Research article

JHEP|Reports

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

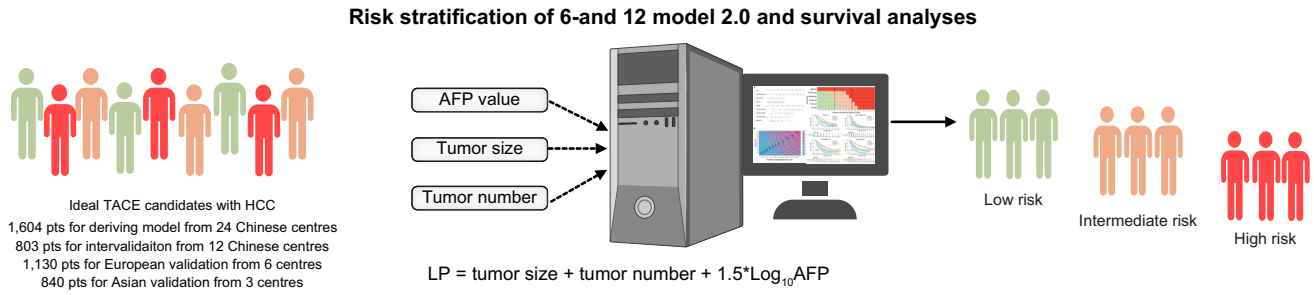
Authors

Dongdong Xia, Wei Bai, Qiuhe Wang, …, Dominik Bettinger, Arndt Vogel, Guohong Han

Correspondence

hangh@fmmu.edu.cn (G. Han).

Graphical abstract



Highlights:

- Development and validation of the 6-and-12 model 2.0 in a large-scale population of 4,377 TACE candidates.
- This model, involving tumor size, tumor number, and AFP, can refine individualized outcome predictions and stratify patients.
- This model outperformed the original 6-and-12 model and other existing metrics.
- This model is an easily accessible, reproducible, and user-friendly tool for simple and accurate clinical applications.

Impact and implications:

In this international multicentre study, we developed and internally and externally validated a novel outcome prediction model for candidates with HCC who would be ideal for TACE. The model, called the 6-and-12 model 2.0, was based on 4,377 patients from 39 centers in five countries. The model offers individualized outcome prediction, outperforming the original 6-and-12 model score and other existing metrics across all datasets and subsets. Based on different levels of alpha-fetoprotein (AFP) and corresponding cut-offs of tumor burden, patients could be stratified into three risk strata with significantly different survival prognoses, which could provide a referential framework to control study heterogeneity and define the target population in future trial designs.

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

© 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/). JHEP Reports, 2025, 7, 1–13

JHEP|Reports

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 Zeng[10], Pimsiri Sripongpun[11], Chunhui Nie[12], Seung Up Kim[13], Ming Huang[14], Wenhao Hu[15], Xiangchun Ding[16], Guowen Yin[17], Hailiang Li[18], Hui Zhao[19], Jean-Pierre Bronowicki[20], Jing Li[21], Jiaping Li[22], Xiaoli Zhu[23], Jianbing Wu[24], Chunqing Zhang[25], Weidong Gong[26], Zixiang Li[27], Zhengyu Lin[28], Tao Xu[29], Tao Yin[30], Rodolphe Anty[31], Jinlong Song[32], Haibin Shi[33], Guoliang Shao[34], Weixin Ren[35], Yongjin Zhang[36], Shufa Yang[37], Yanbo Zheng[38], Jian Xu[39], Wenhui Wang[40], Xu Zhu[41], Ying Fu[9], Chang Liu[10], Apichat Kaewdech[11], Rong Ding[14], Jie Zheng[15], Shuaiwei Liu[16], Hui Yu[17], Lin Zheng[18], Nan You[21], Wenzhe Fan[22], Shuai Zhang[23], Long Feng[24], Guangchuan Wang[25], Peng Zhang[26,42], Xueda Li[27], Jian Chen[28], Feng Zhang[30], Wenbo Shao[32], Weizhong Zhou[33], Hui Zeng[34], Gengfei Cao[35], Wukui Huang[37], Wenjin Jiang[38], Wen Zhang[39], Lei Li[40], Aiwei Feng[41], Enxin Wang[2,43], Zhexuan Wang[2], Dandan Han[2,44], Yong Lv[1,2], Jun Sun[2], Bincheng Ren[45], Linying Xia[46], 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 Zhao[3], Jielai Xia[47], Daiming Fan[1], Kaichun Wu[1], Dominik Bettinger[48], Arndt Vogel[49,50], Guohong Han[1,2,3,*], on behalf of China HCC-TACE study group

JHEP Reports **2025**. vol. 7 | 1–13

Check for updates

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-naive 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×$\log_{10}$ 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-naive 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] seven and eleven 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 $\chi 2$ 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, prodlim, 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 ts, tn, $\log_{10}$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: ts [hazard ratio (HR), 1.107; 95% CI, 1.088–1.126; p <0.001), tn (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 ts, tn, 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 ts and $\log_{10}$ AFP and 0.089 for tn 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:

$$\text{Linear predictor} = \text{tn+ts+1.5} * \log_{10}\text{AFP}$$

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 tn and ts), $\log_{10}$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),

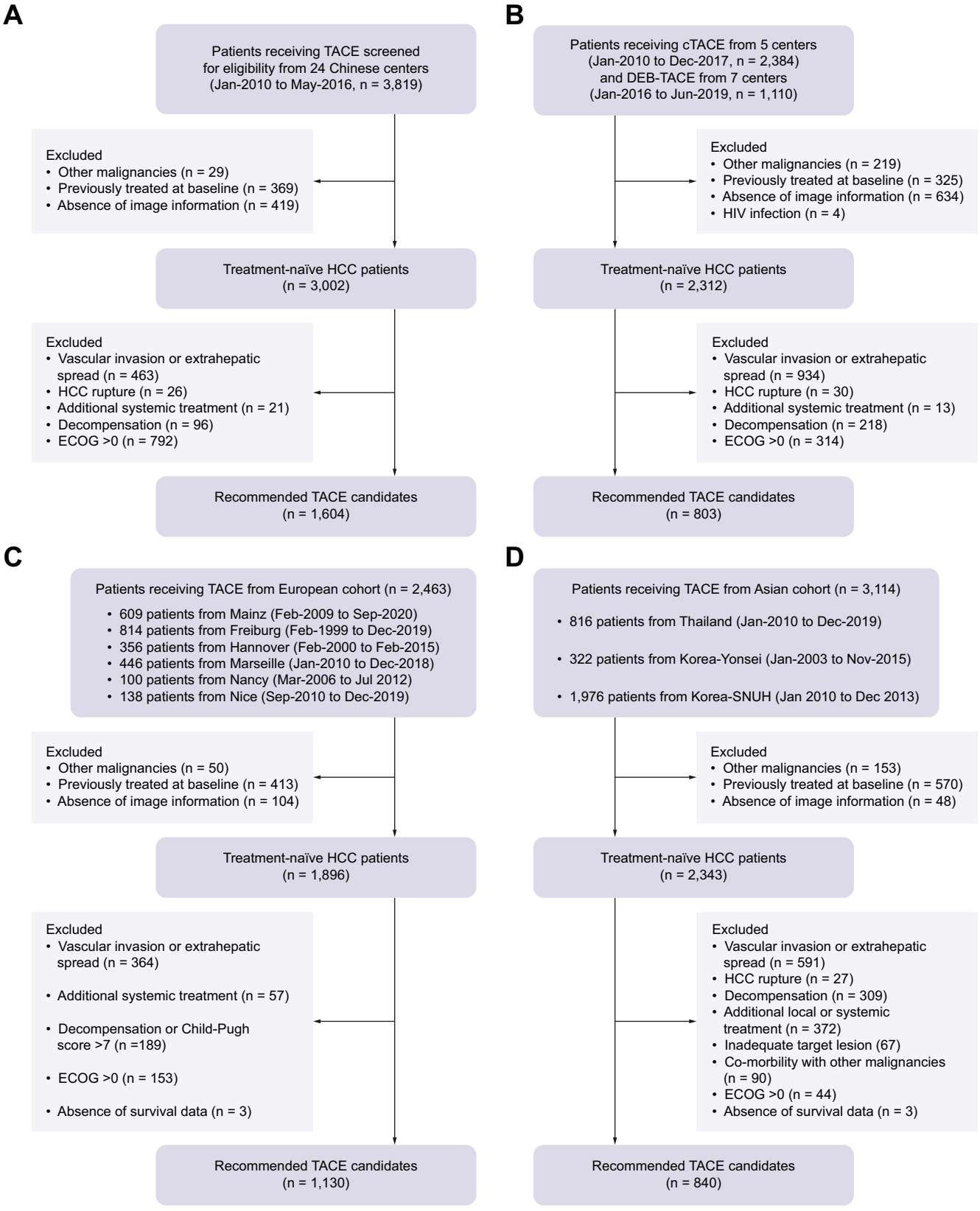


Fig. 1. Flow chart of the current study. (A) Training cohort; (B) Internal validation cohort; (C) European validation cohort; and (D) Asian validation cohort. cTACE, conventional transarterial chemoembolization; DEB, drug-eluting beads; HCC, hepatocellular carcinoma; HIV, human immunodeficiency virus; 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, $\times 10^9$/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, $\times 10^9$/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.

	Univariable Cox regression		Multivariable Cox regression		
Risk factors	**HR (95% CI)**	***p* value**	**HR (95% CI)**	***p* value**	**Beta coefficient**
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.

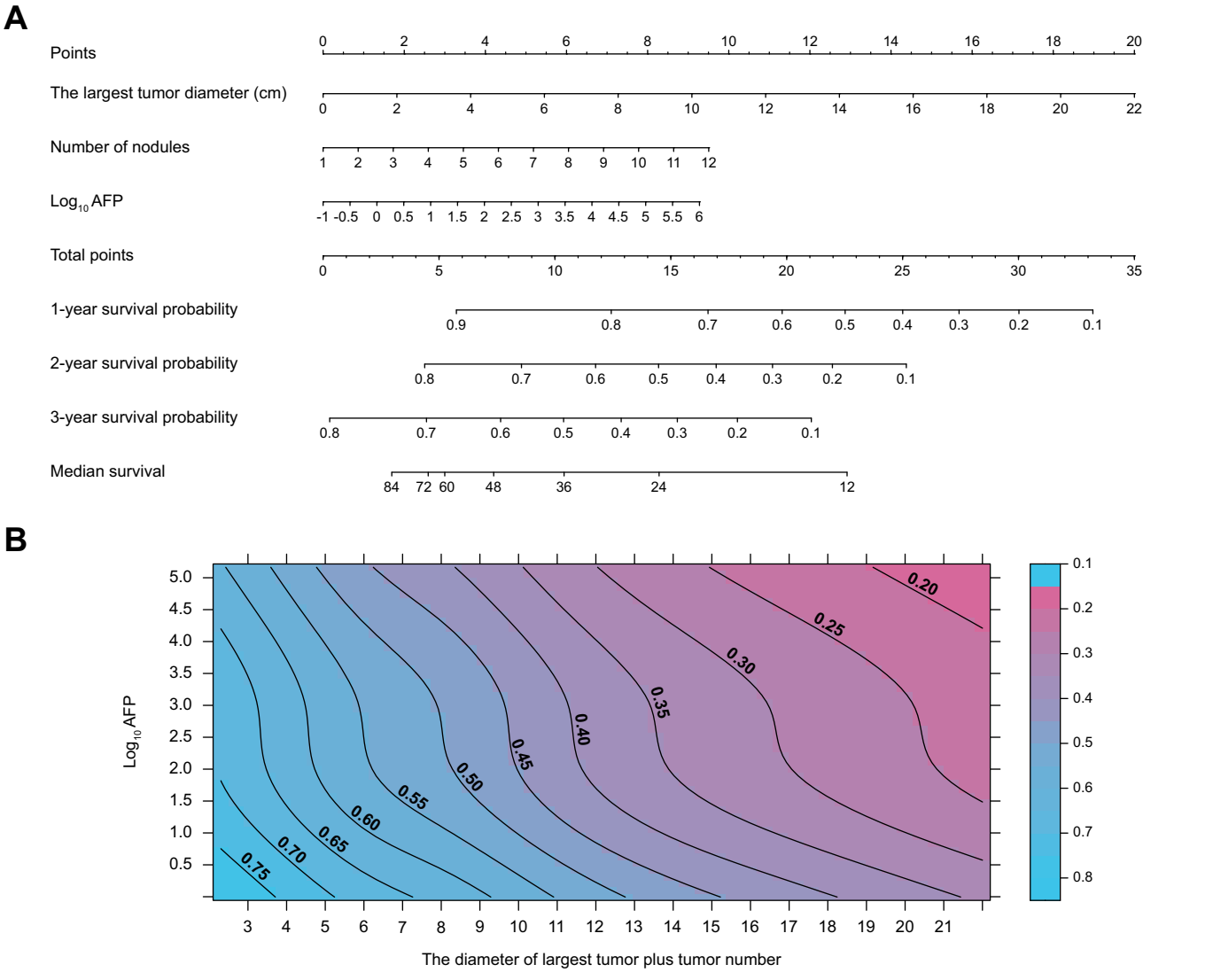


Fig. 2. Development of the prognostic model. (A) Nomogram of 6-and-12 model 2.0 for individual survival prediction; (B) Contour plot of 3-year survival probability according to tumor burden (largest tumor diameter plus tumor number) and Log_{10}AFP. AFP, alpha-fetoprotein.

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	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	0.678	0.013	5,493.00	5,505.42
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	0.660	0.013	5,513.14	5,521.42
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	0.616	0.011	5,550.24	5,554.38
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	0.612	0.011	5,548.71	5,552.85
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	0.573	0.010	5,564.62	5,568.76
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	0.620	0.012	5,536.66	5,540.80
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	0.638	0.013	5,566.06	5,570.19
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	0.671	0.013	5,515.46	5,519.60
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	0.645	0.013	5,554.60	5,558.74
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	0.647	0.010	5,537.54	5,547.68
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	0.614	0.014	5,581.81	5,585.94

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 $\log_{10}$ 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 (ts + tn), 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)	*p* value	NRI (95% CI)	*p* value	IDI (95% CI)	*p* value
6-and-12 model	17.2% (8.60–22.8%)	<0.001	1.6% (0.7–2.9%)	<0.001	11.0% (4.60–17.2%)	<0.001	1.4% (0.5–2.8%)	<0.001
Up to seven criteria	18.0% (7.90–26.0%)	0.006	4.5% (2.7–6.8%)	<0.001	15.4% (3.7–22.3%)	<0.001	4.4% (2.1–6.7%)	<0.001
Four and seven criteria	20.8% (10.0–29.7%)	<0.001	4.5% (2.6–6.7%)	<0.001	19.6% (10.1–27.5%)	<0.001	5.1% (2.9–7.5%)	<0.001
Seven and eleven criteria	14.0% (2.80–21.4%)	0.014	2.4% (0.9–4.1%)	0.002	7.0% (-3.4–15.8%)	0.264	1.6% (0.1–3.4%)	0.022
BCLC subclassification	28.4% (17.1–35.1%)	<0.001	6.0% (3.9–8.5%)	<0.001	15.3% (7.0–23.4%)	0.002	5.6% (3.0–8.4%)	<0.001
HAP score	24.0% (16.2–32.4%)	<0.001	5.3% (3.5–7.5%)	<0.001	26.0% (16.3–33%)	<0.001	7.4% (5.0–9.9%)	<0.001
mHAP III score	24.3% (15.4–32.0%)	<0.001	2.5% (1.2–4.2%)	<0.001	17.6% (9.7–24.9%)	<0.001	3.6% (2.2–5.4%)	<0.001
mHAP II score	25.0% (15.1–32.4%)	<0.001	4.7% (2.9–6.9%)	<0.001	20.5% (11.8–27.9%)	<0.001	6.0% (3.6–8.5%)	<0.001
mHAP score	21.4% (13.4–31.1%)	<0.001	4.6% (2.9–6.8%)	<0.001	21.6% (11.9–30%)	<0.001	5.9% (3.6–8.3%)	<0.001
ALBI score	37.2% (29.6–42.4%)	<0.001	8.3% (6.1–11.1%)	<0.001	26.2% (19.7–31.1%)	<0.001	10.3% (7.4–13.5%)	<0.001

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 prognosis 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

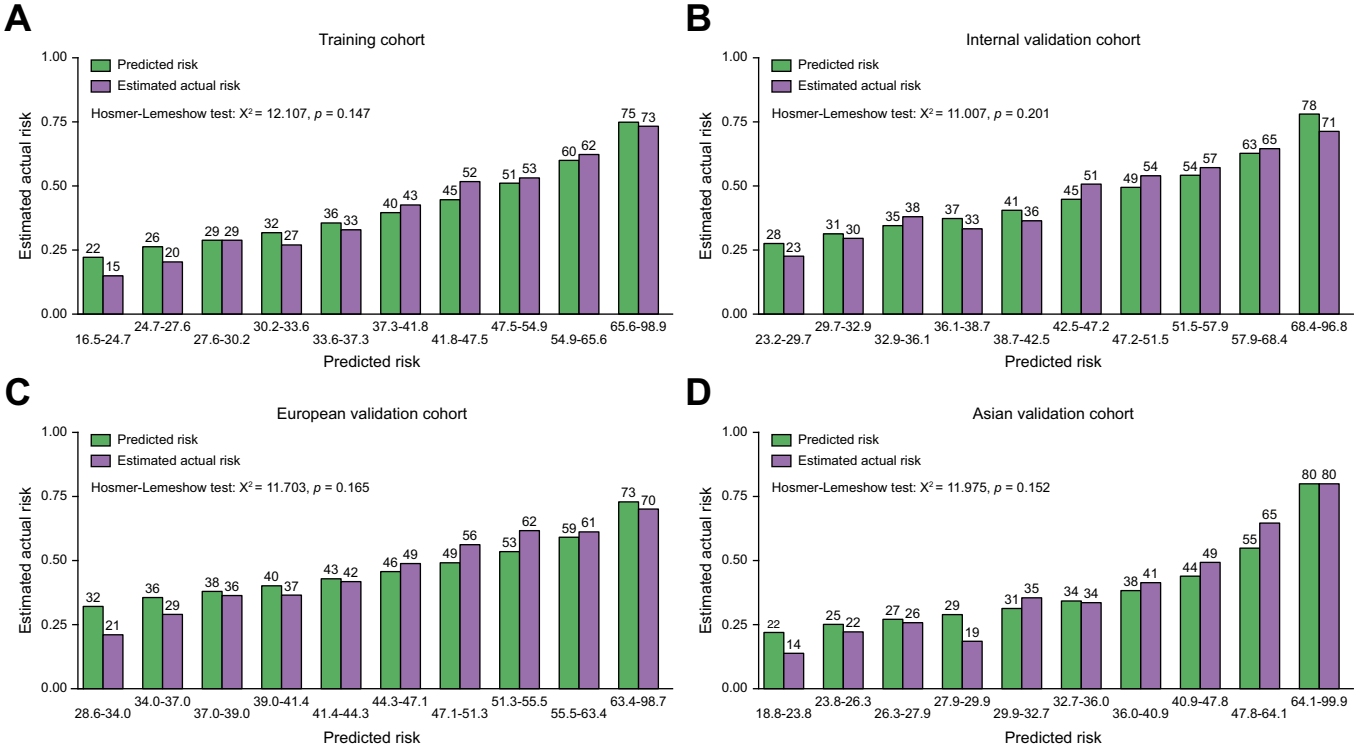


Fig. 3. Calibration analyses of 6-and-12 model 2.0 in candidates ideal for TACE. (A) Training cohort (Hosmer–Lemeshow test $\chi2 = 12.107$, $p = 0.147$); (B) Internal validation cohort (Hosmer–Lemeshow test $\chi2 = 11.007$, $p = 0.201$); (C) European validation cohort (Hosmer–Lemeshow test $\chi2 = 11.703$, $p = 0.165$); and (D) Asian validation cohort (Hosmer–Lemeshow test $\chi2 = 11.975$, $p = 0.152$). TACE, transarterial chemoembolization.

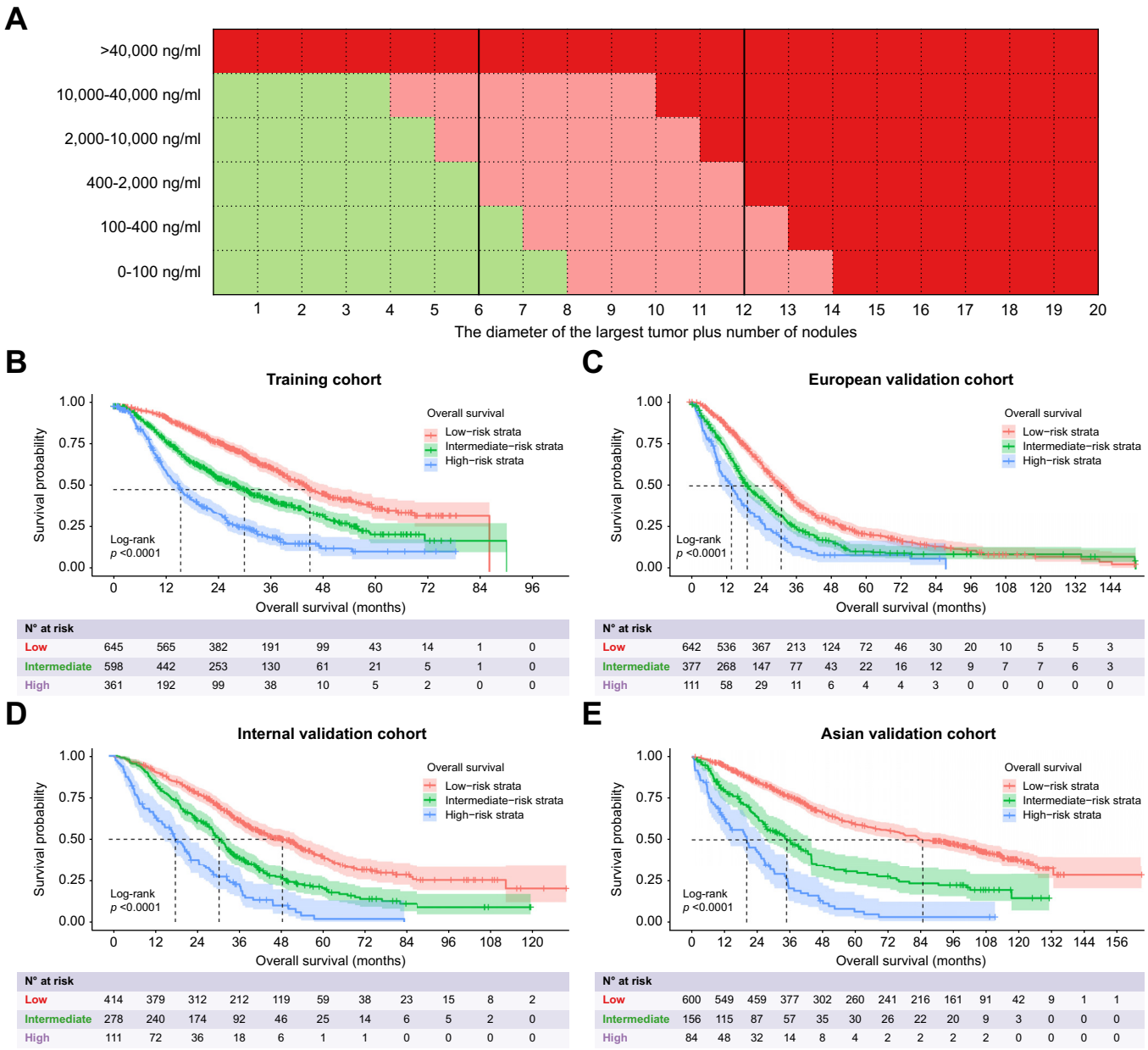


Fig. 4. Risk stratification of 6-and-12 model 2.0 and survival analyses by Kaplan–Meier method according to the stratification. (A) Risk stratification of 6-and-12 model 2.0. (B-E) Kaplan-Meier survival analysis of (B) Training cohort; (C) European validation cohort; (D) Internal validation cohort; and (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 objectivized, 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 underestimation 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/superselective 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

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

Statistics, Fourth Military Medical University, Xi'an, China; [48]*Department of Medicine II, University Medical Center Freiburg, Freiburg, Germany;* [49]*Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany;* [50]*Division 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; SD, Standard deviation; 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 multivariable 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