [16] Cappelli A, Cucchetti A, Cabibbo G, et al. Refining prognosis after transarterial chemo-embolization for hepatocellular carcinoma. *Liver Int* 2016;36:729–736.
- [17] Park Y, Kim SU, Kim BK, et al. Addition of tumor multiplicity improves the prognostic performance of the hepatoma arterial-embolization prognostic score. *Liver Int* 2016;36:100–107.
- [18] **Pinato DJ, Sharma R**, Allara E, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. *J Hepatol* 2017;66:338–346.

- [19] Wang Q, Xia D, Bai W, et al. Development of a prognostic score for recommended TACE candidates with hepatocellular carcinoma: a multicentre observational study. *J Hepatol* 2019;70:893–903.
- [20] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–141.
- [21] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med* 2015;162:735–736.
- [22] Lammer J, Malagari K, Vogl T, et al. Prospective randomized study of doxorubicin-eluting-bead embolization in the treatment of hepatocellular carcinoma: results of the PRECISION V study. *Cardiovasc Intervent Radiol* 2010;33:41–52.
- [23] Golfieri R, Giampalma E, Renzulli M, et al. Randomised controlled trial of doxorubicin-eluting beads vs conventional chemoembolisation for hepatocellular carcinoma. *Br J Cancer* 2014;111:255–264.
- [24] Hung YW, Lee IC, Chi CT, et al. Redefining tumor burden in patients with intermediate-stage hepatocellular carcinoma: the Seven-Eleven Criteria. *Liver Cancer* 2021;10:629–640.
- [25] Han G, Berhane S, Toyoda H, et al. Prediction of survival among patients receiving transarterial chemoembolization for hepatocellular carcinoma: a response-based approach. *Hepatology* 2020;72:198–212.
- [26] Seyal AR, Gonzalez-Guindalini FD, Arslanoglu A, et al. Reproducibility of mRECIST in assessing response to transarterial radioembolization therapy in hepatocellular carcinoma. *Hepatology* 2015;62:1111–1121.
- [27] Wang W, Zhao Y, Bai W, et al. Response assessment for HCC patients treated with repeated TACE: the optimal time-point is still an open issue. *J Hepatol* 2015;63:1530–1531.
- [28] Xia D, Wang Q, Bai W, et al. Optimal time point of response assessment for predicting survival is associated with tumor burden in hepatocellular carcinoma receiving repeated transarterial chemoembolization. *Eur Radiol* 2022;32:5799–5810.
- [29] Hong G, Suh KS, Suh SW, et al. Alpha-fetoprotein and (18)F-FDG positron emission tomography predict tumor recurrence better than Milan criteria in living donor liver transplantation. *J Hepatol* 2016;64:852–859.
- [30] de Baere T, Ronot M, Chung JW, et al. Initiative on superselective conventional transarterial chemoembolization results (INSPIRE). *Cardiovasc Intervent Radiol* 2022;45:1430–1440.
- [31] Craig P, Young S, Golzarian J. Current trends in the treatment of hepatocellular carcinoma with transarterial embolization: variability in technical Aspects. *Cardiovasc Intervent Radiol* 2019;42:1322–1328.
- [32] Llovet JM, De Baere T, Kulik L, et al. Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2021;18:293–313.

Keywords: Hepatocellular carcinoma; Transarterial chemoembolization; Tumor burden; Alpha-fetoprotein; Risk stratification.

Received 15 December 2023; received in revised form 30 August 2024; accepted 5 September 2024; Available online 12 September 2024

Supplemental information

Tumor burden with AFP improves survival prediction for TACE-treated patients with HCC: An international observational study[☆]

Dongdong Xia, Wei Bai, Qiuhe Wang, Jin Wook Chung, Xavier Adhoute, Roman Kloeckner, Hui Zhang, Yong Zeng, Pimsiri Sripongpun, Chunhui Nie, Seung Up Kim, Ming Huang, Wenhao Hu, Xiangchun Ding, Guowen Yin, Hailiang Li, Hui Zhao, Jean-Pierre Bronowicki, Jing Li, Jiaping Li, Xiaoli Zhu, Jianbing Wu, Chunqing Zhang, Weidong Gong, Zixiang Li, Zhengyu Lin, Tao Xu, Tao Yin, Rodolphe Anty, Jinlong Song, Haibin Shi, Guoliang Shao, Weixin Ren, Yongjin Zhang, Shufa Yang, Yanbo Zheng, Jian Xu, Wenhui Wang, Xu Zhu, Ying Fu, Chang Liu, Apichat Kaewdech, Rong Ding, Jie Zheng, Shuaiwei Liu, Hui Yu, Lin Zheng, Nan You, Wenzhe Fan, Shuai Zhang, Long Feng, Guangchuan Wang, Peng Zhang, Xueda Li, Jian Chen, Feng Zhang, Wenbo Shao, Weizhong Zhou, Hui Zeng, Gengfei Cao, Wukui Huang, Wenjin Jiang, Wen Zhang, Lei Li, Aiwei Feng, Enxin Wang, Zhexuan Wang, Dandan Han, Yong Lv, Jun Sun, Bincheng Ren, Linying Xia, Xiaomei Li, Jie Yuan, Zhengyu Wang, Bohan Luo, Kai Li, Wengang Guo, Zhanxin Yin, Yan Zhao, Jielai Xia, Daiming Fan, Kaichun Wu, Dominik Bettinger, Arndt Vogel, Guohong Han, and on behalf of China HCC-TACE study group

**Tumor burden with AFP improves survival prediction for
TACE-treated patients with HCC: An international
observational study**

Dongdong Xia, Wei Bai, Qiuhe Wang, Jin Wook Chung, Xavier Adhoute, Roman Kloeckner, Hui Zhang, Yong Zeng, Pimsiri Sripongpug, Chunhui Nie, Seung up Kim, Ming Huang, Wenhao Hu, Xiangchun Ding, Guowen Yin, Hailiang Li, Hui Zhao, Jean-Pierre Bronowicki, Jing Li, Jiaping Li, Xiaoli Zhu, Jianbing Wu, Chunqing Zhang, Weidong Gong, Zixiang Li, Zhengyu Lin, Tao Xu, Tao Yin, Rodolphe Anty, Jinlong Song, Haibin Shi, Guoliang Shao, Weixin Ren, Yongjin Zhang, Shufa Yang, Yanbo Zheng, Jian Xu, Wenhui Wang, Xu Zhu, Ying Fu, Chang Liu, Apichat Kaewdech, Rong Ding, Jie Zheng, Shuaiwei Liu, Hui Yu, Lin Zheng, Nan You, Wenzhe Fan, Shuai Zhang, Long Feng, Guangchuan Wang, Peng Zhang, Xueda Li, Jian Chen, Feng Zhang, Wenbo Shao, Weizhong Zhou, Hui Zeng, Gengfei Cao, Wukui Huang, Wenjin Jiang, Wen Zhang, Lei Li, Aiwei Feng, Enxin Wang, Zhexuan Wang, Dandan Han, Yong Lv, Jun Sun, Bincheng Ren, Linying Xia, Xiaomei Li, Jie Yuan, Zhengyu Wang, Bohan Luo, Kai Li, Wengang Guo, Zhanxin Yin, Jielai Xia, Daiming Fan, Kaichun Wu, Dominik Bettinger, Arndt Vogel, Guohong Han, on behalf of China HCC-TACE study group.

Table of contents

Methods	4
Statistical analysis.....	7
Table S1	8
Table S2	10
Table S3	11
Table S4	12
Table S5.....	13
Table S6	15
Table S7	17
Table S8	19
Table S9	20
Table S10	21
Table S11	22
Table S12	23
Table S13	24
Table S14	25
Table S15	26
Table S16	27
Table S17	28
Table S18	29
Table S19	30
Table S20	31
Table S21	32
Table S22	33
Fig. S1	37
Fig. S2	34
Fig. S3	35
Fig S4	36
Fig. S5	38
Fig S6	39

Fig S8	42
Fig. S9	43
Fig. S10	44
Fig. S11	45
Fig. S12	46
Fig.S13	48
Fig. S14.	49
References	50

Methods

Detailed primary investigators and number of eligible patients were summarized in **Table S1**, and **Fig. 1** showed the flow chart of this study.

Training dataset. This dataset comprised 1604 eligible cases after screening 3819 patients with HCC undergoing conventional TACE (cTACE) from 24 Chinese academic centres between January 2010 and May 2016. In contrast with the previous study, we used entire cohort to derive the model, not randomized splitting into training and validation cohort. The data of the training cohort have been published in Journal of Hepatology [1].

Internal validation dataset. A total of 3496 consecutive patients who underwent cTACE from another five centres (between January 2010 and December 2017, n=2386) and drug-eluting beads TACE (DEB-TACE) from seven centres (between January 2016 and June 2019, n=1110) were retrospectively screened. Parameters, including baseline demographics, tumor characteristics, laboratory testing and TACE procedures, were collected by two independent investigators using a previously reported method [1]. Finally, a total of 803 patients were enrolled to analysis. These data have never been published previously.

External validation dataset. Finally, as shown in Fig S1-C, European dataset consisted of 1,130 eligible and anonymous cases at 6 centers in two countries, the French cohort of 362 patients was consisted of three datasets from Marseille (252 patients), Nancy (72 patients), and Nice (38 patients); and the Germany cohort of 768 patients was consisted of three datasets from Mainz (113 patients), Hannover (242 patients) and Freiburg (413 patients). The Asian dataset was obtained from three centers with 840 eligible and anonymous cases (442 and 187 patients from SNUH and Yonsei, Korea; 211 patients from Songkla, Thailand). These datasets with the same parameters were collected by the primary investigators and their colleagues at each center, including age, sex, aetiology, previous treatment (yes/no), ECOG score, tumor characteristics (ts and tn), liver function (Child–Pugh score and albumin-bilirubin [ALBI] score), and laboratory tests (including AFP value; the international normalized ratio [INR]; levels of alanine aminotransferase [ALT], aspartate aminotransferase [AST], albumin, total bilirubin, creatinine; white blood cell count (WBC), platelet count (PLT) level), and TACE procedures (DEB-TACE or cTACE, superselective or not, and total sessions of TACE). The French cohort of 362 patients from Marseille (252), Nancy (72), and Nice (38) have been published in the following journal: World journal of hepatology (World J Hepatol 2020 August 27; 12(8): 0-0)[2]; World Journal of Clinical Cases (World J Clin Cases 2021 June 26; 9(18): 4559-4572)[3]; European Journal of Gastroenterology and Hepatology (Eur J Gastroenterol Hepatol. 2019 Nov;31(11):1414-1423)[4]. Part

of Thailand cohort was published in Clinical Translational Gastroenterology (Clin Transl Gastroenterol. 2021 Feb 18;12(2): e00310) [5]. Part of Germany cohort was published in Frontiers in Oncology (Front Oncol. 2022 Feb 23;12:850454.)[6].

Statistical analysis

Multiple imputation by chained equations (MICE) was used to impute missing outcome data after adjustment for all measured variables potentially associated with missing data. Our intention was to include all factors that could be associated with missingness. Pattern and percent of missing value were depicted in **Fig. S1**.

Table S2-S3 summarized correlation coefficient between these indicator variables with missing values, and correlation coefficient between variables with missing values and other observable variables, respectively. The correlation coefficient is not particularly large, indicating that the data is less likely to be pattern of Missing Completed at Random (MCAR) and more likely to be pattern of Missing at Random, which suggests a multiple imputation is needed. Then, we produced 5 datasets (C1-C5, Table S4) with imputed missing values and non-missing values consistent with the observed data using the MICE processes. Each of the 5 datasets were used to analyze the primary outcome. The estimated coefficients and standard errors from the 5 models were combined into a final estimated coefficient and standard error using robust methods. We used R to implement the multiple imputation with packages of "VIM", "survival", "ggplot2", "survminer", and "mice".

Table S1: Summarization of participated centers, primary investigator and number of eligible patients at each center.

Datasets	Participated centers	City	Country	Primary investigator	No.
Training (N=1604)	Xijing Hospital	Xi'an	China	Han GH	211
	First Affiliated Hospital of Fujian Medical University	Fuzhou	China	Lin ZY	36
	Hunan Provincial People's Hospital	Changsha	China	Zhang YJ	25
	The Affiliated Cancer Hospital of Zhengzhou University	Zhengzhou	China	Li HL	90
	The First Affiliated Hospital of Nanjing Medical University	Nanjing	China	Shi HB	29
	The Affiliated Cancer Hospital of Nanjing Medical University	Nanjing	China	Yin GW	117
	The First Affiliated Hospital of Lanzhou University	Lanzhou	China	Wang WH	14
	The Second Affiliated Hospital of Nanchang University	Nanchang	China	Wu JB	48
	Nanjing General Hospital of the Nanjing Military Command	Nanjing	China	Xu J	18
	The Affiliated Hospital of Nantong University	Nantong	China	Zhao H	69
	The Affiliated Hospital of Qingdao University	Qingdao	China	Li ZX	39
	The 910 Hospital of the Chinese People's Liberation Army Joint Logistic Support Force	Quanzhou	China	Xu T	35
	Shandong Province Hospital Affiliated to Shandong University	Jinan	China	Zhang CQ	47
	Shandong Tumor Hospital	Jinan	China	Song JL	31
	The First Affiliated Hospital of Soochow University	Suzhou	China	Zhu XL	49
	Tangdu Hospital, Fourth Military Medical University	Xi'an	China	Gong WD	41
	The Affiliated Tumor Hospital of Xinjiang Medical University	Urumqi	China	Yang SF	21
	Southwest Hospital, Third Military Medical University	Chongqing	China	Zhang H	164
	Xinqiao Hospital, Third Military Medical University	Chongqing	China	Li J	67
	The Third Affiliated Hospital of Kunming University	Kunming	China	Huang M	164
Yantai Yuhuangding Hospital	Yantai	China	Zheng YB	20	

	The First Affiliated Hospital of Zhejiang University	Hangzhou	China	Nie CH	197
	Zhejiang Cancer Hospital	Hangzhou	China	Shao GL	29
	The First Affiliated Hospital of Sun Yat-sen University	Guangzhou	China	Li JP	43
Internal validation (N=633, cTACE)	West China Hospital	Chengdu	China	Zeng Y	278
	Hubei Cancer Hospital	Wuhan	China	Yin T	33
	The Affiliated Tumor Hospital of Xinjiang Medical University	Urumqi	China	Ren WX	26
	General Hospital of Ningxia Medical University	Yinchuan	China	Ding XC	144
	The First Affiliated Hospital of Wenzhou Medical University	Wenzhou	China	Hu WH	152
Internal validation (N=170, DEB-TACE)	Peking University Cancer Hospital	Beijing	China	Zhu X	13
	The Affiliated Hospital of Qingdao University	Qingdao	China	Li ZX	11
	Southwest Hospital, Third Military Medical University	Chongqing	China	Zhang H	31
	The Third Affiliated Hospital of Kunming University	Kunming	China	Huang M	4
	The First Affiliated Hospital of Zhejiang University	Hangzhou	China	Nie CH	91
	The First Affiliated Hospital of Sun Yat-sen University	Guangzhou	China	Li JP	12
	The Second Affiliated Hospital of Nanchang University	Nanchang	China	Wu JB	8
European validation (N=1130)	Hôpital Saint-Joseph	Marseille	France	Adhoute	252
	Centre Hospitalo-Universitaire de Nancy	Nancy	France	Bronowicki	72
	Hôpital Universitaire de l'Archet Nice	Nice	France	Anty	38
	University Medical Center of the Johannes Gutenberg University Mainz	Mainz	Germany	Kloeckner	113
	Hannover Medical School	Hannover	Germany	Vogel	242
	University Medical Center Freiburg	Freiburg	Germany	Bettinger	413
Asian validation (N=840)	Seoul National University Hospital	Seoul	Korea	Chung JW	442
	Yonsei University College of Medicine	Seoul	Korea	Kim SU	187
	Faculty of Medicine, Prince of Songkla University	Songkhla	Thailand	Sripongpun	211

Table S2: Correlation coefficients (r) between these indicator variables with missing values.

Variables	AFP	WBC	PLT	INR	BUN	Cr
AFP	1	0.24	0.24	-0.01	-0.03	-0.02
WBC		1	0.98	-0.01	0.24	0.33
PLT			1	-0.01	0.28	0.39
INR				1	-0.01	-0.004
BUN					1	0.67
Cr						1

Abbreviations: AFP, alpha-fetoprotein; BUN, blood urea nitrogen; Cr, creatinine; INR, international normalized ratio; PLT, platelet; WBC, white blood cell.

Table S3. Correlation coefficient (r) between variables with missing values and other observable variables.

Variables	AFP	WBC	PLT	INR	BUN	Cr
AFP	NA	0.031	0.028	-0.007	-0.003	-0.006
WBC	0.037	NA	0.058	-0.005	-0.008	0.078
PLT	0.023	NA	NA	-0.001	-0.022	0.023
INR	0.013	0.000	-0.002	NA	0.033	0.032
BUN	-0.020	0.099	0.099	0.030	NA	0.005
Cr	-0.048	-0.038	-0.038	0.018	-0.003	NA
ALT	0.029	0.007	0.006	-0.018	0.087	0.048
AST	0.000	0.009	0.006	0.01	0.062	0.037
ALB	-0.035	-0.048	-0.043	0.03	-0.072	-0.056
TBIL	0.038	0.021	0.025	0.047	0.015	0.059
Gender	-0.016	-0.018	-0.001	-0.014	-0.010	0.005
Age	0.012	-0.030	-0.035	0.032	0.015	-0.014
Aetiology	-0.003	-0.025	-0.027	-0.015	0.008	0.001
Tumor size	-0.008	0.038	0.044	-0.004	-0.007	0.018
Tumor number	0.026	-0.012	-0.014	0	0.030	0.017
ECOG	NA	NA	NA	NA	NA	NA
Child-Pugh score	0.044	0.010	0.006	-0.018	0.002	0.047

Abbreviations: AFP, alpha-fetoprotein; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; cm, centimeter; Cr, creatinine; INR, international normalized ratio; NA, not available; PLT, platelet; TBIL, total bilirubin; WBC, white blood cell.

Table S4. Predictors for OS by Cox multivariable regression in each imputed cohort and pooled all cohorts. (SE, standard error; AFP, alpha-fetoprotein).

	Tumor size, per 1cm increase			Tumor number, refer to single			Log ₁₀ AFP, per 1 increase		
	beta coefficient	SE	p value	beta coefficient	SE	p value	beta coefficient	SE	p value
C1	0.103	0.0093	<0.001	0.097	0.0172	<0.001	0.150	0.0280	<0.001
C2	0.102	0.0093	<0.001	0.096	0.0173	<0.001	0.152	0.0279	<0.001
C3	0.102	0.0093	<0.001	0.098	0.0173	<0.001	0.148	0.0279	<0.001
C4	0.101	0.0093	<0.001	0.097	0.0173	<0.001	0.148	0.0279	<0.001
C5	0.101	0.0093	<0.001	0.096	0.0172	<0.001	0.151	0.0278	<0.001
Pooled	0.102	0.0093	<0.001	0.096	0.0173	<0.001	0.150	0.0279	<0.001

*Age, gender, aetiology, ALT, AST, ALBI score, BUN, Cr, and INR were not identified as prognostic factors of overall survival in C1-C5 and pooled cohort.

Table S5. Baseline demographics and clinical characteristics in Chinese DEB-TACE cohort.

Variables	DEB-TACE (n=170)
Sex	
male	149 (87.6%)
female	21 (12.4%)
Age, years	62 (53-69)
Aetiology	
HBV	159 (93.5%)
Others	11 (6.5%)
The largest tumor diameter, cm	4.6 (3.0-7.1)
≤ 3 cm	44 (25.9%)
>3, ≤ 7 cm	83 (48.8%)
>7, ≤ 10 cm	24 (14.1%)
>10 cm	19 (11.2%)
Tumor number	
1	86 (50.6%)
2	47 (27.6%)
≥ 3	37 (21.8%)
Current BCLC staging	
A	107 (62.9%)
B	63 (37.1%)
Child-Pugh score	
5	141 (82.9%)
6	26 (15.3%)
7	3 (1.8%)

ALBI grade	
1	106 (62.4%)
2	64 (37.6%)
AFP, ng/ml	30.8 (5.4-296.5)
ALT, U/L	30.5 (20-49)
AST, U/L	37 (26-54)
ALB, g/L	41.3 (37.7-43.9)
TBIL, $\mu\text{mol/L}$	14.4 (10.9-20.8)
INR	1.1 (1.0-1.1)
WBC, $\times 10^9/\text{L}$	5.3 (4.0-6.5)
PLT, $\times 10^9/\text{L}$	127 (88-185)
Cr $\mu\text{mol/L}$	74 (66-82)
Sessions of TACE	2 (2-3)
Follow-up time, months	30.6 (23.1-38.1)

Abbreviations: AFP, alpha-fetoprotein; ALB, albumin; ALBI, albumin-bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; cm, centimeter; Cr, creatinine; cTACE, conventional transarterial chemoembolization; DEB-TACE, drug-eluting beads transarterial chemoembolization; HBV, hepatic B virus; HCV, hepatic C virus; INR, international normalized ratio; PLT, platelet; TACE, transarterial chemoembolization; TBIL, total bilirubin; WBC, white blood cell.

Table S6. Comparison of the performance and discrimination among current available prognostic metrics in different subgroups of gender and age.

Prognostic metrics	Male				Female				Age≤60 years				Age>60 years			
	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC
6-and-12 model 2.0	0.673	0.011	8965.11	8978.76	0.675	0.027	1010.05	1018.20	0.672	0.013	5939.51	5952.09	0.673	0.015	3597.17	3608.49
6-and-12 model	0.666	0.010	8984.12	8993.22	0.649	0.026	1020.09	1025.53	0.661	0.013	5964.45	5972.84	0.668	0.015	3601.71	3609.26
Up to seven criteria	0.613	0.009	9039.55	9044.10	0.605	0.024	1031.01	1033.73	0.612	0.011	5999.38	6003.57	0.61	0.013	3634.03	3637.80
Four and seven criteria	0.616	0.010	9045.40	9049.95	0.579	0.026	1036.11	1038.83	0.605	0.012	6014.46	6018.65	0.619	0.014	3627.99	3631.76
Seven and eleven criteria	0.646	0.010	8997.04	9001.59	0.634	0.026	1022.64	1025.36	0.639	0.012	5979.62	5983.81	0.652	0.015	3598.68	3602.46
BCLC subclassification	0.586	0.088	9067.83	9072.38	0.592	0.023	1029.30	1032.01	0.599	0.01	6002.03	6006.23	0.567	0.013	3659.14	3662.92
HAP score	0.607	0.011	9074.50	9079.05	0.587	0.029	1037.57	1040.29	0.599	0.013	6025.34	6029.53	0.615	0.016	3646.61	3650.39
mHAP III score	0.656	0.011	9014.25	9018.80	0.640	0.027	1020.20	1022.92	0.661	0.013	5966.33	5970.52	0.64	0.016	3632.80	3636.58
mHAP II score	0.617	0.011	9056.38	9060.93	0.601	0.03	1033.62	1036.34	0.615	0.013	6008.0	6012.19	0.614	0.016	3643.96	3647.53
mHAP score	0.615	0.011	9062.24	9066.79	0.613	0.027	1029.63	1032.35	0.612	0.013	5979.62	5983.81	0.62	0.016	3641.45	3645.24
ALBI score	0.532	0.012	9134.41	9138.06	0.530	0.029	1041.25	1043.97	0.510	0.014	6060.15	6064.35	0.564	0.017	3673.15	3676.94

Abbreviations: AIC, Akaike Information Criterion; ALBI, albumin-bilirubin; BIC, Bayesian Information Criterion; BCLC, Barcelona

Clinic Liver Cancer; C-index, concordance index; HAP, Hepatoma arterial-embolization prognostic; SD, standard deviation;

Table S7. Comparison of the performance and discrimination among current available prognostic metrics in different subgroups of ALBI grade and aetiology.

Prognostic metrics	ALBI grade 1				ALBI grade 2				HBV				Other aetiology			
	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC	C-index	SD	AIC	BIC
6-and-12 model 2.0	0.676	0.015	4460.37	4472.18	0.674	0.013	4846.94	4859.04	0.669	0.011	8834.25	8847.82	0.694	0.024	1117.41	1125.75
6-and-12 model	0.656	0.015	4485.71	4493.58	0.672	0.013	4852.87	4860.94	0.658	0.011	8860.21	8869.29	0.692	0.023	1118.93	1128.49
Up to seven criteria	0.609	0.012	4510.96	4514.90	0.614	0.012	4894.0	4898.03	0.611	0.009	8910.85	8915.39	0.616	0.019	1134.97	1137.75
Four and seven criteria	0.616	0.013	4508.70	4512.64	0.605	0.013	4910.32	4914.36	0.603	0.010	8931.07	8935.61	0.654	0.02	1122.79	1129.57
Seven and eleven criteria	0.635	0.014	4496.31	4500.25	0.654	0.013	4855.08	4859.11	0.64	0.01	8873.87	8878.41	0.671	0.022	1119.20	1121.97
BCLC subclassification	0.580	0.011	4520.76	4524.68	0.589	0.011	4916.83	4920.86	0.588	0.008	8929.30	8933.84	0.578	0.019	1143.07	1145.85
HAP score	0.617	0.015	4515.66	4519.60	0.585	0.014	4937.9	4941.94	0.606	0.011	8942.74	8947.28	0.600	0.027	1143.90	1146.68
mHAP III score	0.660	0.015	4492.53	4496.47	0.671	0.013	4949.95	4853.99	0.648	0.011	8887.16	8891.70	0.683	0.024	1124.52	1127.30
mHAP II score	0.623	0.015	4510.48	4514.42	0.60	0.014	4921.58	4925.61	0.617	0.011	8922.03	8926.57	0.602	0.026	1142.06	1144.84
mHAP score	0.633	0.015	4500.13	4504.07	0.592	0.014	4930.04	4934.07	0.617	0.011	8922.83	8927.36	0.603	0.026	1143.50	1146.28
ALBI score	0.489	0.016	4555.00	4558.94	0.496	0.015	4958.10	4962.14	0.536	0.012	8996.49	9001.03	0.502	0.029	1153.66	1156.44

Abbreviations: AIC, Akaike Information Criterion; ALBI, albumin-bilirubin; BIC, Bayesian Information Criterion; BCLC, Barcelona

Clinic Liver Cancer; C-index, concordance index; HAP, Hepatoma arterial-embolization prognostic; HBV, Hepatic B virus; SD, standard deviation;

Table S8. Comparison of the performance and discrimination among current available prognostic metrics in Chinese DEB-TACE cohort.

Prognostic metrics	C-index	SD	AIC	BIC
6-and-12 model 2.0	0.639	0.033	664.1	666.4
6-and-12 model	0.607	0.037	669.6	674.1
Up to seven criteria	0.584	0.031	670.2	672.4
Four and seven criteria	0.606	0.029	667.3	669.6
Seven and eleven criteria	0.592	0.033	670.4	672.7
BCLC subclassification	0.583	0.031	659.8	662.1
HAP score	0.585	0.034	671.0	673.3
mHAP III score	0.632	0.033	664.4	666.6
mHAP II score	0.592	0.033	669.1	671.4
mHAP score	0.591	0.031	669.8	672.1
ALBI score	0.507	0.037	678.2	680.5

Abbreviations: AIC, Akaike Information Criterion; ALBI, albumin-bilirubin; BIC, Bayesian Information Criterion; BCLC, Barcelona Clinic Liver Cancer; C-index, concordance index; HAP, Hepatoma arterial-embolization prognostic; HBV, Hepatic B virus; SD, standard deviation;

Table S9. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in internal validation cohort.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	18.3% (7.10%-27.1%)	0.002	0.7% (-0.1%-1.7%)	0.092	11.5% (4.0%-18.5%)	0.004	1.3% (0.2%-2.7%)	0.004
Up to seven criteria	30.0% (16.1%-39.6%)	<0.001	4.0% (2.1%-6.6%)	<0.001	15.5% (4.9%-24.5%)	0.012	4.5% (2.0%-7.0%)	0.002
Four and seven criteria	22.5% (10.1%-35.3%)	0.002	3.4% (1.5%-5.7%)	<0.001	10.9% (-0.1%-21.1%)	0.052	3.4% (0.8%-5.7%)	0.022
Seven and eleven criteria	20.5% (9.80%-32.2%)	0.002	2.5% (1.0%-4.6%)	0.002	16.6% (6.3%-25.3%)	0.004	3.4% (1.3%-5.5%)	0.002
BCLC subclassification	23.5% (6.70%-32.8%)	0.012	2.6% (1.2%-3.9%)	<0.001	15.8% (4.2%-28.4%)	0.002	2.9% (0.8%-5.2%)	0.016
HAP score	13.1% (0.30%-25.2%)	0.044	2.9% (1.0%-5.5%)	<0.001	8.7% (-1.3%-20.0%)	0.098	4.5% (1.6%-7.5%)	0.002
mHAP III score	22.0% (10.4%-36.0%)	<0.001	1.9% (0.5%-4.2%)	0.004	16.5% (5.9%-24.8%)	0.008	1.6% (0.3%-3.6%)	0.02
mHAP II score	14.6% (-0.1%-29.3%)	0.056	2.5% (0.3%-5.1%)	0.026	6.9% (-6.2%-18.1%)	0.304	3.0% (0.0%-6%)	0.05
mHAP score	9.30% (-2.0%-21%)	0.100	2.1% (0.2%-4.6%)	0.026	10.2% (0.0%-19%)	0.044	3.7% (1%-6.3%)	0.002
ALBI score	28.7% (13.6%-36.3%)	<0.001	5.3% (3.0%-8.4%)	<0.001	26.9% (15.2%-33.8%)	<0.001	8.7% (5.5%-12.2%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S10. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in European validation cohort.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	23.1% (15.1% - 32.0%)	<0.001	1.6% (0.7% - 2.5%)	<0.001	17.9% (9.9% - 24.1%)	<0.001	1.6% (0.5% - 2.6%)	<0.001
Up to seven criteria	20.7% (8.80% - 31.0%)	<0.001	3.5% (2.0% - 5.5%)	<0.001	18% (6.4% - 23.8%)	<0.001	3.0% (1.1% - 4.5%)	<0.001
Four and seven criteria	21.2% (11.7% - 27.8%)	<0.001	3.8% (2.0% - 5.7%)	<0.001	14.9% (7.1% - 23.7%)	<0.001	3.5% (1.45 - 5.5%)	0.004
Seven and eleven criteria	21.2% (9.5% - 31%)	<0.001	2.9% (1.4% - 4.4%)	<0.001	15.6% (5.1% - 23.2%)	0.004	2.3% (0.8% - 3.7%)	<0.001
BCLC subclassification	21.4% (9.6% - 31.3%)	<0.001	3.3% (1.5% - 5.5%)	<0.001	18.5% (7.4% - 25.3%)	0.004	2.9% (1.1% - 4.7%)	0.004
HAP score	2.4% (-9.6% - 12.7%)	0.758	1.5% (-0.6% - 3.7%)	0.172	-1.5% (-12.7% - 8.9%)	0.802	0.9% (-1.9% - 3.1%)	0.527
mHAP III score	16.7% (3.2% - 28.3%)	0.012	1.3% (0.1% - 2.5%)	0.044	6.3% (-8.4% - 15.2%)	0.427	0.7% (-0.6% - 2.0%)	0.315
mHAP II score	9.3% (-3.9% - 20.5%)	0.168	2.3% (0.3% - 4.6%)	0.028	1.4% (-11.7% - 8.4%)	0.958	0.4% (-2.6% - 2.8%)	0.798
mHAP score	5.1% (-7.6% - 16%)	0.431	1.6% (-0.4% - 3.5%)	0.120	6.7% (-5.6% - 15.1%)	0.319	0.2% (-0.2% - 4.0%)	0.064
ALBI score	17% (6.5% - 26.2%)	<0.001	3.3% (1.5% - 5.5%)	<0.001	7.1% (0.3% - 18.1%)	0.06	3.7% (0.8%-6.2%)	0.028

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S11. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in Asian validation cohort.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	19.3% (0.61% - 29.0%)	<0.001	1.3% (-2.0% - 2.7%)	0.088	17.1% (8.8% - 24.5%)	<0.001	2.2% (0.7% - 4.1%)	0.004
Up to seven criteria	22.9% (9.9% - 34.9%)	<0.001	4.8% (1.7% - 8.4%)	<0.001	18.5% (4.3% - 26.8%)	0.020	4.8% (1.9% - 7.7%)	0.008
Four and seven criteria	24.7% (9.5% - 35.3%)	0.004	4.6% (1.7% - 8.7%)	0.004	17.7% (4.9% - 26.6%)	0.004	4.1% (0.8% - 7.7%)	0.024
Seven and eleven criteria	30.2% (17.4% - 40.5%)	<0.001	3.6% (1.2% - 6.4%)	0.008	28.4% (19.8% - 35.1%)	<0.001	6.4% (4.2% - 9.1%)	<0.001
BCLC subclassification	20.9% (3.3% - 32.8%)	0.02	3.1% (-0.7% - 10%)	0.100	18.2% (2.7% - 26.8%)	0.016	3.1% (-0.4% - 6.8%)	0.072
HAP score	12.6% (-3.8% - 27.2%)	0.136	4.2% (0.8% - 8%)	0.016	10.2% (-2.0% - 23%)	0.100	5.9% (2% - 9.4%)	<0.001
mHAP III score	24.4% (8.3% - 35.4%)	0.020	2.8% (0.8% - 6%)	0.008	13.4% (2.2% - 21.6%)	0.032	2.4% (0.05% - 4.7%)	0.012
mHAP II score	11.3% (-3.4% - 24.5%)	0.144	4.5% (0.8% - 8.4%)	0.008	6.8% (-4.1% - 19.0%)	0.208	5% (1% - 8.1%)	0.012
mHAP score	7.7% (-10.0% - 24.6%)	0.383	2.3% (-0.8% - 5.9%)	0.128	5.7% (-6.3% - 19.3%)	0.319	3.1% (-0.3% - 6.6%)	0.068
ALBI score	16.9% (2.5% - 31.4%)	0.02	6.2% (2.0% - 11.1%)	0.004	16.2% (3% -26.8%)	0.020	7.8% (3.3% - 12.6%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S12. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in patients with age > 60years.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	8.7% (0.2% - 20.9%)	0.048	0.9% (0.1% - 2.7%)	0.028	5.5% (-3.6% - 15.6%)	0.236	-0.5% (-0.2% - 2.1%)	0.251
Up to seven criteria	9.3% (-11.4% - 24%)	0.439	3.5% (0.2% - 7.4%)	0.024	10.4% (-7.9% - 27.0%)	0.323	4.2% (0.0% - 8.1%)	0.048
Four and seven criteria	8.1% (-7.5% - 27.4%)	0.303	2.9% (0.2% - 6.5%)	0.028	14.9% (-5.7% - 30.3%)	0.132	3.6% (0.5% - 7.5%)	0.046
Seven and eleven criteria	-2.2% (-17.4% - 18.6%)	0.886	0.4% (-1.8% - 3.5%)	0.651	-10.9% (-24.1% - 9.1%)	0.315	-0.9% (-4.2% - 2.3%)	0.659
BCLC subclassification	27.9% (17.7% - 42.9%)	0.004	6.2% (3.5% - 10.3%)	<0.001	21.2% (7.1% - 33.2%)	0.004	8.4% (4.3% - 12.5%)	0.004
HAP score	15.9% (0.5% - 31.9%)	0.046	3.6% (0.6% - 7.8%)	0.012	19.1% (3.7% - 31.9%)	0.008	6.8% (2.0% - 10.9%)	0.004
mHAP III score	20.9% (10.2% - 33.3%)	<0.001	3.1% (1.0% - 6.4%)	<0.001	22.6% (12% - 32.8%)	<0.001	5% (2.6% - 7.8%)	<0.001
mHAP II score	22.2% (5.4% - 34.3%)	0.008	3.6% (0.4% - 7.3%)	0.032	15.4% (4.2% - 29.7%)	0.016	6.3% (1.9% - 10.7%)	0.004
mHAP score	19.9% (2.6% - 33.8%)	0.024	3.3% (0.5% - 7.0%)	0.024	15% (-2.5% - 29.1%)	0.096	5.6% (1.4% - 9.8%)	0.008
ALBI score	32.8% (19.5% - 44.2%)	0.004	6.9% (3.7% - 11.3%)	0.004	21.3% (6.5% - 31.9%)	0.004	9.3% (4.9% - 14.2%)	0.004

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S13. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in patients with age≤ 60years.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	24.9% (5.8% - 37.7%)	0.012	4.1% (0.7% - 8.7%)	0.008	11.6% (4.1% - 33.7%)	0.012	4.1% (0.8% - 10.0%)	0.008
Up to seven criteria	29.5% (5.9% - 47.4%)	0.012	7.3% (2.5% - 14.3%)	<0.001	17.6% (3.1% - 41.3%)	0.016	8.4% (3.4% - 15.3%)	<0.001
Four and seven criteria	35.1% (16.6% - 51.0%)	<0.001	9.7% (4.7% - 16.8%)	<0.001	20.1% (0.2% - 46.9%)	0.046	9.6% (3.4% - 18.1%)	0.008
Seven and eleven criteria	24.8% (0.8% - 41.8%)	0.046	4.3% (0.5% - 10.1%)	0.048	17.6% (-4.4% - 35.6%)	0.144	4.6% (0.0% - 11.6%)	0.050
BCLC subclassification	25.1% (1.4% - 45.7%)	0.032	7.5% (1.4% - 14.5%)	0.012	4.2% (-10.7% - 32.8%)	0.383	7.6% (0.1% - 16.0%)	0.024
HAP score	34.1% (14.0% - 49.4%)	0.004	9.4% (3.9% - 17.2%)	<0.001	28.5% (0.2% - 42.6%)	0.048	10.4% (3.4% - 18.1%)	0.008
mHAP III score	29.8% (9.0% - 49.0%)	0.004	5.1% (2.2% - 10.3%)	<0.001	27.9% (0.4% - 43.2%)	0.048	5.1% (1.7% - 10.6%)	0.004
mHAP II score	27.1% (7.6% - 46.2%)	0.024	7.9% (2.1% - 15.3%)	0.020	27.1% (7.6% - 46.2%)	0.024	7.9% (2.1% - 15.3%)	0.020
mHAP score	22.3% (1.4% - 41.0%)	0.048	6.5% (0.3% - 13.7%)	0.036	17.4% (-6.3% - 38.3%)	0.116	6.3% (0.5% - 13.7%)	0.048
ALBI score	33.9% (15.7% - 50.4%)	<0.001	11.1% (5% - 19.3%)	<0.001	29.1% (0.6% - 47.2%)	0.044	12.2% (3.9% - 21.1%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S14. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in male patients.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	13.6% (2.8% - 24.5%)	0.004	1.4% (0.4%-2.4%)	0.012	9.3% (0.1% - 18.4%)	0.044	1.1% (-0.2% - 2.3%)	0.092
Up to seven criteria	18.1% (4.5% - 25.3%)	0.004	4.3% (2.1% - 6.4%)	<0.001	13.5% (2.9% - 22.2%)	0.016	3.9% (1.2% - 6.4%)	0.008
Four and seven criteria	16.8% (3.7% - 28.3%)	0.016	3.9% (1.8% - 5.9%)	<0.001	17.7% (6.7% - 27.4%)	<0.001	4.5% (1.9% - 7.0%)	<0.001
Seven and eleven criteria	11.4% (1.2% - 21.1%)	0.046	2.3% (0.6% - 4.0%)	0.004	3.9% (0.6% - 15.1%)	0.047	1.2% (0.1% -3.1%)	0.024
BCLC subclassification	28.4% (18% - 36.1%)	<0.001	5.9% (3.8% - 8.4%)	<0.001	16.3% (6.8% - 24.1%)	<0.001	5.4% (2.8% - 8.3%)	<0.001
HAP score	25.2% (14.7% - 34.0%)	<0.001	5.6% (3.5% - 7.9%)	<0.001	23.8% (16.7% - 30.7%)	<0.001	7.7% (4.8% - 10.2%)	<0.001
mHAP III score	23.6% (13.1% - 34.0%)	<0.001	2.2% (0.6% - 3.6%)	0.012	16.6% (6.4% - 24.5%)	<0.001	3.4% (1.5% - 5.0%)	<0.001
mHAP II score	26.2% (14.3% - 33.9%)	<0.001	5.0% (2.9% - 7.3%)	<0.001	24.0% (14.6% - 29.7%)	<0.001	6.7% (3.8% - 9.1%)	<0.001
mHAP score	21.2% (13.5% - 32.4%)	<0.001	4.5% (2.5% - 6.5%)	<0.001	21.2% (10.8% - 29.2%)	<0.001	1.2% (0.1% - 3.1%)	0.024
ALBI score	38.0% (28.6% - 44.4%)	<0.001	8.1% (5.5% - 10.7%)	<0.001	25.7% (18.5% - 32.5%)	<0.001	10.2% (6.6% - 13.1%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S15. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in female patients.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	20.3% (0.6% - 46.1%)	0.048	3.9% (0.6% - 7%)	0.012	12.6% (4.7% - 31.8%)	0.002	3.9% (0.2% - 7.5%)	0.032
Up to seven criteria	27.3% (3.5% - 47.2%)	0.012	7.1% (2.3% - 12.9%)	<0.001	21.4% (4.6% - 42.6%)	0.020	8.2% (2.9% - 13.4%)	<0.001
Four and seven criteria	33.3% (17.5% - 50.7%)	<0.001	9.5% (4.3% - 15.7%)	<0.001	32.9% (5.7% - 47.8%)	0.020	9.4% (3.1% - 15.7%)	<0.001
Seven and eleven criteria	24.6% (1.9% - 45.4%)	0.036	4.2% (-0.4% - 8.6%)	0.080	21.7% (-3.7% - 39.1%)	0.100	4.4% (-0.1% - 8.7%)	0.068
BCLC subclassification	24.7% (1.5% - 48.0%)	0.036	7.4% (1.1% - 14.0%)	0.012	15.7% (-9.5% - 35.4%)	0.271	7.5% (0.3% - 14.1%)	0.044
HAP score	32.5% (11.6% - 47.2%)	0.004	9.2% (3.5% - 16.3%)	<0.001	31.7% (1.4% - 45.3%)	0.036	10.2% (2.2% - 17.6%)	0.012
mHAP III score	25.1% (-1.5% - 46.2%)	0.076	5.0% (1.1% - 9.7%)	0.008	26.5% (1.1% - 40.2%)	0.032	4.9% (0.1% - 9.8%)	0.046
mHAP II score	20.1% (5.8% - 41.1%)	0.016	7.7% (1.3% - 14.9%)	0.020	17.1% (3.4% - 39.7%)	0.014	8.5% (0.5% - 16.0%)	0.032
mHAP score	20.4% (3.9% - 38.9%)	0.048	6.4% (0.5% - 12.5%)	0.012	19.9% (6.0% - 36.1%)	0.012	6.1% (2.4% - 13.3%)	0.028
ALBI score	32.3% (16.1% - 50.3%)	<0.001	11% (4.8% - 18.3%)	<0.001	32.3% (2.1% - 47.9%)	0.005	12% (2.9% - 20.6%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S16. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in patients with ALBI grade 1.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	18.3% (6.7% - 31.4%)	<0.001	2.3% (1.1% - 4.0%)	<0.001	14.2% (3.2% - 23.6%)	<0.001	2.2% (0.6% - 3.7%)	0.008
Up to seven criteria	21.9% (2.0% - 31.3%)	0.040	4.9% (1.8% - 8.2%)	<0.001	15.2% (0.4% - 26.0%)	0.048	4.4% (1.4% - 8.0%)	0.012
Four and seven criteria	19.2% (2.7% - 32.7%)	0.024	4.3% (1.4% - 7.7%)	<0.001	15.7% (2.0% - 27.5%)	0.032	4.1% (0.9% - 7.5%)	0.008
Seven and eleven criteria	20.2% (5.4% - 30.9%)	0.020	3.3% (0.9% - 6.1%)	0.008	11.4% (-2.1% - 23.2%)	0.092	2.8% (0.2% - 5.5%)	0.036
BCLC subclassification	25.8% (10.3% - 37.1%)	<0.001	6.1% (2.8% - 9.8%)	<0.001	18.7% (4.0% - 28.4%)	0.004	5.8% (2.2% - 9.3%)	0.004
HAP score	16.6% (0.5% - 33.4%)	0.048	3.9% (1.0% - 7.3%)	0.008	24.0% (8.5% - 34.7%)	<0.001	6.8% (3.5% - 10.2%)	<0.001
mHAP III score	31.5% (20.5% - 41.8%)	<0.001	3.2% (1.3% - 5.7%)	0.004	23.4% (14.2% - 32.2%)	<0.001	3.6% (1.7% - 5.8%)	<0.001
mHAP II score	16.8% (3.8% - 28.9%)	0.024	3.8% (1.0% - 7.0%)	0.012	23.4% (14.2% - 32.2%)	<0.001	6.0% (2.7% - 9.5%)	<0.001
mHAP score	16.3% (0.5% - 29.2%)	0.048	3.1% (0.3% - 5.9%)	0.016	17.5% (2.0% - 30.3%)	0.024	4.3% (1.3% - 7.3%)	0.008
ALBI score	38.3% (27.5% - 47.0%)	<0.001	8.1% (4.4% - 12.1%)	<0.001	24.4% (15.6% - 33.7%)	<0.001	9.7% (5.5% - 13.9%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S17. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in patients with ALBI grade 2.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	13.1% (-0.1% - 25.9%)	0.052	1.0% (-0.5% - 2.6%)	0.156	4.1% (-6.7% - 17.8%)	0.431	0.5% (-1.3% - 2.3%)	0.635
Up to seven criteria	16.3% (2.3% - 29.8%)	0.024	4.3% (1.8% - 7.2%)	<0.001	14.8% (0.0% - 25.9%)	0.050	4.3% (0.6% - 7.3%)	0.012
Four and seven criteria	26.6% (8.4% - 36.2%)	0.004	5.0% (2.3% - 29.8%)	0.024	23.6% (7.3% - 32.8%)	<0.001	6.1% (2.5% - 9.3%)	<0.001
Seven and eleven criteria	5.6% (-6.5% - 18.6%)	0.311	1.3% (-0.6% - 3.7%)	0.168	2.3% (-10% - 13.9%)	0.731	0.0% (-2.7% - 2.3%)	1.034
BCLC subclassification	29.1% (16.6% - 40.5%)	<0.001	6.3% (3.5% - 9.6%)	<0.001	14.0% (1.4% - 24.3%)	0.028	5.8% (2.0% - 9.0%)	0.004
HAP score	35.4% (22.4% - 43.0%)	<0.001	6.8% (4.2% - 10.1%)	<0.001	25.6% (17.3% - 34.3%)	<0.001	9.1% (5.7% - 12.6%)	<0.001
mHAP III score	10.8% (-1.7% - 26.3%)	0.124	0.1% (-1.7% - 2.0%)	0.850	10.2% (-1.0% - 23.1%)	0.080	0.9% (-0.6% - 2.7%)	0.259
mHAP II score	27.4% (14.0% - 37.9%)	<0.001	5.7% (3.1% - 9.0%)	<0.001	22.2% (9.3% - 30.3%)	<0.001	7.0% (3.5% - 10.4%)	<0.001
mHAP score	27.9% (16.7% - 39.5%)	<0.001	6.0% (3.4% - 8.8%)	<0.001	25.0% (11.3% - 34.2%)	<0.001	8.1% (4.9% - 11.2%)	<0.001
ALBI score	37.9% (29.0% - 46.2%)	<0.001	8.8% (5.7% - 12.6%)	<0.001	28.9% (20.7% - 38.5%)	<0.001	11.5% (7.3% - 15.7%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S18. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in patients with HBV.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	16.7% (4.7% - 27.7%)	0.004	1.6% (0.5% - 2.6%)	0.004	13.2% (4.4% - 23.1%)	0.004	1.6% (0.3% - 2.9%)	0.004
Up to seven criteria	17.7% (6.3% - 26.9%)	0.004	4.4% (2.3% - 6.9%)	<0.001	14.5% (1.7% - 22.6%)	0.016	4.2% (1.3% - 6.6%)	0.008
Four and seven criteria	23.8% (11.1% - 33.1%)	<0.001	4.7% (2.7% - 7.5%)	<0.001	23.2% (12.1% - 31.1%)	<0.001	5.6% (3.1% - 8.1%)	<0.001
Seven and eleven criteria	16.3% (3.9% - 23.4%)	0.024	2.6% (0.9% - 4.4%)	0.004	7.0% (-5.0% - 17.6%)	0.359	1.7% (-0.3% - 3.5%)	0.124
BCLC subclassification	26.0% (14.8% - 34.1%)	<0.001	5.7% (3.2% - 8.4%)	<0.001	14.0% (3.3% - 23.5%)	0.004	5.2% (1.9% - 7.8%)	<0.001
HAP score	24.7% (13.3% - 31.3%)	<0.001	5.0% (2.7% - 7.3%)	<0.001	24.5% (13.1% - 32.9%)	<0.001	6.8% (4.0% - 9.3%)	<0.001
mHAP III score	23.0% (13.4% - 34.0%)	<0.001	2.3% (0.9% - 3.8%)	0.004	18.0% (8.5% - 25.9%)	<0.001	3.7% (2.2% - 5.4%)	<0.001
mHAP II score	21.8% (7.5% - 31.5%)	<0.001	4.2% (2.1% - 6.8%)	<0.001	19.0% (9.2% - 27.0%)	<0.001	5.2% (2.4% - 8.0%)	<0.001
mHAP score	16.8% (8.5% - 28.8%)	<0.001	4.1% (2.1% - 6.4%)	0.004	18.2% (5.3% - 27.6%)	0.008	5.1% (2.3% - 7.7%)	0.004
ALBI score	34.7% (26.3% - 41.5%)	<0.001	8.0% (5.4% - 10.9%)	<0.001	26.4% (20.7% - 33.9%)	<0.001	9.8% (6.5% - 13.1%)	<0.001

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S19. Comparison of NRI and IDI between 6-and-12 model 2.0 and other current available prognostic metrics (standard model) at 1-year and 3-year timepoint in Chinese DEB-TACE cohort.

	1-year survival timepoint				3-year survival timepoint			
	NRI (95% CI)	p value	IDI (95% CI)	p value	NRI (95% CI)	p value	IDI (95% CI)	p value
6-and-12 model	9.90% (-19.1%-37.4%)	0.527	0.0% (-2.2%-1.4%)	0.951	13.7% (-8.50%-37.4%)	0.242	2.5% (-0.6%-6.1%)	0.148
Up to seven criteria	31.5% (-12.5%-48.8%)	0.206	1.2% (-0.3%-4.5%)	0.126	6.4% (-23.8%-31.4%)	0.703	3.5% (-3.1%-9.9%)	0.298
Four and seven criteria	1.5% (-31.8%-34.9%)	0.683	0.1% (-2.8%-2.6%)	0.929	-0.8% (-28.6%-29.5%)	0.987	1.7% (-5.4%-8.1%)	0.613
Seven and eleven criteria	9.9% (-22%-38.5%)	0.625	0.6% (-1.3%-3%)	0.523	12% (-16.8%-38.6%)	0.322	4.1% (-0.7%-9.5%)	0.098
BCLC subclassification	29.3% (-12.4%-50%)	0.18	1.3% (-0.3%-4.4%)	0.098	3.5% (-24.7%-28.6%)	0.755	2.8% (-3.8%-8.9%)	0.354
HAP score	-1.3% (-26.7%-35.7%)	1.137	1.1% (-1.4%-4.8%)	0.408	6.5% (-25.5%-33.4%)	0.713	3.1% (-4.2%-11.2%)	0.478
mHAP III score	5.9% (-29.2%-33.2%)	0.799	0.5% (-1.3%-2.7%)	0.505	5.1% (-24.3%-38.2%)	0.821	0.3% (-3.9%-4.7%)	0.901
mHAP II score	17.5% (-19.5%-44%)	0.354	1.8% (-0.2%-0.54%)	0.078	-5.8% (-30.4%-34%)	0.877	1.4% (-6.4%-9.5%)	0.727
mHAP score	29.3% (-12%-46.8%)	0.200	1.5% (-0.1%-5%)	0.076	4.8% (-19.9%-33.2%)	0.659	2.8% (-4.4%-9.9%)	0.440
ALBI score	20.2% (-10.6%-42.7%)	0.244	1.7% (-0.7%-5.8%)	0.210	28.5% (-10.4%-49.3%)	0.140	7.8% (-0.8%-16.9%)	0.09

Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; HAP, Hepatoma arterial-embolization prognostic; IDI, integrated discrimination improvement; NRI, net reclassification improvement.

Table S20. Subgroup analyses of OS according to the current risk stratification and its' hazard ratio by COX multivariable analysis.

Subgroups	Low-risk strata	Intermediate-risk strata	High-risk strata	p value	HR, 95% CI	p value	Adjusted variables
Age≤60 years	45.0 (40.2-49.8) months	30.9 (25.5-36.3) months	15.1 (11.9-18.3) months	<0.001	1.78 (1.58-2.01)	<0.001	WBC, AST
Age>60 years	46.8 (37.8-55.8) months	28.6 (23.4-33.9) months	16.1 (14.5-17.7) months	<0.001	2.02 (1.73-2.36)	<0.001	ALT, AST, ALB, TBIL
Male	45.0 (37.3-52.3) months	30.1 (26.0-34.2) months	15.8 (13.7-17.9) months	<0.001	1.91 (1.72-2.12)	<0.001	WBC, PLT, ALT, AST, TBIL
Female	46.3 (40.4-52.2) months	29.4 (17.2-41.6) months	13.6 (9.30-17.9) months	<0.001	1.92 (1.49-2.47)	<0.001	None
ALBI grade 1	48.9 (40.7-57.1) months	30.9 (25.0-36.8) months	17.5 (12.9-22.1) months	<0.001	1.84 (1.61-2.11)	<0.001	Age, WBC, Cr
ALBI grade 2	42.6 (36.4-48.8) months	28.4 (23.2-33.6) months	14.8 (13.0-16.6) months	<0.001	1.90 (1.65-2.18)	<0.001	PLT, AST
HBV	44.4 (39.8-49.0) months	30.8 (27.1-34.5) months	15.5 (13.4-17.6) months	<0.001	1.83 (1.66-2.03)	<0.001	WBC, AST, ALB
Other etiologies	56.0 (NE-NE) months	26.6 (19.5-33.7) months	14.9 (9.40-20.4) months	<0.001	2.02 (1.59-2.57)	<0.001	TBIL

Abbreviations: ALBI, albumin-bilirubin; CI, confidence interval; HBV, hepatic B virus; HR, hazard ratio; NE, not estimated.

Table S21. Subgroup analyses of overall survival according to the current risk stratification in patients with BCLC-A and BCLC-B

HCC among these four cohorts.

Datasets	BCLC stage	Low-risk strata	Intermediate-risk strata	High-risk strata	p value
Training	A	44.3 (40.0-50.1) months	31.2 (28.2-38.2) months	17.3 (13.2-24.8) months	<0.001
	B	48.0 (39.6 - NR) months	21.6 (18.2-25.4) months	13.8 (12.2-16.0) months	<0.001
Internal validation	A	51.1 (43.2-57.5) months	32.0 (27.7-37.4) months	17.6 (9.90-33.3) months	<0.001
	B	38.3 (35.5-59.6) months	30.4 (28.9-34.4) months	21.0 (17.2-25.5) months	<0.001
European validation	A	34.5 (31.5-37.6) months	23.3 (18.2-32.9) months	14.8 (12.4-32.7) months	<0.001
	B	26.1 (24.2-30.8) months	19.2 (17.2-22.3) months	13.6 (10.2-17.8) months	<0.001
Asian validation	A	96.3 (81.7-108) months	33.9 (21.7 - NR) months	19.5 (7.87 - NR) months	<0.001
	B	55.4 (47.3-91.5) months	34.7 (27.0-43.7) months	20.7 (13.6-26.7) months	<0.001

Abbreviations: BCLC, Barcelona Clinic Liver Cancer; NR, not reached.

Table S22. Summarization of the pivotal randomized controlled trials related to TACE.

Publication (year)	Trial	Country	Treatment	Primary endpoint	Outcomes	P
Okusaka et al[7]. 2009	NA	Japan	TAI (n = 82) cTACE (n = 79)	OS	22.3 21.2	0.383
Kudo et al[8]. 2011	POST-TACE	Japan, Korea	cTACE (responders) plus sorafenib (n = 229) cTACE plus placebo (n = 229)	TTP	5.4 3.7	0.252
Yu et al[9]. 2014	NA	China	TEA (n = 49) cTACE (n = 49)	OS	24.3 20.1	0.513
Golfieri et al[10]. 2014	PRECISION ITALIA	Italy	DEB- TACE (n = 89) cTACE (n = 88)	OS (2 years)	56.80% 55.40%	0.949
Kudo et al[11]. 2014	BRISK- TA	Global	cTACE or DEB- TACE plus brivanib (n = 249) cTACE plus placebo (n = 253)	OS	26.4 26.1	0.53
Lencioni et al[12]. 2016	SPACE	Global	DEB- TACE plus sorafenib (n = 154) DEB- TACE plus placebo (n = 153)	TTP	5.6 5.5	0.072
Meyer et al[13]. 2017	TACE-2	UK	DEB- TACE plus sorafenib (n = 157) DEB- TACE plus placebo (n = 156)	PFS	7.8 7.7	0.85
Kudo et al[14]. 2018	ORIENTAL	Japan, Korea, Taiwan	cTACE plus orantinib (n = 445) cTACE plus placebo (n = 444)	OS	31.1 32.3	0.435
Ikeda et al[15]. 2018	NA	Japan	cTACE with miriplatin (n = 129) cTACE with epirubicin (n = 128)	OS	36.5 37.1	0.946
Kudo et al[16]. 2022	TACTICS	Japan	cTACE plus sorafenib (n = 80) cTACE (n = 76)	OS	36.2 30.8	0.40

Abbreviations: cTACE, conventional transarterial chemoembolization; DEB-TACE, drug-eluting beads transarterial chemoembolization; OS, overall survival; PFS, progression-free survival; TAI, transarterial infusion; TEA, transarterial ethanol ablation; TTP, time to progression;

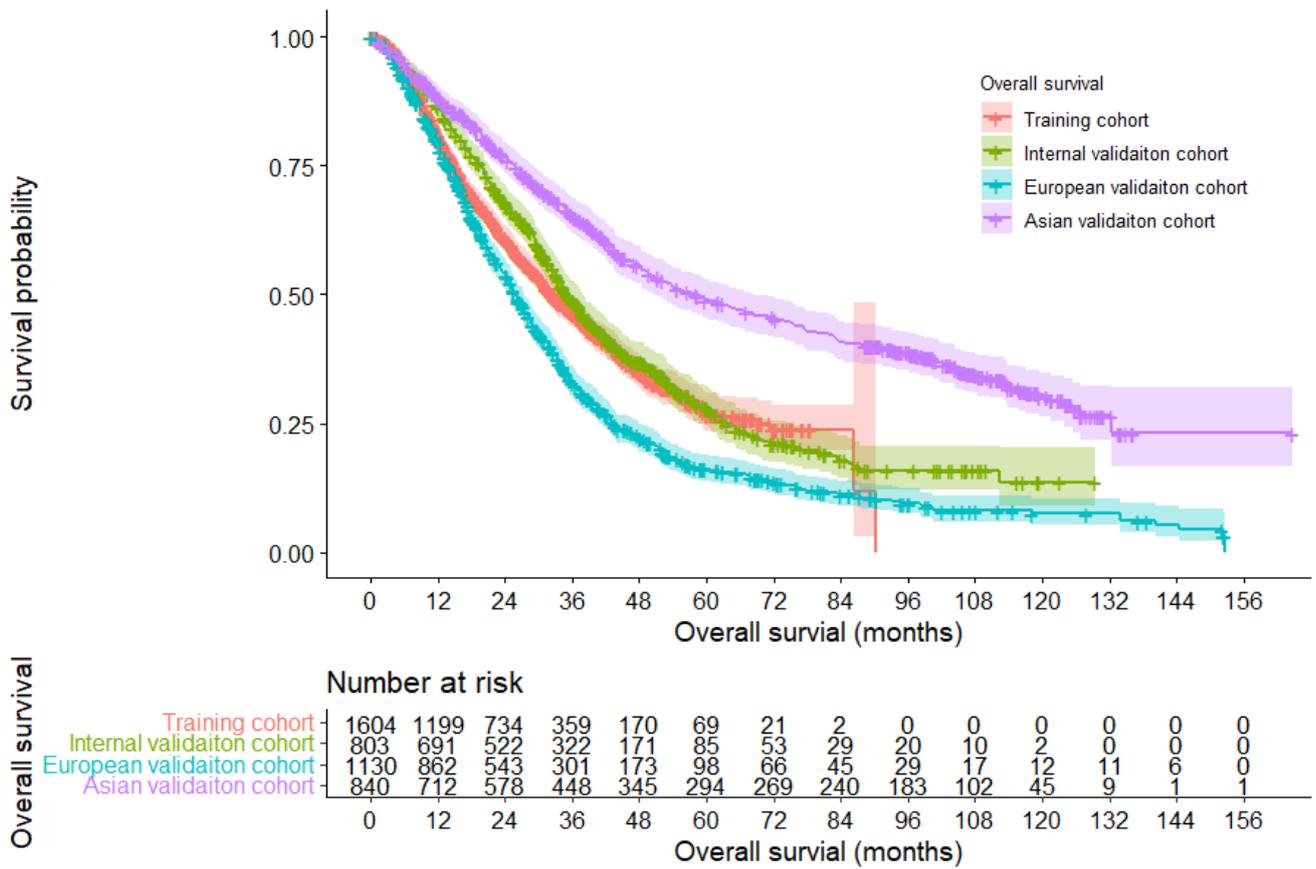


Fig. S2. Overall survival analysis by Kaplan-Meier method in training, internal, European and Asian validation cohorts. (median overall survival time was 32.9 (95% CI, 30.4–35.4) in the training cohort, 35.1 (95% CI, 32.9–37.3) in the internal validation cohort, 24.9 (95% CI, 22.0–27.9) in the European validation cohort, and 57.9 (95% CI, 48.7–67.1) months in the Asian validation cohort, $p < 0.001$ for overall comparison by log-rank test)

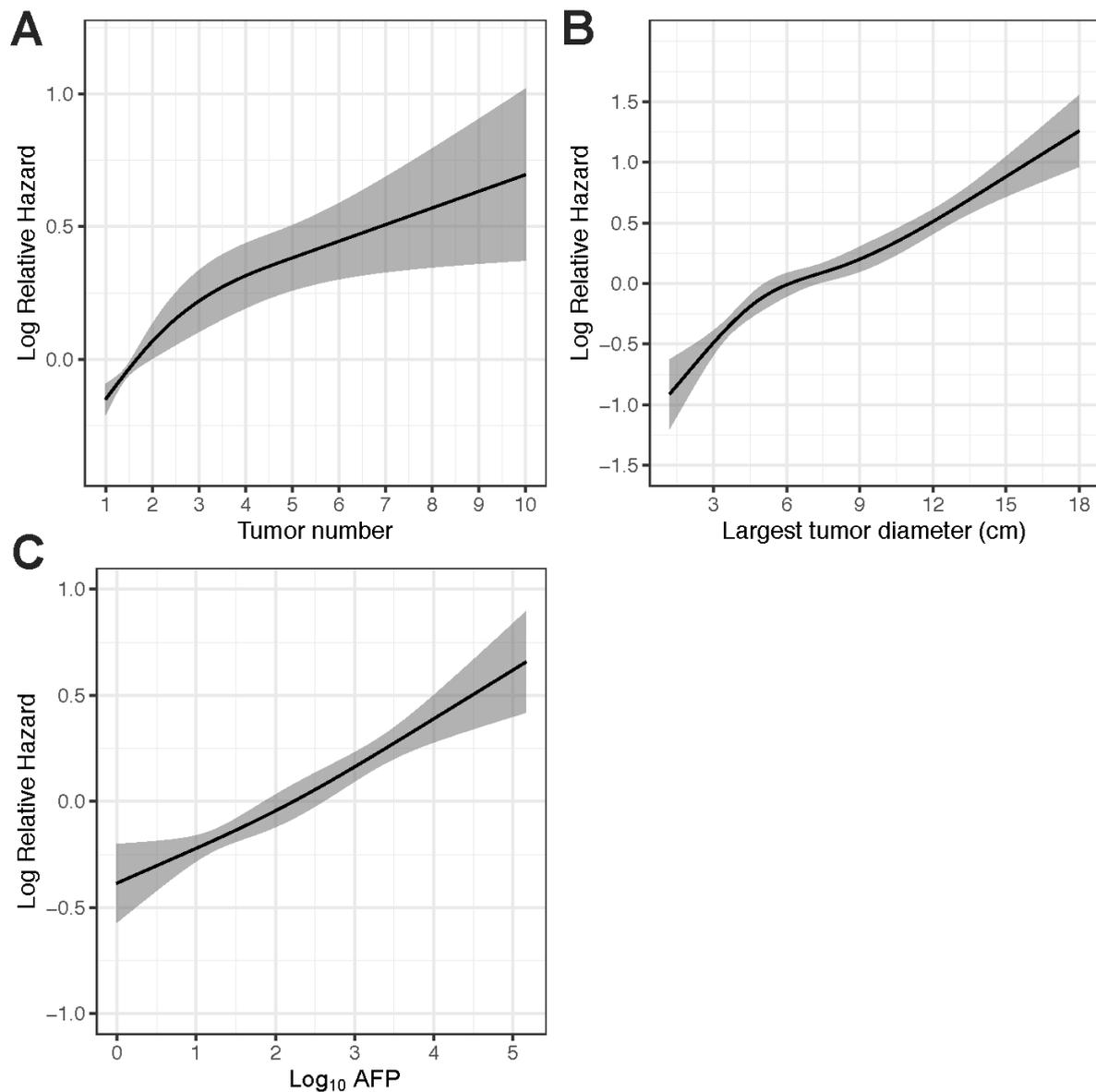


Fig S3. Relation between tumor number, largest tumor diameter, log₁₀AFP and relative hazard. (A, Restricted cubic spline of tumor number in training cohort (non-linear $p = 0.05$); B, Restricted cubic spline of largest tumor diameter in training cohort (non-linear $p = 0.11$); C, Restricted cubic spline of log₁₀AFP in training cohort (non-linear $p = 0.40$).

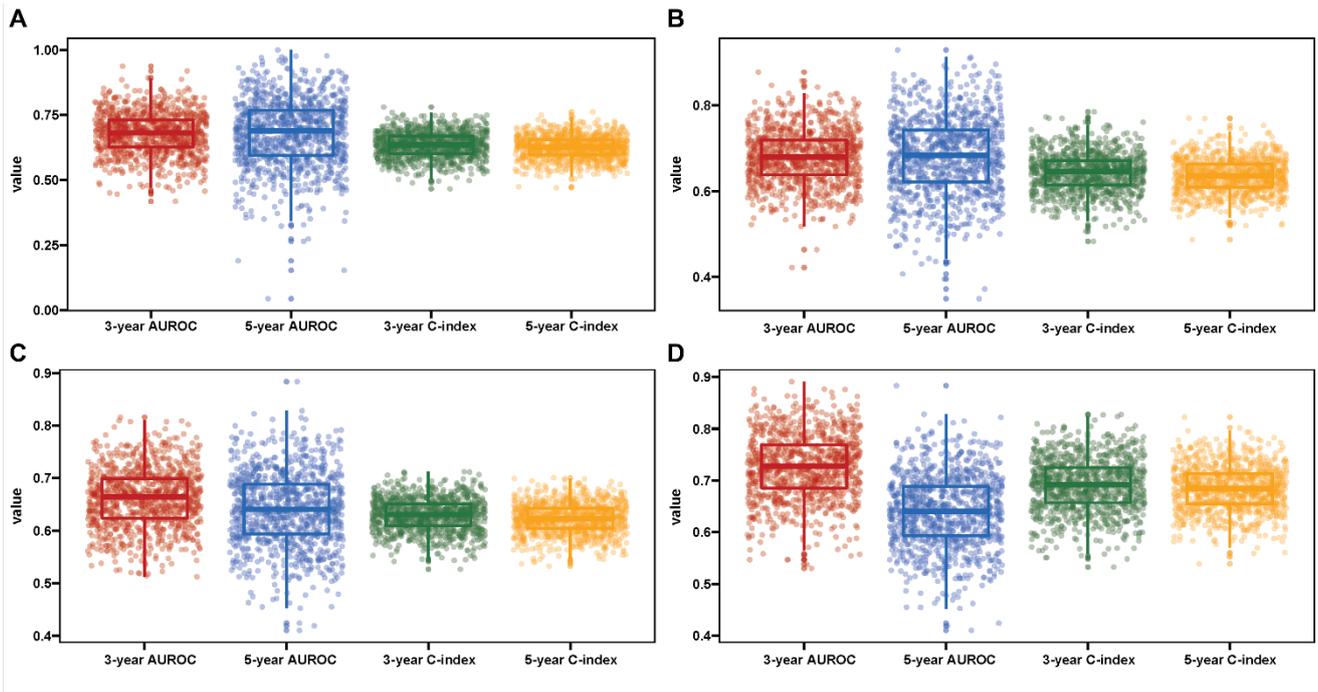


Fig. S4. Discrimination analyses of 6-and-12 model 2.0 using the concordance index (C-index) and the area under the receiver operating characteristics curve (AUROC) with a 10-fold-100-times cross validation approach in ideal TACE candidates. (Each scatter represents each cross-validation result, bars represent interquartile range and bold lines inside the box plot median levels. **A**, training cohort; **B**, internal validation cohort; **C**, European validation cohort; **D**, Asian validation cohort)

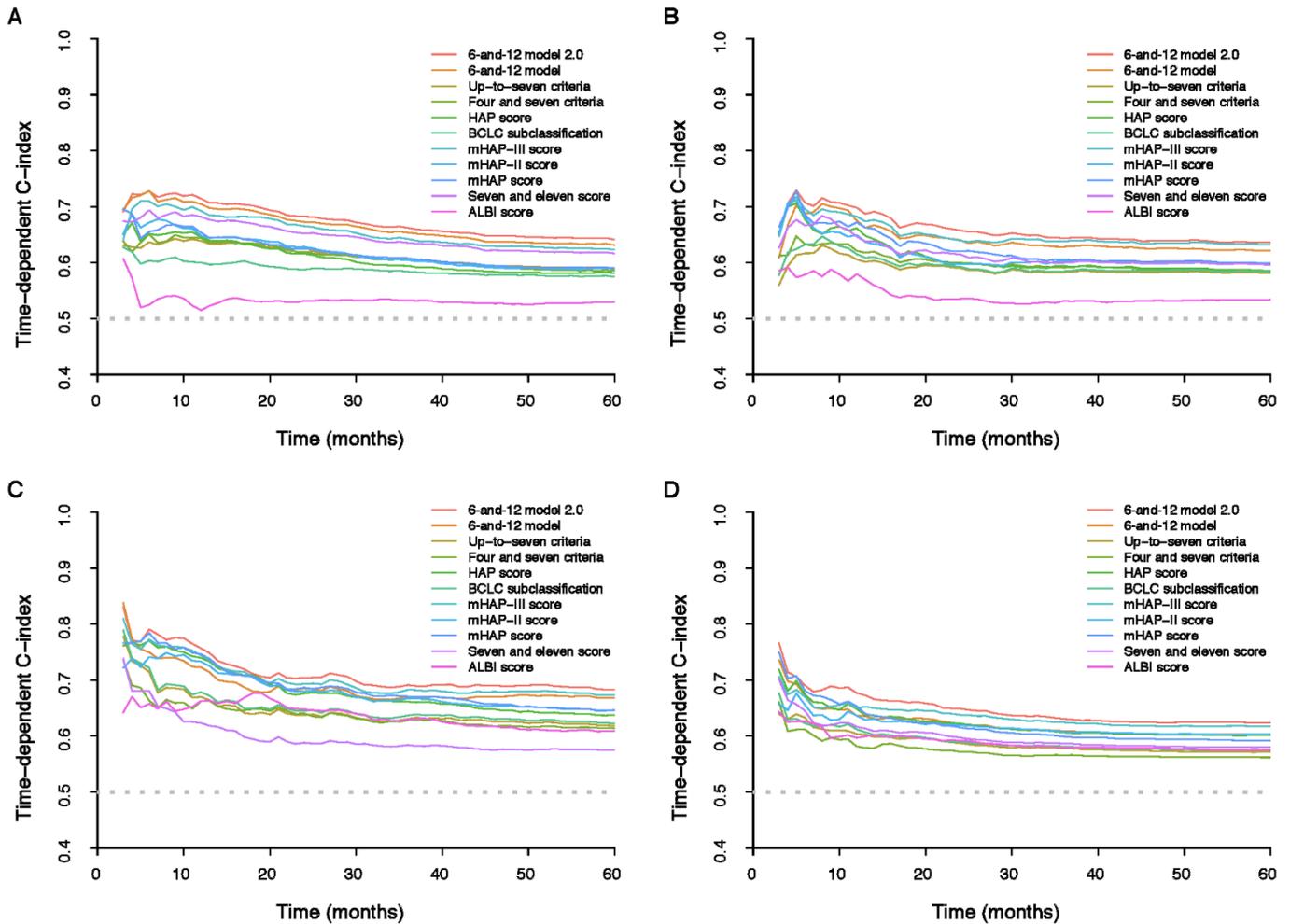


Fig. S5. Time-dependent C-index values of 6-and-12 model 2.0 and other available models. (A) training cohort; (B) internal validation cohort; (C) Asian validation cohort; (D) European validation cohort. Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; C-index, concordance index; HAP, hepatoma arterial-embolization prognostication.

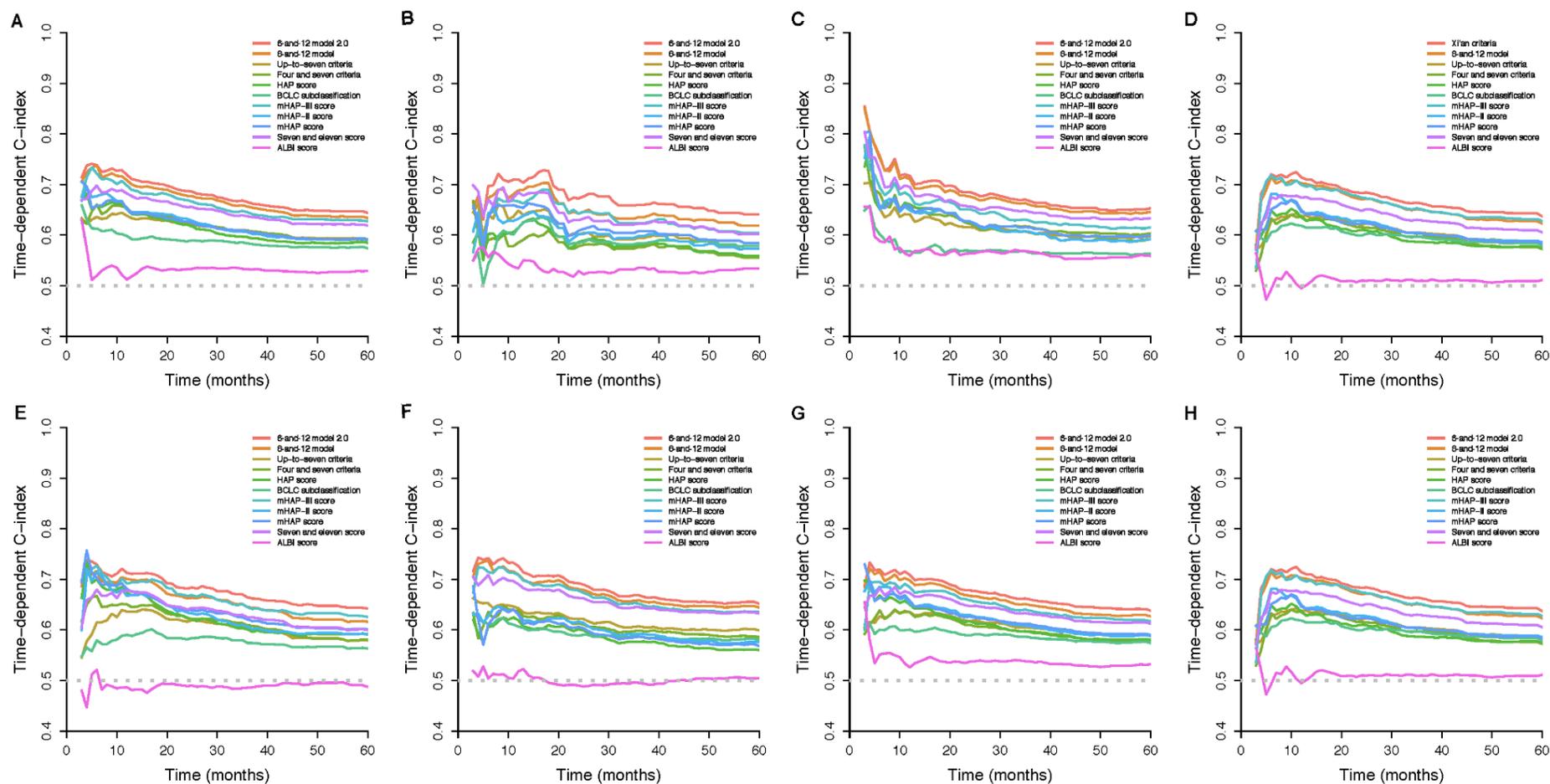


Fig S6. Time-dependent C-index values of 6-and-12 model 2.0 and other available models in different subgroups. (A) male; (B) female; (C) Age>60 years; (D) Age≤60 years; (E) ALBI grade 1; (F) ALBI grade 2; (G) HBV; (H) Other aetiology. Abbreviations: ALBI, albumin-bilirubin; BCLC, Barcelona Clinic Liver Cancer; C-index, concordance index; HAP, hepatoma arterial-embolization

prognostication.

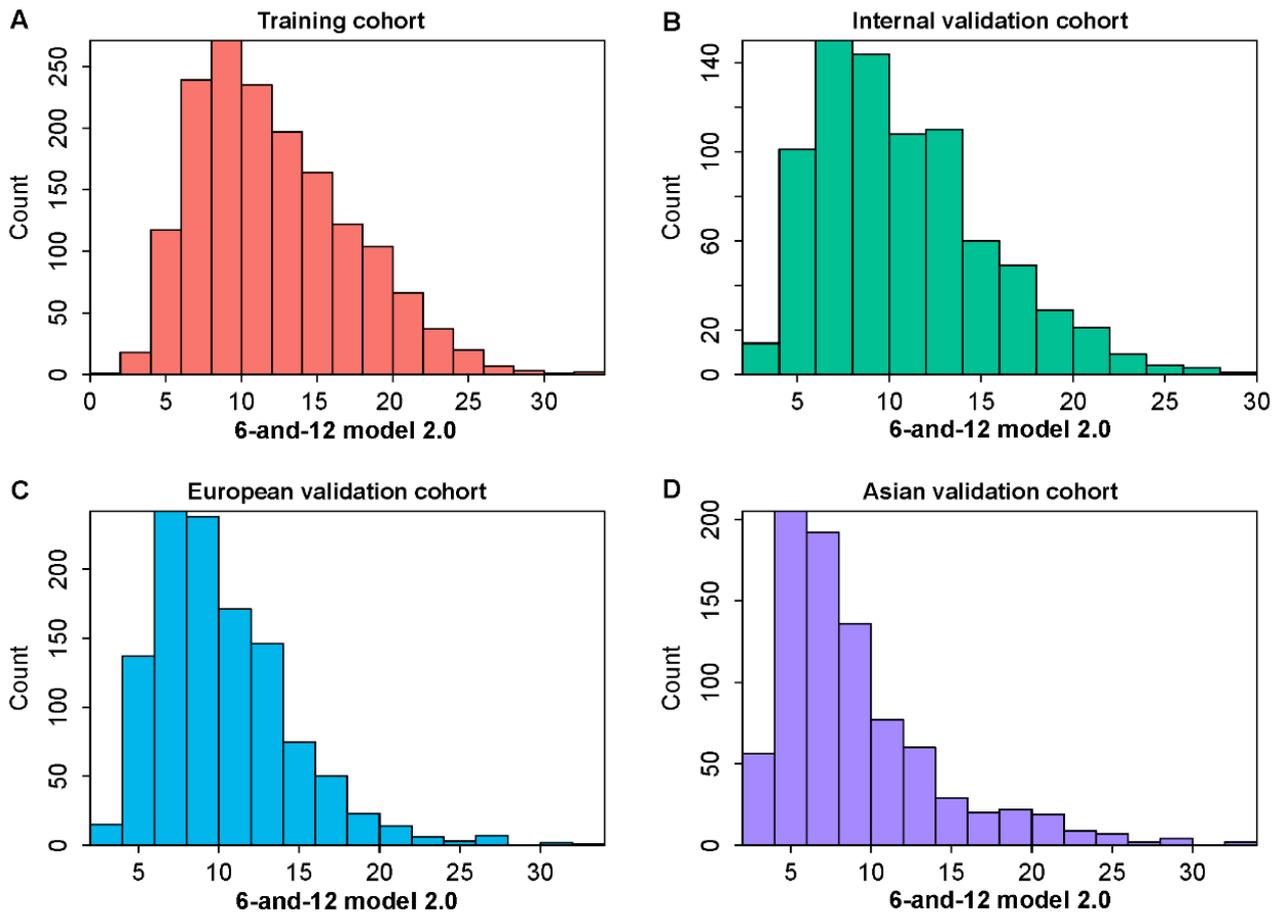


Fig. S7. Overall distribution of cases according to 6-and-12 model 2.0 in training cohort (A), internal validation cohort (B), European validation cohort (C), and Asian validation cohort (D).

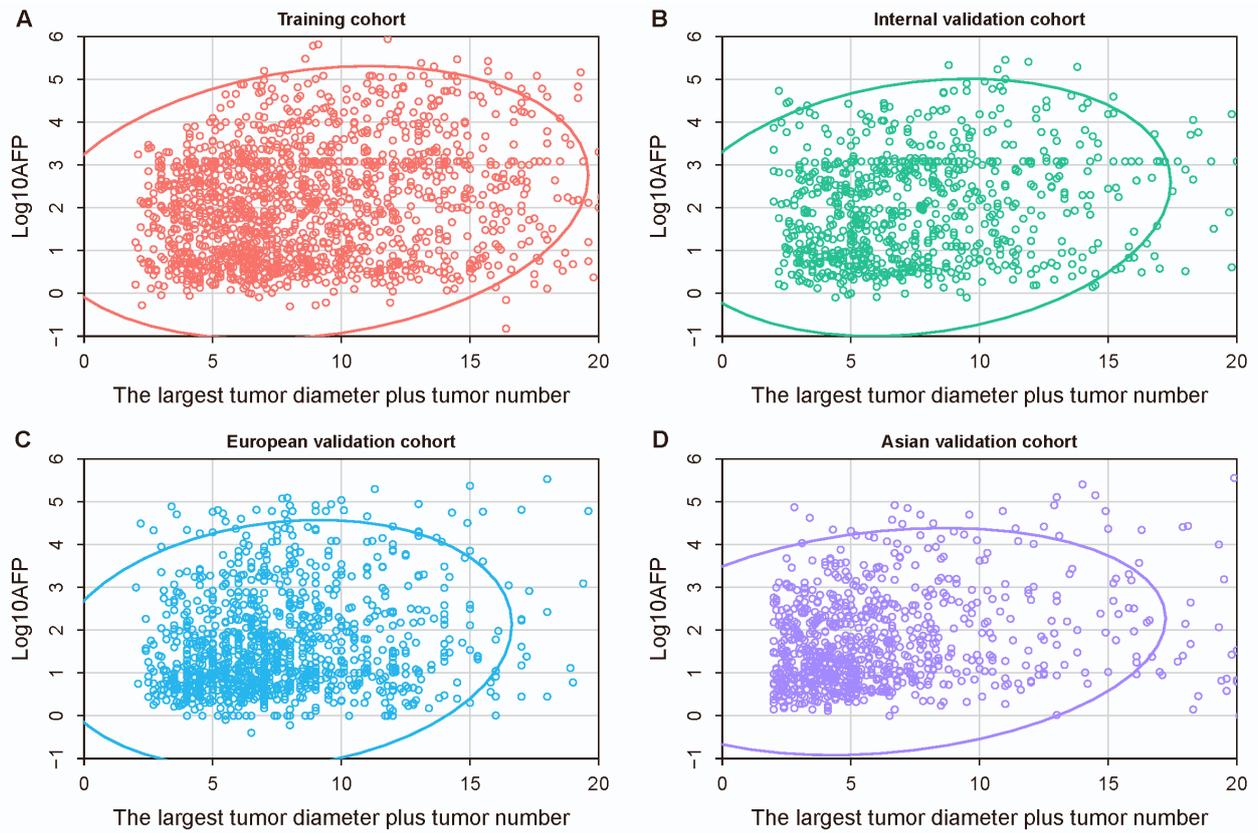


Fig S8. Overall distribution of cases according to baseline $\log_{10}\text{AFP}$ and tumor burden in training cohort (A), internal validation cohort (B), European validation cohort (C), and Asian validation cohort (D).

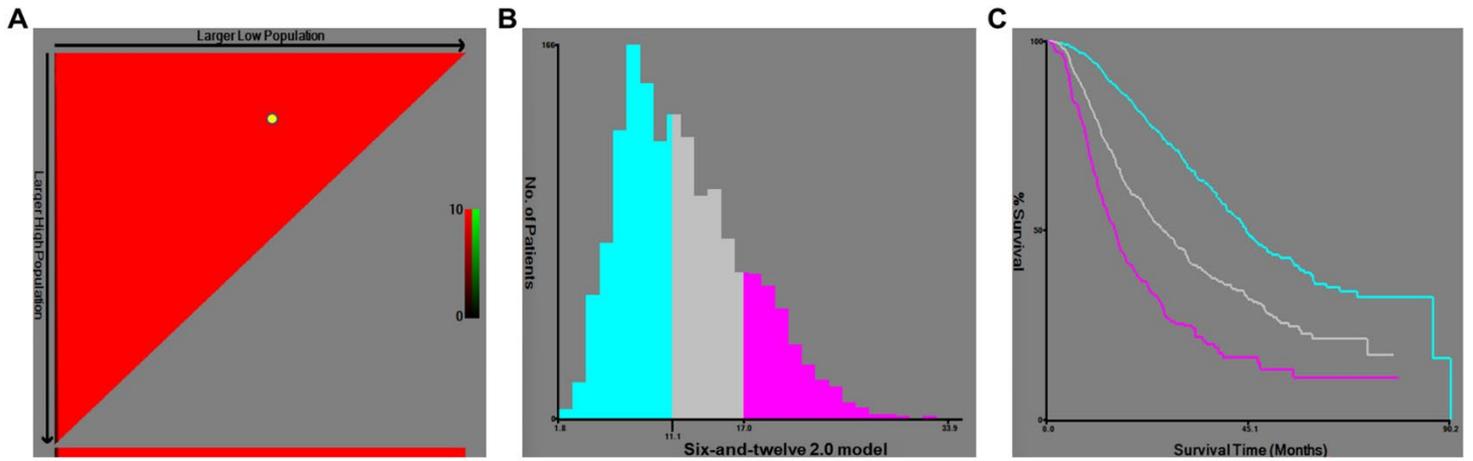


Fig. S9. Determination of the cut-offs of 6-and-12 model 2.0 by X-tile software.

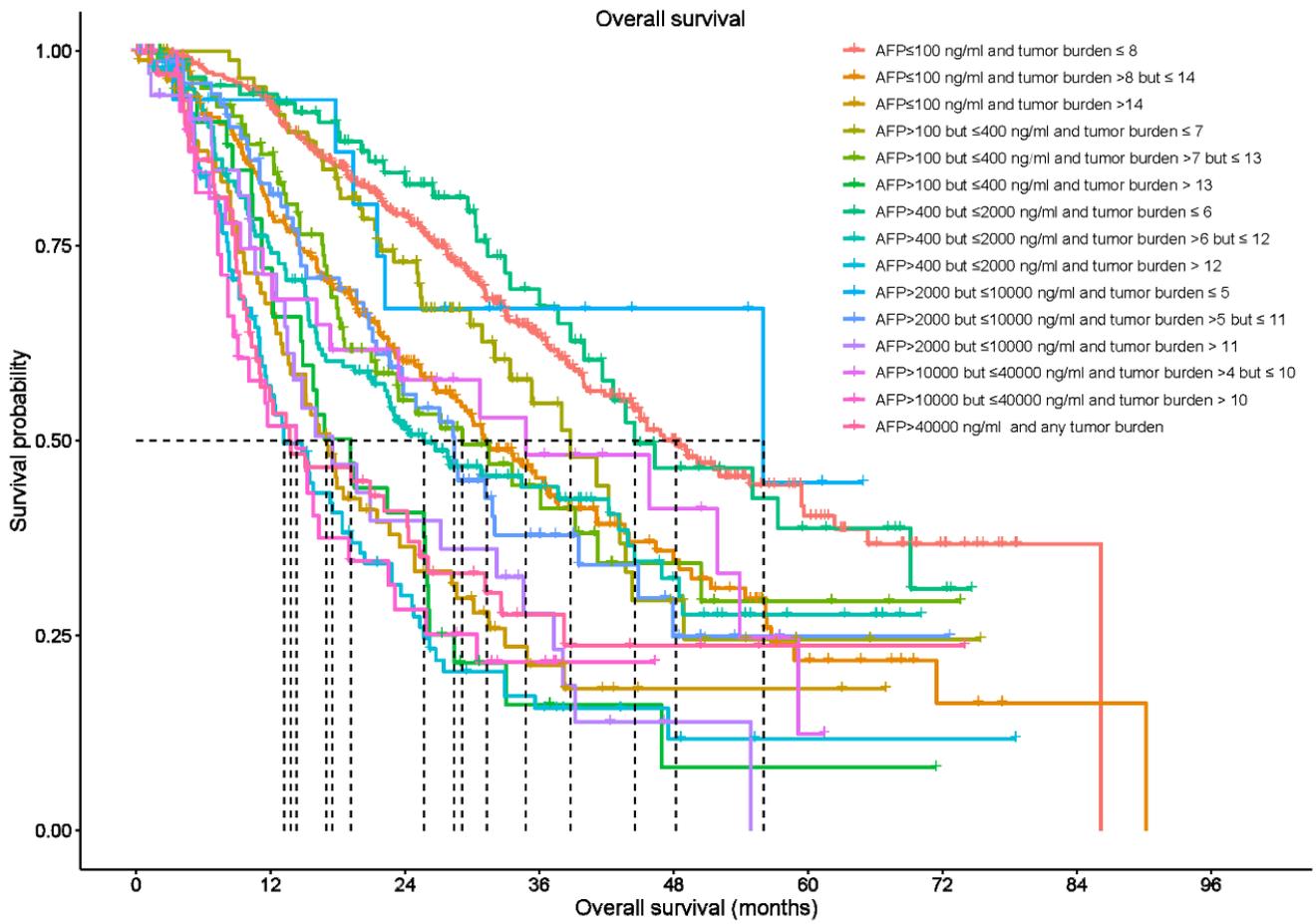


Fig. S10. Overall survival by Kaplan-Meier curve according to the risk stratification in different level of AFP value in training cohort.

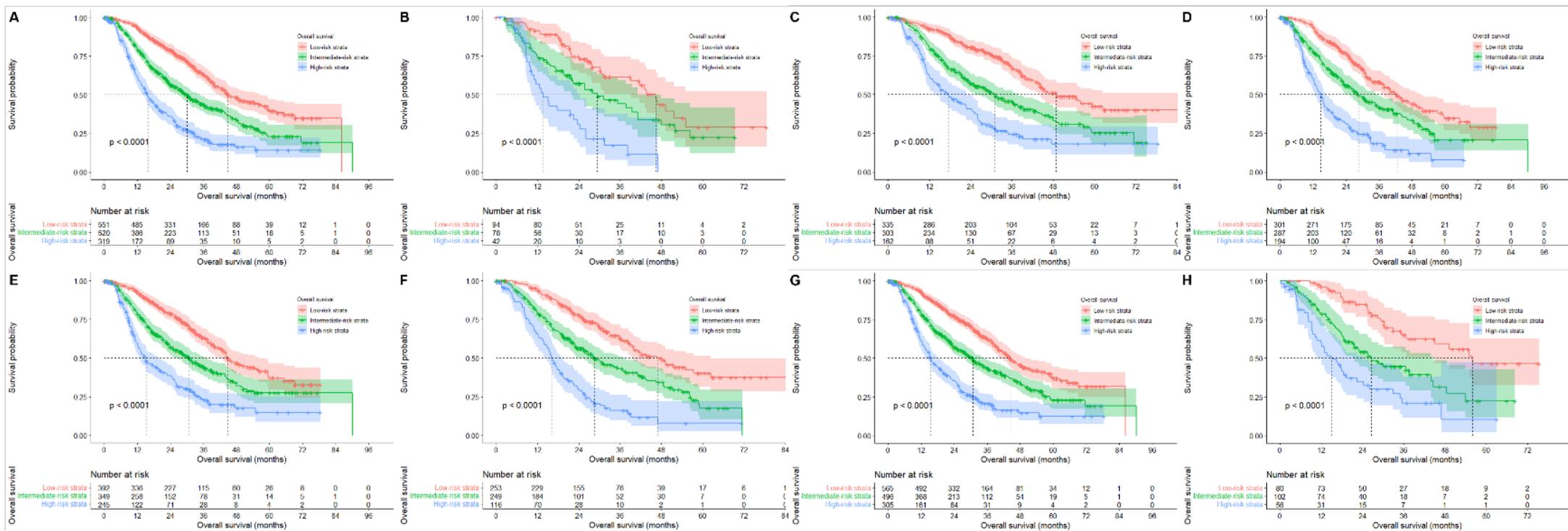


Fig. S11. Survival analyses by Kaplan-Meier method according to the risk stratification of 6-and-12 model 2.0 in different subgroups. (A, male; B, female; C, ALBI grade 1; D, ALBI grade 2; E, age ≤ 60 years; F, age > 60 years; G, HBV; H, other aetiologies, all p < 0.001 by log-rank test).

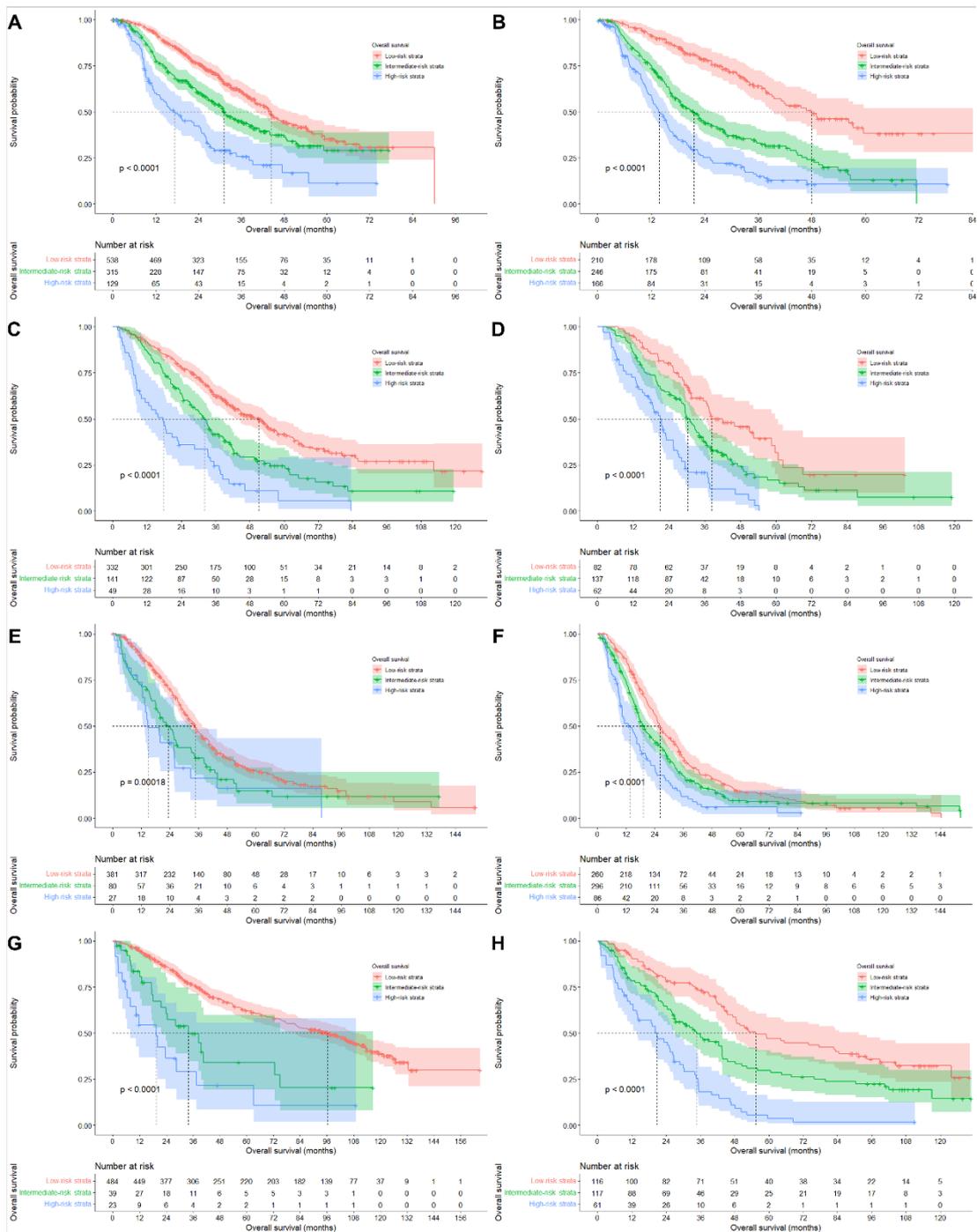


Fig. S12. Survival analyses by Kaplan-Meier method according to the risk stratification of 6-and-12 model 2.0 in BCLC-A and BCLC-B HCC among these four cohorts. (A, BCLC-A in training cohort; B, BCLC-B in training cohort; C, BCLC-A in internal validation cohort; D, BCLC-B in internal validation cohort; E, BCLC-A in European validation cohort; F, BCLC-B in European validation

cohort; **G**, BCLC-A in Asian validation cohort; **H**, BCLC-B in Asian validation cohort, all $p < 0.001$ by log-rank test).

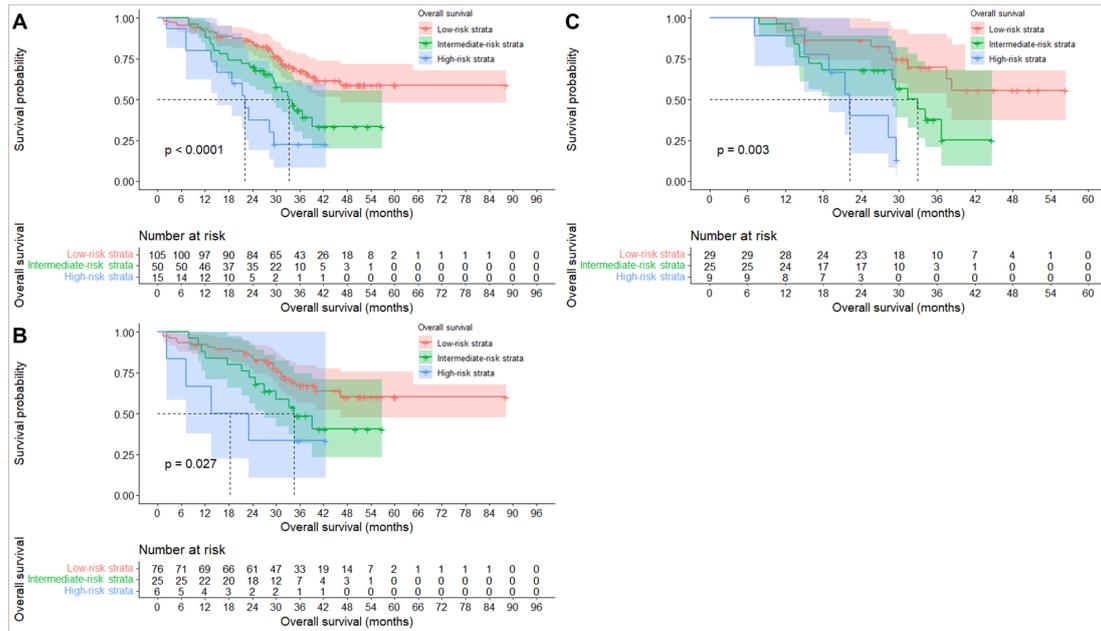


Fig.S13. Survival analyses by Kaplan-Meier method according to the risk stratification of 6-and-12 model 2.0 in Chinese DEB-TACE cohort. (A, whole cohort, $p < 0.001$ by log-rank test; B, BCLC stage A, $p = 0.027$ by log-rank test; C, BCLC stage B, $p = 0.003$ by log-rank test).

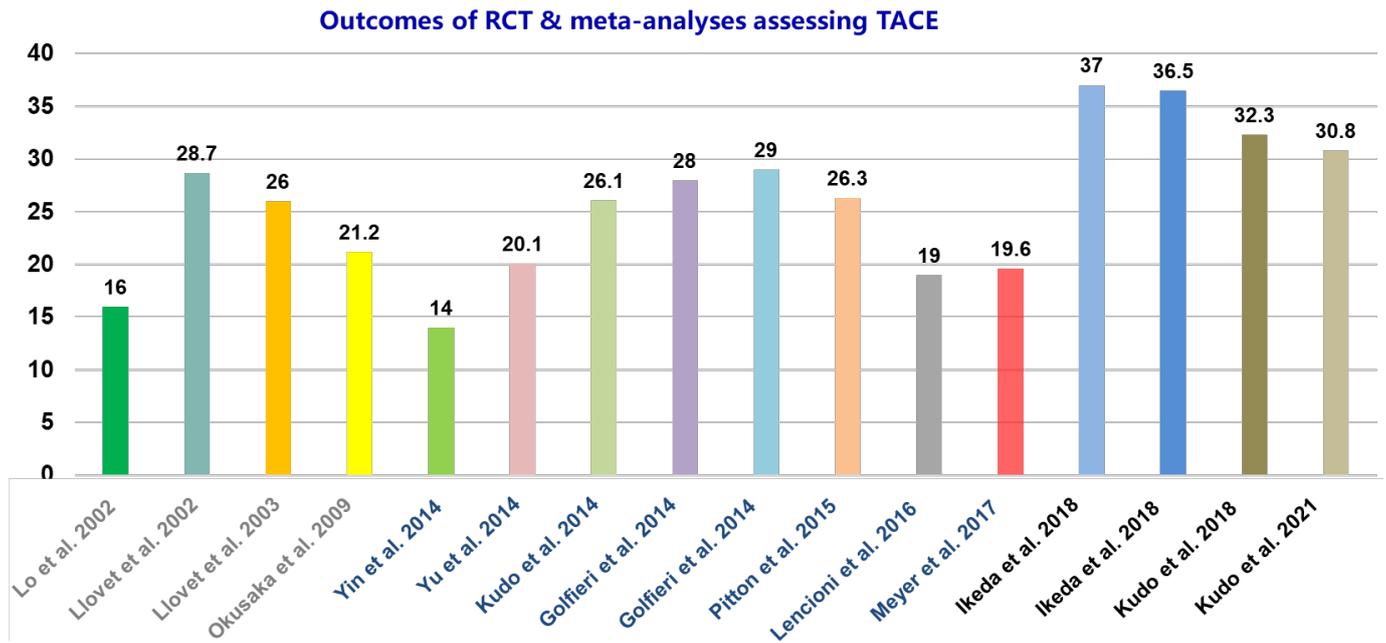


Fig. S14. The main outcomes of OS of TACE in pivotal randomized controlled trials and meta-analysis

References

- [1] Wang Q, Xia D, Bai W, et al. Development of a prognostic score for recommended TACE candidates with hepatocellular carcinoma: A multicentre observational study. *J Hepatol* 2019;70:893-903.
- [2] Adhoute X, Penaranda G, Raoul JL, et al. "Six-and-twelve" score for outcome prediction of hepatocellular carcinoma following transarterial chemoembolization. In-depth analysis from a multicenter French cohort. *World J Hepatol* 2020;12:525-532.
- [3] Adhoute X, Larrey E, Anty R, et al. Expected outcomes and patients' selection before chemoembolization-"Six-and-Twelve or Pre-TACE-Predict" scores may help clinicians: Real-life French cohorts results. *World J Clin Cases* 2021;9:4559-4572.
- [4] Adhoute X, Penaranda G, Raoul JL, et al. Hepatocellular carcinoma macroscopic gross appearance on imaging: predictor of outcome after transarterial chemoembolization in a real-life multicenter French cohort. *Eur J Gastroenterol Hepatol* 2019;31:1414-1423.
- [5] Kaewdech A, Sripongpun P, Cheewasereechon N, et al. Validation of the "Six-and-Twelve" Prognostic Score in Transarterial Chemoembolization-Treated Hepatocellular Carcinoma Patients. *Clin Transl Gastroenterol* 2021;12:e00310.
- [6] Muller L, Hahn F, Auer TA, et al. Tumor Burden in Patients With Hepatocellular Carcinoma Undergoing Transarterial Chemoembolization: Head-to-Head Comparison of Current Scoring Systems. *Front Oncol* 2022;12:850454.
- [7] Okusaka T, Kasugai H, Shioyama Y, et al. Transarterial chemotherapy alone versus transarterial chemoembolization for hepatocellular carcinoma: a randomized phase III trial. *J Hepatol* 2009;51:1030-1036.
- [8] Kudo M, Imanaka K, Chida N, et al. Phase III study of sorafenib after transarterial chemoembolisation in Japanese and Korean patients with unresectable hepatocellular carcinoma. *Eur J Cancer* 2011;47:2117-2127.
- [9] Yu SC, Hui JW, Hui EP, et al. Unresectable hepatocellular carcinoma: randomized controlled trial of transarterial ethanol ablation versus transcatheter arterial chemoembolization. *Radiology* 2014;270:607-620.
- [10] Golfieri R, Giampalma E, Renzulli M, et al. Randomised controlled trial of doxorubicin-eluting beads vs conventional chemoembolisation for hepatocellular carcinoma. *Br J Cancer* 2014;111:255-264.
- [11] Kudo M, Han G, Finn RS, et al. Brivanib as adjuvant therapy to transarterial chemoembolization in patients with hepatocellular carcinoma: A randomized phase III trial. *Hepatology* 2014;60:1697-1707.
- [12] Lencioni R, Llovet JM, Han G, et al. Sorafenib or placebo plus TACE with doxorubicin-eluting beads for intermediate stage HCC: The SPACE trial. *J Hepatol* 2016;64:1090-1098.
- [13] Meyer T, Fox R, Ma YT, et al. Sorafenib in combination with transarterial chemoembolisation in patients with unresectable hepatocellular carcinoma (TACE 2): a randomised placebo-controlled, double-blind, phase 3 trial. *Lancet Gastroenterol Hepatol* 2017;2:565-575.
- [14] Kudo M, Cheng AL, Park JW, et al. Orantinib versus placebo combined with transcatheter arterial chemoembolisation in patients with unresectable hepatocellular carcinoma

(ORIENTAL): a randomised, double-blind, placebo-controlled, multicentre, phase 3 study. *Lancet Gastroenterol Hepatol* 2018;3:37-46.

[15] Ikeda M, Kudo M, Aikata H, et al. Transarterial chemoembolization with miriplatin vs. epirubicin for unresectable hepatocellular carcinoma: a phase III randomized trial. *J Gastroenterol* 2018;53:281-290.

[16] Kudo M, Ueshima K, Ikeda M, et al. Final Results of TACTICS: A Randomized, Prospective Trial Comparing Transarterial Chemoembolization Plus Sorafenib to Transarterial Chemoembolization Alone in Patients with Unresectable Hepatocellular Carcinoma. *Liver Cancer* 2022;11:354-367.