

# Prognostic models for predicting overall survival and cancer-specific mortality in hepatocellular carcinoma: A competing risk analysis

**Kongying Lin**

Mengchao Hepatobiliary Hospital of Fujian Medical University

**Qizhen Huang**

Mengchao Hepatobiliary Hospital of Fujian Medical University

**Zongren Ding**

Mengchao Hepatobiliary Hospital of Fujian Medical University

**Yongyi Zeng**

Mengchao Hepatobiliary Hospital of Fujian Medical University

**Jingfeng Liu** (✉ [drjingfeng@126.com](mailto:drjingfeng@126.com))

Mengchao Hepatobiliary Hospital of Fujian Medical University

---

## Research article

**Keywords:** hepatocellular carcinoma, SEER, competing risk nomogram, cancer-specific mortality, overall survival

**Posted Date:** December 3rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-46064/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Mortality due to hepatocellular carcinoma (HCC), which is the most common liver cancer, is often overestimated because of deaths from other causes. This study was conducted to estimate the probability of cancer-specific mortality (CSM) of patients with HCC and establish a competing risk nomogram for predicting the CSM of among these patients HCC using a large population-based cohort.

**Methods:** Patients diagnosed with HCC between 2004 and 2015 were identified from the Surveillance Epidemiology and End Results Program. CSM and overall survival (OS) were the endpoints of the study. A competing risk nomogram for predicting CSM was built using the Fine and Gray regression model, and the nomogram for predicting OS was constructed with the Cox proportional hazard regression model, and 10-fold cross-validation was performed for the entire set.

**Results:** A total of 34,957 patients were included in the study and randomly divided into a training set and validation set at a ratio of 7:3. Multivariate analysis identified age, race, surgical therapy, chemotherapy, radiotherapy, tumour diameter, and tumour staging as the independent predictive factors of CSM. In addition to these factors, sex and marital status were also identified as independent predictive factors of OS. Using these factors, corresponding nomograms were constructed for CSM and OS. In the validation set, the 5 year concordance-indices of the two nomogram models were estimated as 0.746 and 0.74. Calibration curves revealed good consistency between model predictions and observed outcomes. Furthermore, based on the results of cumulative incidence function analysis and Kaplan-Meier analysis, patients were categorised into four distinct risk subgroups, supporting the predictive performance of the models.

**Conclusions:** In this population-based analysis, we developed and validated nomograms for individualized prediction of CSM and OS in patients with HCC.

## Background

Hepatocellular carcinoma (HCC) is the most common liver cancer and the fourth leading cause of cancer-related death, with approximately 841,000 new cases and 782,000 deaths occurring annually worldwide[1]. The global incidence of HCC adjusted by age is approximately 10.1 cases per 100,000 person-year and is expected to increase in the future[2]. Most patients with HCC are diagnosed in the advanced stages owing to the lack of specific symptoms and exhibit poor prognosis because of unfavourable tumour biology[3, 4].

Cancer-specific death and death from other causes are frequent in HCC, similar to other types of cancer, and are considered competitive events as they are mutually exclusive, i.e., the occurrence of one event prevents the occurrence of the other. However, Kaplan-Meier methods and Cox regression models used in traditional survival analysis can only handle one endpoint, and thus may overestimate the risk of the event of interest [5-7]. The Fine and Gray model based on the sub-distribution hazard has been recommended to overcome this problem [5, 6, 8]. Several studies have analysed the independent predictive factors that predict related malignant diseases by utilizing competing risk analysis and have established models with good predictive performance for cancer-specific death [9-12]. However, little studies have analysed and constructed a prognostic model for cancer-specific mortality (CSM) in patients with HCC using the competing risk method.

Therefore, we conducted this study to evaluate CSM and overall survival (OS) of patients with HCC and to develop corresponding nomograms that can offer physicians a quantitative tool to facilitate risk stratification and make management-related decisions.

## Materials And Methods

### Data source

Data on patients diagnosed with HCC between 2004 and 2015 were extracted from the Surveillance, Epidemiology, and End Results (SEER) database. Data were extracted using SEER\*Stat 8.3.5 from the Surveillance Research Program of the Division

of Cancer Control and Population Science, National Cancer Institute.

## Patients

Patients with HCC were identified based on the International Classification of Disease for Oncology, third edition primary site code C22.0 and Histologic type code 8170-8175. We excluded patients with missing data, which were recorded as 'blank' in the database. Additionally, patients aged < 18 years or those with history of other primary malignancies, with invalid follow-up data, or with undefined data recorded as 'unknown' in the database were also excluded from the study. The detailed inclusion and exclusion criteria are shown in Figure 1.

## Variable selection

Information on demographic factors (race, age, sex, and marital status), tumour diameter, tumour stage according to American Joint Committee on Cancer (AJCC) staging system, therapeutic factors (surgery, chemotherapy, and radiotherapy) and follow-up data were collected from the SEER database.

Marital status was recorded as single (never married), separated, divorced, widowed, unmarried or domestic partner, and married (including common law) in the SEER database. We grouped single (never married), separated, divorced, widowed, and unmarried or domestic partner into the unmarried category. No information was available on tumour staging according to the eighth edition AJCC staging system, and only patients diagnosed after 2010 showed information on diagnosis according to the seventh edition AJCC staging system in the SEER database. To enrol as many patients as possible, the sixth edition AJCC staging system was accepted for further analysis. Based on the Surgery Codes of the SEER program, we divided surgical procedures into four categories: no surgery, local tumour destruction (e.g., heat-radio-frequency ablation, and percutaneous ethanol injection), resection, and transplantation.

The endpoints of the study were CSM and OS. The specific cause of death was determined based on the code of 'SEER cause-specific death classification' in the SEER database. OS was calculated from the date of diagnosis to the date of death due to any cause or the most recent follow-up. CSM was defined as the interval between the date of diagnosis and the date of the most recent follow-up or of the death caused by HCC alone. The median follow-up time was calculated by the reverse Kaplan-Meier method.

## Statistical analysis

Demographic and clinical variables were summarized by descriptive statistics. Categorical variables were expressed as percentages (%). The entire set was randomly divided into training and validation sets at a ratio of 7:3. The comparison of categorical variables between training and validation sets were performed using the chi-square test. CSM and death from other causes were regarded as two competing endpoint events, and the associations between variables and the risk of CSM were evaluated by Fine and Gray's competing risk analysis[6]. The corresponding cumulative incidence of CSM in different groups were depicted by the cumulative incidence function (CIF) and were compared using the Gray's test [6, 8, 13]. Variables with  $p < 0.05$  as determined by univariate analysis or with clinically relevant results were then evaluated using multivariate analysis based on the Fine and Gray's proportional sub-distribution hazard ratio model. The independent predictive factors in the multivariate analysis were incorporated in the nomogram model to predict the 1-, 3-, and 5-year CSM probability[7]. In addition, with aim to compare the difference between competing risk analysis and traditional survival analysis, we also performed the traditional Kaplan-Meier and Cox proportional hazard regression analysis for CSM in present study. In traditional Kaplan-Meier and Cox proportional hazard regression analysis, the competing event was regarded as censored data.

For OS, the independent risk factors were identified by univariate and multivariate Cox proportional hazard regression analyses, and the corresponding nomogram model was then constructed to predict 1-, 3-, and 5-year OS probabilities.

The predictive performance of nomogram models was analysed from perspectives of discrimination and calibration. The discriminative ability of the models was tested using the concordance index (C-index), and calibration of models was tested by

calibration curves [14, 15]. In addition, 10-fold cross-validation was conducted on the entire set, which was randomly partitioned into 10 equal sized subsamples[11]. Furthermore, CIF curves with Gray's test or Kaplan-Meier curves with log-rank test were also conducted to measure the performance of the models. The risk groups were classified according to previously recommended cut-points for predictive models (16th, 50th, and 84th)[16], which classified patients into good, fairly good, fairly poor, and poor risk groups based on their personalized total points determined using the nomogram models.

All statistical analyses were two-sided, and results with  $p < 0.05$  were regarded as statistically significant. SPSS version 20 (SPSS, Inc., Chicago, IL, USA) and R version 3.5.2 (R Project, Vienna, Austria) (R packages 'table1', 'cmprsk', 'mstate', 'rms', 'survival', 'survminer', 'riskRegression' and 'pec') were used to perform statistical analysis.

## Results

### Basic characteristics of patients

According to the inclusion and exclusion criteria, 34,957 patients diagnosed with HCC between 2004 and 2015 were included for analysis (Figure 1). The entire cohort was then randomly divided into a training set ( $n = 24,469$ ) and validation set ( $n = 10,488$ ) at a ratio of 7:3. The basic clinicopathological features of the entire cohort as well as those of the training and validation sets are shown in Table 1. In the entire cohort, most patients were Caucasian (67.9%), male (77.2%), and younger than 60 years (46.0%). In terms of therapy for HCC, 67.3% of patients did not undergo surgery for the primary nodules, 13.2% of patients were treated by local tumour destruction, 11.8% of patients underwent liver resection, and 7.7% of patients underwent liver transplantation surgery. Regarding tumour characteristics, tumours diameter smaller than 3 cm (33.3%) and stage I per AJCC criteria (41.9%) were most common. There were no significant differences among clinicopathological features between the training and validation sets.

### Follow-up and survival analysis of patients

The median follow-up time was 63 months (range: 1–155 months) months. Of the 34,957 patients, 9,840 had survived throughout the follow-up, 21,044 died from HCC, and 4073 died from other causes. For the training set, the 5-year OS, CSM, and death due to other causes were 24.2%, 63.9%, and 11.9%, respectively. The 3- and 5-year cumulative incidence of mortality and CIF curves corresponding to each clinicopathological variable are shown in Table 2 and Figure 2.

In the traditional survival analysis for CSM, the cumulative incidence of CSM estimated by Kaplan-Meier function were higher than that estimated by CIF, the 5-year cumulative incidence of CSM estimated by Kaplan-Meier analysis in the training set was 69.1%. (Table S1)

### Identification of risk factors and construction of nomograms

Univariate and multivariate analysis were performed in the training set to identify independent risk factors associated with CSM and OS (Table 3, 4 and Table S3).

For CSM, multivariate analysis based on Fine-Gray's competing risk analysis showed that age, race, surgical therapy, chemotherapy, radiotherapy, tumour diameter, and tumour stage were independent risk factors for CSM. Specifically, old age, white race, absence of surgical therapy, absence of radiotherapy, absence of chemotherapy, larger tumour diameter, and advanced tumour stage were associated with increased probability of CSM. In the multivariate traditional Cox regression analysis, these seven variables were again identified as independent risk factors of CSM. Moreover, sex and marital status were identified as additional independent risk factors.

Multivariate Cox regression analysis showed that all the aforementioned nine variables were independent risk factors for OS. Specifically, old age, male sex, white race, unmarried status, absence of surgical therapy, absence of radiotherapy, absence of chemotherapy, larger tumour diameter, and more advanced tumour stage were associated with poorer OS.

Based on the associated independent risk factors in corresponding multivariate analysis, the nomogram for predicting OS and competing risk nomogram for predicting CSM were constructed (Figure 3). Nomogram, which integrate various prognostic factors, is an easy-to-apply graphical tool for personalized prediction of survival probability of patients. In the nomogram, each variable has corresponding points that cross the scale. After adding the points of all variables, we can obtain the total points, which can be used to estimate the probability of event occurrence by drawing a line downward from the location of the total points to the survival axes.

### **Predictive performance of nomogram models**

The predictive performance of nomogram models was verified via the C-index and calibration curve in the training and validation sets.

For the competing risk nomogram for CSM, the median unadjusted C-index for different years of the model reached 0.749(Range, 0.740--0.781) in the training set and 0.754 (Range, 0.746-0.784) in the validation set, respectively. Furthermore, the median 10-fold cross-validation adjusted C-index in the entire cohort reached 0.750 (Range, 0.741–0.781), which indicated the robustness of the model. The calibration plots also displayed good agreement between the predictions of the competing risk nomogram model and observation in the probability of 3- and 5-year CSM in the training and validation sets (Figure 4 and Figure S1). For the traditional model, the median unadjusted C-index reached 0.758(Range, 0.753-0.783) in the training set and 0.759(Range, 0.754-0.780) in the validation set, respectively, and the adjusted C-index reached 0.755(Range, 0.751-0.779) in the entire set. (Table S2)

For the OS nomogram, the C-index values were 0.745 (Range, 0.740-0.768) in the training set and 0.743(Range, 0.740-0.764) in the validation set, and the adjusted C-index reached 0.742(Range, 0.738-0.764) in the entire set (Table S2). Calibration curves for 3- and 5-year were also well-matched with the standard lines (Figure 4 and Figure S1).

Based on the nomograms, each patient was assigned corresponding total points for probability of CSM and OS. The median total points calculated for CSM and OS were 148 (range: 7–254) and 169 (range: 16–298) in the training set and 148 (range: 7–254) and 169 (range: 16–296) in the validation set. Based on previously reported cut-off points (16th, 50th, and 84th of total points in the training set)[16], patients were categorised into four distinct risk groups. CIF and Kaplan-Meier analysis also showed that the curves corresponding to the four risk groups were clearly separated in the training and validation sets (both  $p < 0.001$ ), further supporting the good predictive performance of the nomogram models (Figure 5).

## **Discussion**

Cancer-specific death and other cause-specific death are mutually exclusive events in oncology research. The traditional Kaplan–Meier and Cox regression methods, which take competitive events as censoring data, tend to overestimate the incidence of CSM[17, 18]. Therefore, there may be deviation in the prognosis assessment of patients by clinicians, creating a substantial psychological burden on patients and negatively affecting their lives. In the present study, we conducted a real-world study based on the SEER database to identify independent predictive factors of CSM of patients diagnosed with HCC using the competing risk method and established a competing nomogram model for individualized prediction of CSM. We also constructed a model for predicting OS. Both models achieved excellent predictive efficiency, which could potentially help clinicians assess prognosis more accurately.

In the present study, we also compared the traditional Kaplan-Meier and Cox regression analysis with CIF and Fine-Gray's competing risk regression analysis. To our knowledge, this is the first study to compare the two algorithms in the prognostic analysis of HCC. Although there is no obvious difference in the predictive performance of the models, there are some problems that deserve our attention. Firstly, in the presence of competing events, traditional survival analysis regarded competing events as censoring data, which may overestimate the incidence of interest event. Our results showed that the cumulative incidence of CSM estimated by Kaplan-Meier method was higher than that of CIF function, which was consistent with previous research on other malignancies[19, 20]. Similarly, the Cox regression model also overestimated the incidence of CSM, and the degree of

overestimation became more obvious with the extension of follow-up. (Figure S2) Overestimating the incidence of CSM will bring huge psychological burden to patients and affect their lives. Therefore, in the case of competitive events, the competing risk model will be a better choice for accurately predicting the incidence of interested events, especially for those elderly patients or those with early tumour stage, whose competing events will occur more frequently. In addition, the results of multivariate analysis were also different in the Fine-Gray's analysis and Cox regression analysis. Sex and marital status was identified as independent risk factor for CSM in Cox regression model, however, these two variables were not significant in the Fine-Gray's analysis.

The two models constructed for prediction of CSM and OS included nine parameters: age, race, sex, marital status, surgical therapy, chemotherapy, radiotherapy, tumour diameter, and AJCC staging. Marital status and sex were the only difference between the two models. Previous research showed that marital status is one factor affecting prognosis[21, 22]; this may be because a close and cohesive family increases the likelihood of adherence, and psychological and economic support from spouses may contribute to improvements in survival in married patients[23-25]. Furthermore, several studies based on the SEER database indicated that HCC patients who were married had a better prognosis[26-28]. However, our competitive risk analysis showed that marital status was significantly associated with well OS but not CSM. Therefore, the marital status mainly associated with other cause-specific death for HCC but has little association with cancer-specific death.

By comparing the prognosis outcomes of different surgical treatments, we found that liver transplantation remains the most effective treatment. Interestingly, patients who underwent liver transplantation had comparable 5-year CSM and other causes-specific mortality. In the entire set, 2686 patients were administered liver transplantation and 784 died during follow-up, including 393 from other causes and 391 from HCC, with 5-year CSM and other causes-specific mortality of 13.4% (95%CI, 12.0–14.8%) and 10.5% (95%CI, 9.2–11.7%), respectively. Therefore, survival analysis of liver transplantation patients with HCC should consider the influence of competing events.

Although using population-based data from SEER can reduce selection or treatment biases associated with small sample sizes or single-centre data analysis, there were several limitations to this study. First, this was a retrospective study. Second, although we included a large number of multicentre queues, all patients were from the United States. As there may be differences in treatments and the management of HCC among different countries, international multicentre studies are needed to accurately estimate and generalise the predictive performance of the models. Third, not all previously reported factors were recorded in the SEER database, such as the aetiology of HCC and genomic data. Considering the heterogeneity of HCC, the inclusion of these variables may improve the predictive power of the models.

## Conclusion

Overall, in this population-based study, we developed and validated nomogram models for individualized prediction of CSM and OS in patients with HCC. These simple tools can help clinicians in identifying high-risk groups and can guide clinical decision making. For the patients, the models will help answer consultation questions from patients and provide personalized prognosis assessments.

## Abbreviations

HCC, hepatocellular carcinoma; CSS, cancer-specific survival; OS, Overall survival; CSM, cancer-specific mortality; SEER, Surveillance, Epidemiology, and End Results; AJCC, American Joint Committee on Cancer; CIF, cumulative incidence function; OCSS, other cause-specific survival.

## Declarations

**Ethics approval and consent to participate:**

As all the data of this study were derived from the SEER database, institutional review board approval and consent to participate were not required.

#### **Consent for publication:**

Not applicable.

#### **Availability of data and materials:**

All the data of this study were derived from the SEER database ([www.seer.cancer.gov](http://www.seer.cancer.gov)).

#### **Competing Interests:**

The authors declare no conflict of interest.

#### **Funding:**

This study was supported by Key Clinical Specialty Discipline Construction Program of Fuzhou, Fujian, P.R.C (Grant number: 201912002) and Fujian Provincial Medical Center of Hepatobiliary. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in the writing of the manuscript.

#### **Author Contributions:**

KL, YZ, and JL designed the project. KL, QH, and ZD performed the experiments and KL contributed statistical analysis. KY, QH, and ZD wrote the manuscript. All authors read and approved the final manuscript.

#### **Acknowledgements:**

Not applicable.

## **References**

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Forner A, Reig M, Bruix J: **Hepatocellular carcinoma.** *Lancet* 2018, **391**(10127):1301-1314.
3. Benson AB, D'Angelica MI, Abbott DE, Abrams TA, Alberts SR, Anaya DA, Anders R, Are C, Brown D, Chang DT *et al*: **Guidelines Insights: Hepatobiliary Cancers, Version 2.2019.** *J Natl Compr Canc Netw* 2019, **17**(4):302-310.
4. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, Zhu AX, Murad MH, Marrero JA: **AASLD guidelines for the treatment of hepatocellular carcinoma.** *Hepatology* 2018, **67**(1):358-380.
5. Carmona R, Zakeri K, Green G, Hwang L, Gulaya S, Xu B, Verma R, Williamson CW, Triplett DP, Rose BS *et al*: **Improved Method to Stratify Elderly Patients With Cancer at Risk for Competing Events.** *J Clin Oncol* 2016, **34**(11):1270-1277.
6. Pintilie M: **Competing Risks: A Practical Perspective, 1st edn.** Chichester: John Wiley & Sons; 2006.
7. Zhang Z, Geskus RB, Kattan MW, Zhang H, Liu T: **Nomogram for survival analysis in the presence of competing risks.** *Ann Transl Med* 2017, **5**(20):403.
8. Fine JP, Gray RJ: **A Proportional Hazards Model for the Subdistribution of a Competing Risk.** *J Am Stat Assoc* 1999, **94**(446):496-509.
9. He C, Zhang Y, Cai Z, Lin X: **Competing risk analyses of overall survival and cancer-specific survival in patients with combined hepatocellular cholangiocarcinoma after surgery.** *BMC Cancer* 2019, **19**(1):178.
10. Wang Y, Wu J, He H, Ma H, Hu L, Wen J, Lyu J: **Correction to: Nomogram predicting cancer-specific mortality in early-onset rectal cancer: a competing risk analysis.** *Int J Colorectal Dis* 2020, **35**(6):1167-1168.

11. Wu X, Yu W, Petersen RH, Sheng H, Wang Y, Lv W, Hu J: **A competing risk nomogram predicting cause-specific mortality in patients with lung adenosquamous carcinoma.** *BMC Cancer* 2020, **20**(1):429.
12. Xu YB, Liu H, Cao QH, Ji JL, Dong RR, Xu D: **Evaluating overall survival and competing risks of survival in patients with early-stage breast cancer using a comprehensive nomogram.** *Cancer Med* 2020.
13. Gray RJ: **A class of K-sample tests for comparing the cumulative incidence of a competing risk.** *Ann Stat* 1988, **16**(2):1141-1154.
14. Huang XD, Zhou GQ, Lv JW, Zhou HQ, Zhong CW, Wu CF, Zheng ZQ, He XJ, Peng L, Ma J *et al*: **Competing risk nomograms for nasopharyngeal carcinoma in the intensity-modulated radiotherapy era: A big-data, intelligence platform-based analysis.** *Radiother Oncol* 2018, **129**(2):389-395.
15. Zhou H, Zhang Y, Qiu Z, Chen G, Hong S, Chen X, Zhang Z, Huang Y, Zhang L: **Nomogram to Predict Cause-Specific Mortality in Patients With Surgically Resected Stage I Non-Small-Cell Lung Cancer: A Competing Risk Analysis.** *Clin Lung Cancer* 2018, **19**(2):e195-e203.
16. Royston P, Altman DG: **External validation of a Cox prognostic model: principles and methods.** *BMC Med Res Methodol* 2013, **13**:33.
17. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP: **A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions.** *J Clin Epidemiol* 2013, **66**(6):648-653.
18. Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M: **Interpreting and comparing risks in the presence of competing events.** *BMJ* 2014, **349**:g5060.
19. Austin PC, Lee DS, Fine JP: **Introduction to the Analysis of Survival Data in the Presence of Competing Risks.** *Circulation* 2016, **133**(6):601-609.
20. Wolbers M, Koller MT, Witteman JC, Steyerberg EW: **Prognostic models with competing risks: methods and application to coronary risk prediction.** *Epidemiology* 2009, **20**(4):555-561.
21. **Comparative mortality of married and single persons.** *JAMA* 2013, **310**(3):326.
22. Kissane DW: **Marriage is as protective as chemotherapy in cancer care.** *J Clin Oncol* 2013, **31**(31):3852-3853.
23. Baine M, Sahak F, Lin C, Chakraborty S, Lyden E, Batra SK: **Marital status and survival in pancreatic cancer patients: a SEER based analysis.** *PLoS One* 2011, **6**(6):e21052.
24. DiMatteo MR: **Social support and patient adherence to medical treatment: a meta-analysis.** *Health Psychol* 2004, **23**(2):207-218.
25. Goldzweig G, Andritsch E, Hubert A, Brenner B, Walach N, Perry S, Baider L: **Psychological distress among male patients and male spouses: what do oncologists need to know?** *Ann Oncol* 2010, **21**(4):877-883.
26. Wu W, Fang D, Shi D, Bian X, Li L: **Effects of marital status on survival of hepatocellular carcinoma by race/ethnicity and gender.** *Cancer Manag Res* 2018, **10**:23-32.
27. Yan B, Bai DS, Qian JJ, Zhang C, Jin SJ, Jiang GQ: **Does marital status impact postoperative survival in patients with less differentiated hepatocellular carcinoma? A population-based study.** *Cancer Med* 2019, **8**(14):6272-6279.
28. Zhang W, Wang X, Huang R, Jin K, Zhangyuan G, Yu W, Yin Y, Wang H, Xu Z, Sun B: **Prognostic value of marital status on stage at diagnosis in hepatocellular carcinoma.** *Sci Rep* 2017, **7**:41695.

## Tables

**Table1.** Basic characteristics of patients with hepatocellular carcinoma.

Characteristics	Entire set	Training Set	Validation Set	P-value
		(n=24469)	(n=10488)	
<b>Age, years</b>				
≤60	16088 (46.0%)	11293 (46.2%)	4795 (45.7%)	0.475
61-65	6623 (18.9%)	4581 (18.7%)	2042 (19.5%)	
66-70	4409 (12.6%)	3104 (12.7%)	1305 (12.4%)	
71-75	3253 (9.3%)	2274 (9.3%)	979 (9.3%)	
76-80	2451 (7.0%)	1716 (7.0%)	735 (7.0%)	
81-85	1462 (4.2%)	1013 (4.1%)	449 (4.3%)	
≥86	671 (1.9%)	488 (2.0%)	183 (1.7%)	
<b>Sex</b>				
Female	23741 (67.9%)	16574 (67.7%)	7167 (68.3%)	0.352
Male	4683 (13.4%)	3274 (13.4%)	1409 (13.4%)	
Others	6533 (18.7%)	4621 (18.9%)	1912 (18.2%)	
<b>Race</b>				
White	7957 (22.8%)	5593 (22.9%)	2364 (22.5%)	0.526
Black	27000 (77.2%)	18876 (77.1%)	8124 (77.5%)	
<b>Marital status</b>				
Unmarried	15932 (45.6%)	11101 (45.4%)	4831 (46.1%)	0.237
Married	19025 (54.4%)	13368 (54.6%)	5657 (53.9%)	
<b>Surgical therapy</b>				
Resection	23533 (67.3%)	16521 (67.5%)	7012 (66.9%)	0.675
Local tumor destruction	4629 (13.2%)	3218 (13.2%)	1411 (13.5%)	
Transplant	4109 (11.8%)	2865 (11.7%)	1244 (11.9%)	
Chemotherapy	2686 (7.7%)	1865 (7.6%)	821 (7.8%)	
<b>Chemotherapy</b>				
Yes	18594 (53.2%)	12990 (53.1%)	5604 (53.4%)	0.562
No	16363 (46.8%)	11479 (46.9%)	4884 (46.6%)	
<b>Radiotherapy</b>				
Yes	32187 (92.1%)	22546 (92.1%)	9641 (91.9%)	0.505
No	2770 (7.9%)	1923 (7.9%)	847 (8.1%)	
<b>Tumor diameter, cm</b>				
≤5	11640 (33.3%)	8123 (33.2%)	3517 (33.5%)	0.162
6-7	9081 (26.0%)	6302 (25.8%)	2779 (26.5%)	
8-10	9707 (27.8%)	6877 (28.1%)	2830 (27.0%)	
>10	4529 (13.0%)	3167 (12.9%)	1362 (13.0%)	
<b>CC staging system</b>				
Stage I	14661 (41.9%)	10245 (41.9%)	4416 (42.1%)	0.925
Stage II	8323 (23.8%)	5835 (23.8%)	2488 (23.7%)	
Stage III	8115 (23.2%)	5673 (23.2%)	2442 (23.3%)	
Stage IV	3858 (11.0%)	2716 (11.1%)	1142 (10.9%)	

Abbreviations: AJCC, American Joint Committee on Cancer.

**Table2.** Overall survival rate and cumulative incidence of mortality in patients with hepatocellular carcinoma.

Characteristics	Overall survival rate (%) <sup>*</sup>		P-value	Cancer-specific mortality (%) <sup>#</sup>		P-value	Other causes-specific mortality (%) <sup>#</sup>		P-value
	3-year (95%CI)	5-year (95%CI)		3-year (95%CI)	5-year (95%CI)		3-year (95%CI)	5-year (95%CI)	
Age	33.5(32.9-34.1)	24.2(23.6-24.8)		56.6(56.0-57.3)	63.9(63.3-64.6)		9.8(9.5-10.2)	11.9(11.4-12.3)	
5 years			<0.001	53.7(52.7-54.6)	60.2(59.3-61.2)	<0.001	9.9(9.3-10.4)	11.9(11.2-12.5)	<0.001
65	36.5(35.6-37.4)	27.9(27.0-28.8)		53.1(51.5-54.6)	60.9(59.3-62.6)		9.9(8.9-10.8)	11.8(10.8-12.9)	
75	37.1(35.6-38.6)	27.2(25.7-28.8)		57.1(55.2-59.0)	65.6(63.7-67.5)		9.1(8.0-10.2)	10.8(9.6-12.0)	
85	33.8(32.0-35.7)	23.6(21.9-25.5)		61.4(59.3-63.5)	70.5(68.4-72.6)		7.8(6.7-9.0)	9.8(8.5-11.1)	
95	30.8(28.8-32.9)	19.7(17.9-21.7)		65.3(62.9-67.6)	73.7(71.3-76.0)		10.9(9.4-12.5)	13.1(11.4-14.8)	
105	23.8(21.8-26.1)	13.2(11.4-15.3)		71.1(68.1-74.0)	76.8(74.0-79.6)		12.7(10.6-14.8)	15.5(13.1-17.8)	
115	16.2(14.0-18.9)	7.7(6.0-10.0)		73.9(69.8-78.0)	78.9(74.9-82.8)		13.9(10.7-17.0)	16.3(12.8-19.8)	
125	12.3(9.5-15.9)	4.9(3.0-7.9)		53.7(52.7-54.6)	60.2(59.3-61.2)		9.9(9.3-10.4)	11.9(11.2-12.5)	
135			<0.001			<0.001			<0.001
145	32.6(31.9-33.4)	23.6(22.9-24.4)		57.0(56.2-57.8)	63.9(63.1-64.7)		10.4(9.9-10.9)	12.5(12.0-13.1)	
155	27.7(26.1-29.4)	17.6(16.2-19.2)		61.1(59.4-62.9)	69.1(67.3-70.8)		11.2(10.0-12.3)	13.3(12.0-14.6)	
165	40.8(39.3-42.3)	31.0(29.5-32.5)		52.3(50.8-53.8)	60.5(58.9-62.1)		6.9(6.2-7.7)	8.5(7.7-9.4)	
175			<0.001			<0.001			0.047
185	36.9(35.6-38.3)	27.7(26.3-29.0)		54.0(52.6-55.4)	61.3(59.9-62.8)		9.1(8.3-9.9)	11.0(10.1-11.9)	
195	32.5(31.8-33.2)	23.2(22.5-23.9)		57.4(56.7-58.2)	64.7(63.9-65.4)		10.1(9.6-10.5)	12.1(11.6-12.6)	
205			<0.001			<0.001			<0.001
215	28.8(27.9-29.7)	19.4(18.5-20.3)		60.0(59.1-61.0)	67.3(66.3-68.3)		11.2(10.6-11.8)	13.3(12.6-14.0)	
225	37.4(36.6-38.3)	28.1(27.3-29.0)		53.9(53.0-54.8)	61.2(60.3-62.1)		8.7(8.2-9.2)	10.7(10.1-11.3)	
235			<0.001			<0.001			<0.001
245	19.8(19.1-20.5)	11.9(11.3-12.5)		69.6(68.9-70.3)	76.0(75.3-76.7)		10.6(10.1-11.0)	12.1(11.5-12.6)	
255	48.7(46.9-50.6)	32.7(30.8-34.7)		40.0(38.2-41.8)	52.3(50.2-54.3)		11.3(10.1-12.4)	15.1(13.7-16.5)	
265	61.1(59.2-63.0)	47.1(45.1-49.3)		32.9(31.1-34.7)	44.4(42.3-46.5)		6.0(5.1-6.9)	8.5(7.3-9.6)	
275	82.9(81.2-84.7)	75.9(73.8-78.0)		9.7(8.3-11.1)	13.5(11.8-15.1)		7.4(6.1-8.6)	10.7(9.2-12.2)	
285			<0.001			0.356			<0.001
295	34.0(33.2-34.9)	25.2(24.4-26.1)		55.3(54.4-56.2)	61.9(61.0-62.8)		10.7(10.2-11.3)	12.9(12.2-13.5)	
305	33.0(32.1-33.9)	23.0(22.1-23.9)		58.2(57.2-59.1)	66.2(65.3-67.2)		8.9(8.3-9.4)	10.8(10.1-11.4)	
315			<0.001			<0.001			<0.001
325	34.5(33.9-35.2)	25.1(24.5-25.8)		55.4(54.7-56.1)	62.8(62.1-63.5)		10.0(9.6-10.4)	12.1(11.7-12.6)	
335	20.3(18.3-22.5)	12.0(10.2-14.2)		72.2(69.9-74.4)	79.0(76.8-81.3)		7.6(6.3-8.8)	9.0(7.5-10.4)	
345			<0.001			<0.001			<0.001
355	51.9(50.7-53.1)	39.9(38.7-41.2)		36.6(35.5-37.7)	45.4(44.1-46.6)		11.5(10.8-12.2)	14.7(13.9-15.6)	
365	35.3(34.1-36.6)	24.8(23.6-26.1)		53.3(52.0-54.6)	61.6(60.3-63.0)		11.4(10.6-12.2)	13.6(12.7-14.5)	
375	19.6(18.6-20.6)	12.5(11.7-13.5)		72.3(71.2-73.4)	78.1(77.0-79.2)		8.1(7.5-8.8)	9.4(8.6-10.1)	

; staging m	13.2(12.0- 14.5)	8.6(7.6- 9.7)	80.5(79.1- 81.9)	84.7(83.3- 86.0)	6.3(5.4- 7.1)	6.7(5.9- 7.6)
	<0.001		<0.001		<0.001	
	47.9(46.9- 49.0)	35.6(34.5- 36.7)	40.7(39.7- 41.7)	50.0(48.9- 51.1)	11.3(10.7- 12.0)	14.4(13.7- 15.2)
	41.0(39.6- 42.3)	30.0(28.7- 31.4)	47.8(46.5- 49.1)	56.4(55.0- 57.8)	11.2(10.4- 12.1)	13.6(12.6- 14.5)
	14.0(13.0- 14.9)	8.5(7.7- 9.3)	78.4(77.3- 79.5)	83.3(82.3- 84.4)	7.6(6.9- 8.3)	8.2(7.5- 9.0)
	4.3(3.6- 5.2)	2.2(1.6- 2.9)	89.8(88.6- 91.0)	91.5(90.4- 92.6)	5.9(5.0- 6.8)	6.3(5.4- 7.3)

Abbreviations: AJCC, American Joint Committee on Cancer.

\*: Overall survival rate was estimated by Kaplan-Meier method and tested by log-rank test.

#: Cancer-specific mortality and Other causes-specific mortality were estimated by cumulative incidence function and tested by Gray's test.

**Table 3.** Univariate analysis for OS and CSM.

Characteristics	OS*		CSM#	
	HR (95% CI)	P-value	sdHR (95% CI)	P-value
<b>Age, years</b>		Ref		Ref
<65	1.011(0.969-1.054)	0.616	0.991(0.948-1.04)	0.71
65-70	1.091(1.041-1.144)	<0.001	1.107(1.053-1.16)	<0.001
70-75	1.249(1.186-1.316)	<0.001	1.302(1.234-1.37)	<0.001
75-80	1.478(1.397-1.565)	<0.001	1.397(1.316-1.48)	<0.001
80-85	1.727(1.611-1.851)	<0.001	1.550(1.440-1.67)	<0.001
≥85	2.111(1.918-2.324)	<0.001	1.781(1.598-1.99)	<0.001
<b>Sex</b>		Ref		Ref
Male	1.184(1.134-1.235)	<0.001	1.16(1.104-1.211)	<0.001
Female	0.829(0.797-0.863)	<0.001	0.92(0.882-0.959)	<0.001
<b>Marital status</b>		Ref		Ref
Married	1.13(1.09-1.171)	<0.001	1.1(1.06-1.14)	<0.001
Unmarried				
<b>Surgical therapy</b>		Ref		Ref
Radical tumor destruction	0.438(0.418-0.46)	<0.001	0.443(0.422-0.465)	<0.001
Resection	0.327(0.31-0.345)	<0.001	0.378(0.357-0.399)	<0.001
Transplant	0.141(0.13-0.154)	<0.001	0.108(0.097-0.121)	<0.001
<b>Chemotherapy</b>		Ref		Ref
Yes	0.915(0.888-0.942)	<0.001	0.983(0.952-1.01)	0.29
No				
<b>Radiotherapy</b>		Ref		Ref
Yes	1.332(1.263-1.405)	<0.001	1.4(1.33-1.48)	<0.001
No				
<b>Tumor diameter, cm</b>		Ref		Ref
≤1	1.532(1.47-1.597)	<0.001	1.61(1.54-1.69)	<0.001
1-2	2.49(2.395-2.59)	<0.001	2.78(2.67-2.90)	<0.001
>2	3.428(3.269-3.594)	<0.001	3.87(3.67-4.08)	<0.001
<b>C staging system</b>		Ref		Ref
Stage I	1.153(1.107-1.2)	<0.001	1.19(1.14-1.24)	<0.001
Stage II	2.709(2.609-2.813)	<0.001	2.90(2.79-3.02)	<0.001
Stage III	4.632(4.418-4.857)	<0.001	4.71(4.46-4.97)	<0.001

Abbreviations: AJCC, American Joint Committee on Cancer; OS, overall survival, CSM, cancer-specific mortality; HR, hazard ratio; CI, confidence interval; sdHR, subdistribution hazard ratio; ref, reference.

\*: based on Cox proportional hazard regression analysis.

#: based on Fine and Gray's competing risk regression analysis.

**Table 4.** Multivariate analysis for OS and CSM.

Characteristics	OS*		CSM#	
	HR (95% CI)	P-value	sdHR (95% CI)	P-value
<b>Age, years</b>				
≤60		Ref		
61-65	0.991(0.95-1.033)	0.672	0.970(0.925-1.018)	0.22
66-70	1.088(1.037-1.141)	0.001	1.076(1.020-1.135)	0.0072
71-75	1.124(1.067-1.185)	<0.001	1.157(1.091-1.226)	<0.001
76-80	1.248(1.178-1.323)	<0.001	1.130(1.055-1.211)	<0.001
81-85	1.284(1.196-1.379)	<0.001	1.137(1.043-1.239)	0.0035
≥86	1.304(1.182-1.437)	<0.001	1.132(0.999-1.282)	0.052
<b>Sex</b>				
Female		Ref		
Male	1.039(0.995-1.085)	0.084	1.01(0.959-1.064)	0.7
<b>Race</b>				
White		Ref		
Black	0.827(0.794-0.861)	<0.001	0.909(0.869-0.951)	<0.001
<b>Marital status</b>				
Married		Ref		
Unmarried	0.908(0.88-0.936)	<0.001	0.972(0.938-1.007)	0.12
<b>Surgical therapy</b>				
No surgery		Ref		
Local tumor destruction	0.537(0.511-0.564)	<0.001	0.604(0.574-0.636)	<0.001
Resection	0.265(0.25-0.281)	<0.001	0.348(0.327-0.370)	<0.001
Transplant	0.177(0.162-0.193)	<0.001	0.157(0.139-0.176)	<0.001
<b>Chemotherapy</b>				
No		Ref		
Yes	0.622(0.603-0.642)	<0.001	0.745(0.718-0.772)	<0.001
<b>Radiotherapy</b>				
No		Ref		
Yes	0.775(0.734-0.818)	<0.001	0.862(0.814-0.912)	<0.001
<b>Tumor diameter, cm</b>				
≤5		Ref		
5.1-6	1.424(1.365-1.485)	<0.001	1.432(1.366-1.500)	<0.001
6.1-7	1.712(1.633-1.795)	<0.001	1.801(1.706-1.901)	<0.001
≥7.1	2.268(2.145-2.399)	<0.001	2.333(2.182-2.495)	<0.001
<b>CC staging system</b>				
CC1		Ref		
CC2	1.341(1.286-1.397)	<0.001	1.363(1.303-1.426)	<0.001
CC3	1.871(1.791-1.955)	<0.001	1.890(1.799-1.987)	<0.001
CC4	2.948(2.8-3.103)	<0.001	2.859(2.684-3.045)	<0.001
<b>Akaike information criterion</b>	320667.9		274678.6	

Abbreviations: AJCC, American Joint Committee on Cancer; OS, overall survival; CSM, cancer-specific mortality; HR, hazard ratio; CI, confidence interval; sdHR, subdistribution hazard ratio; ref, reference.

\*: based on Cox proportional hazard regression analysis.

#: based on Fine and Gray's competing risk regression analysis.

## Supporting Information

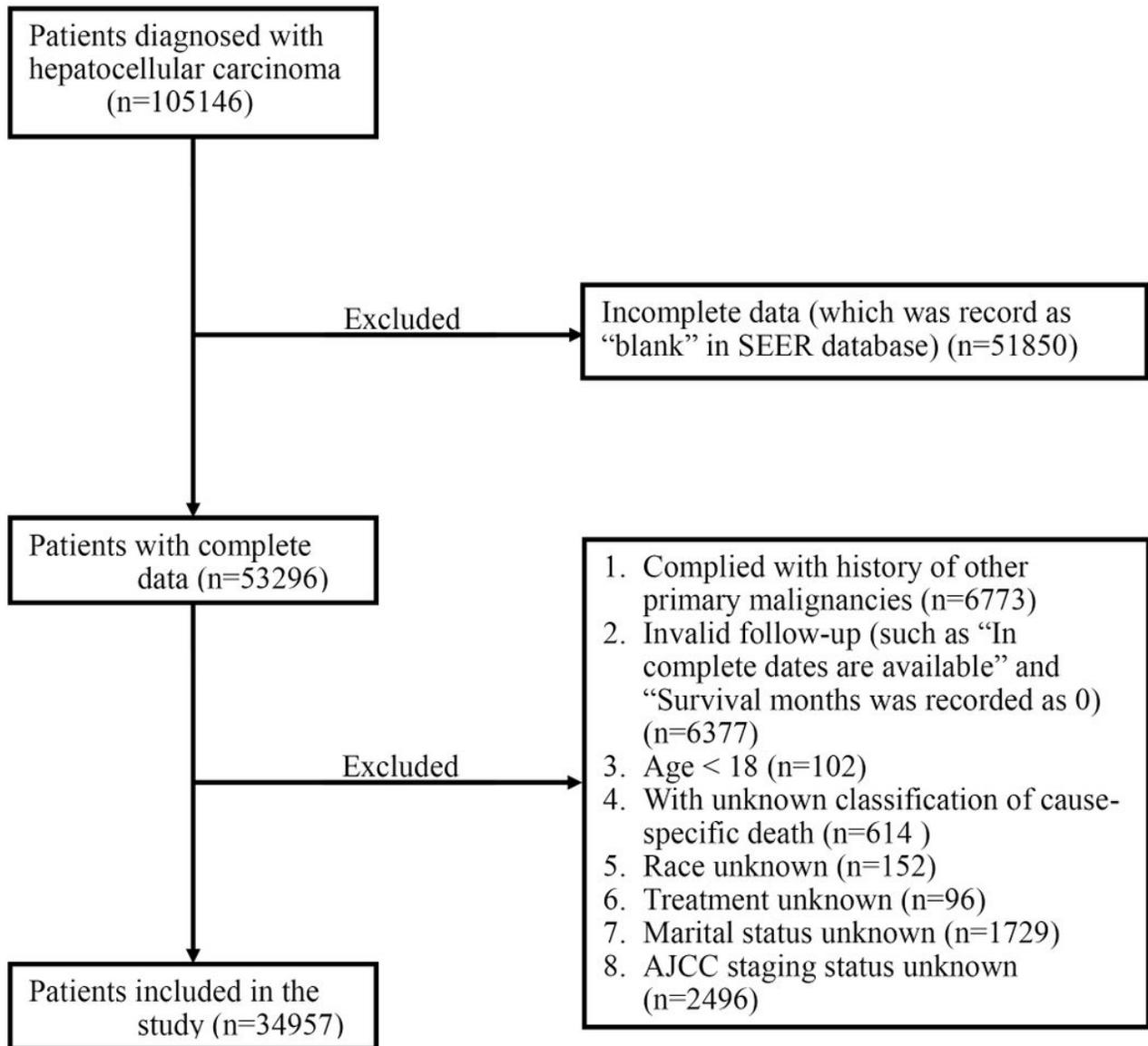
**Figure S1.** The calibration curves for predicting the 3-year CSM and OS in the training and validation sets. (A) CSM, training set; (B) CSM, validation set; (C) OS, training set; (D) OS, validation set.

CSM, cancer-specific mortality; OS, overall survival.

**Figure S2.** Difference of cumulative incidence of CSM predicted by the competing risk model and tradition Cox regression model. A. 3 year; B, 5 year. X-axis represent the patient, and the Y-axis means the difference of predicted cumulative incidence of CSM (Cox regression model minus Fine-Gray competing risk model)

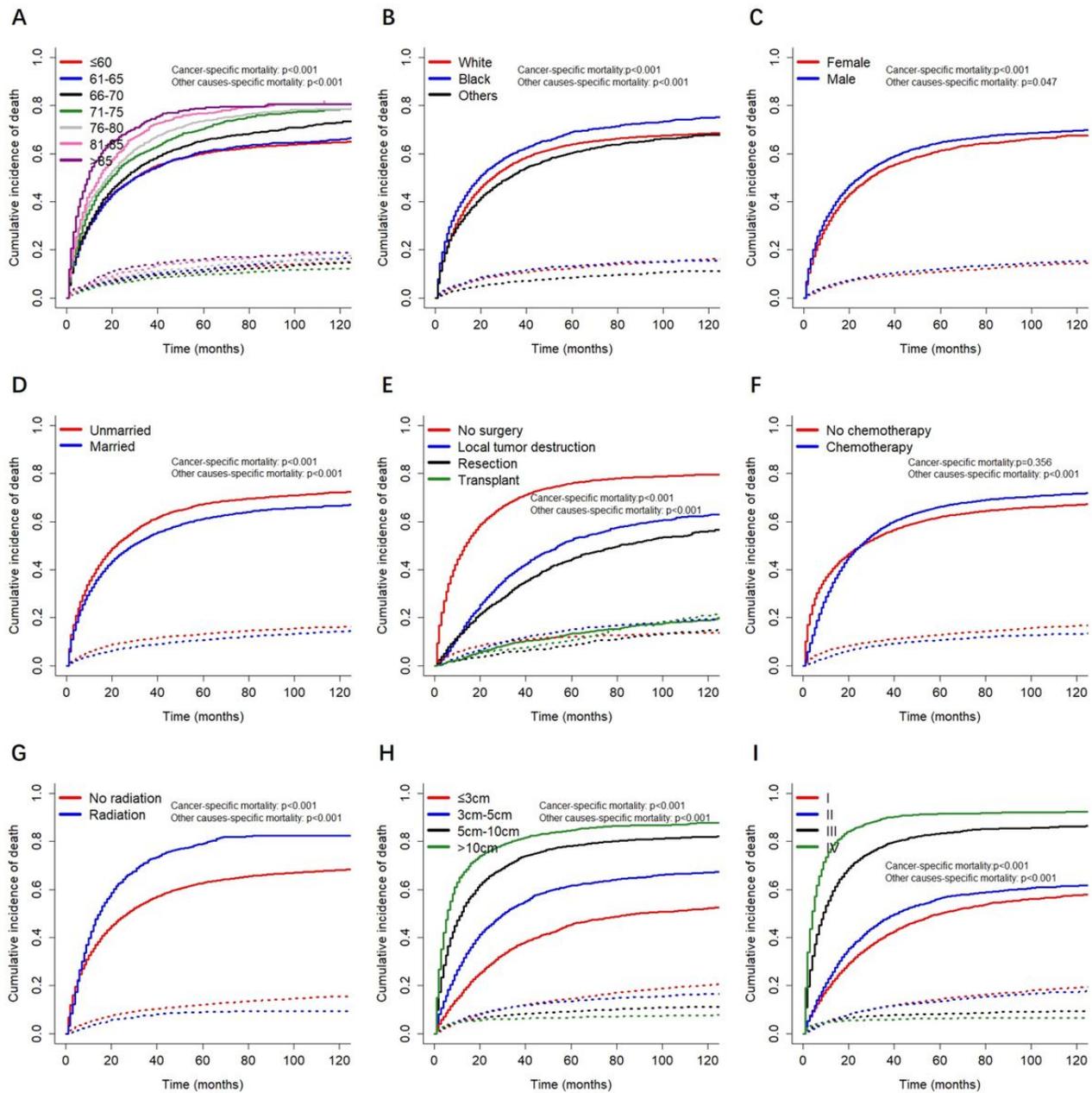
CSM, cancer-specific mortality.

## Figures



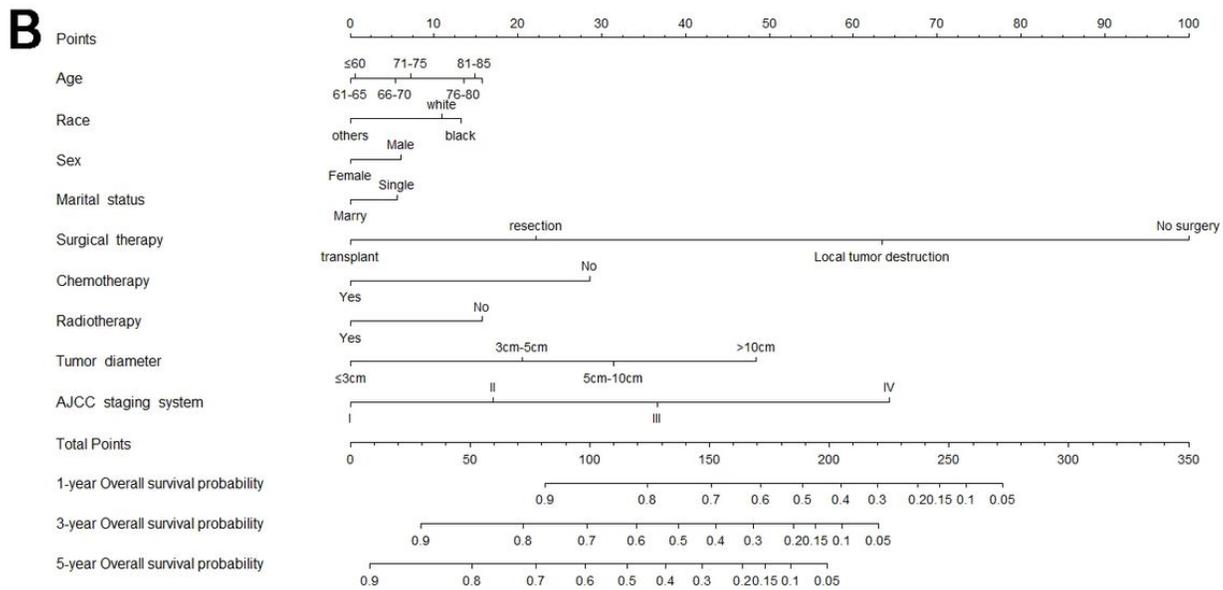
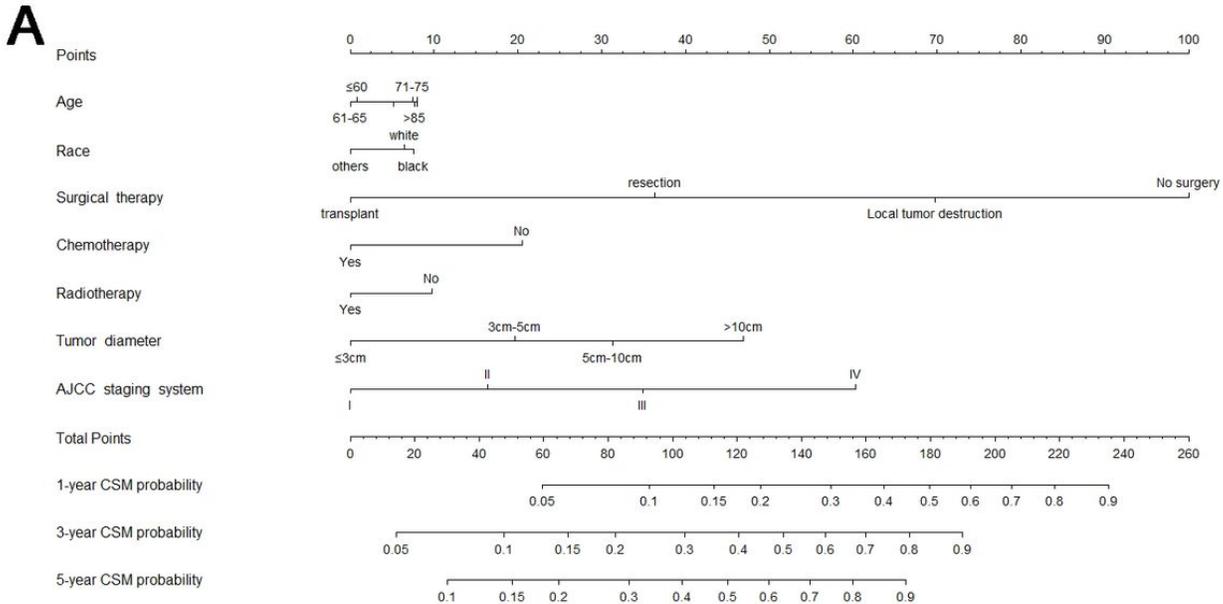
**Figure 1**

Flowchart for patient selection. SEER, Surveillance, Epidemiology, and End Results (SEER) program.



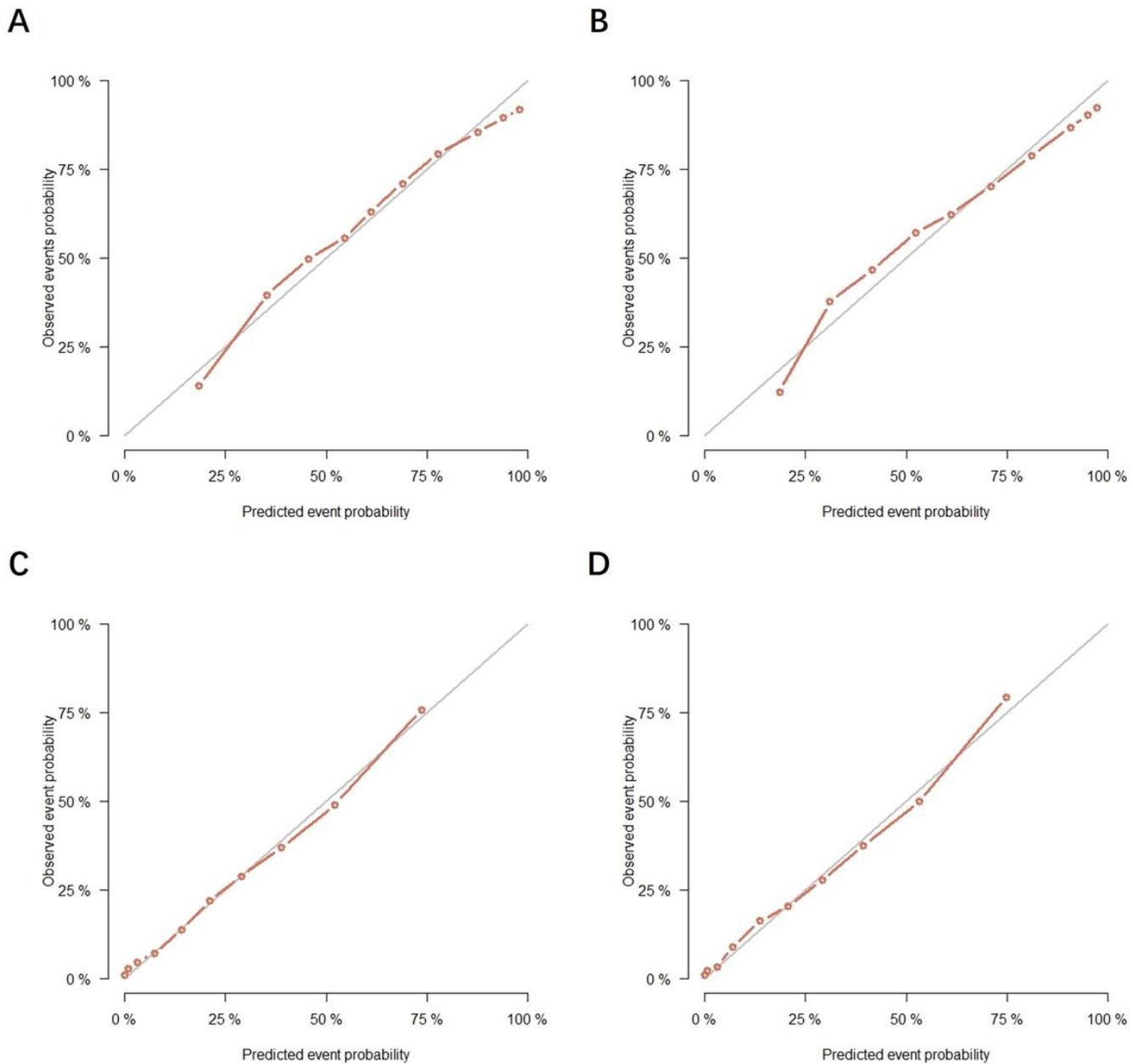
**Figure 2**

Cumulative incidence function for cancer-specific mortality and other causes-specific mortality according to patient characteristics (A) age; (B) race; (C) sex; (D) marital status; (E) surgical therapy; (F) chemotherapy; (G) radiotherapy; (H) tumour diameter; (I) AJCC staging system (Solid line: cancer-specific mortality; Dotted line: other causes-specific mortality). AJCC, American Joint Committee on Cancer.



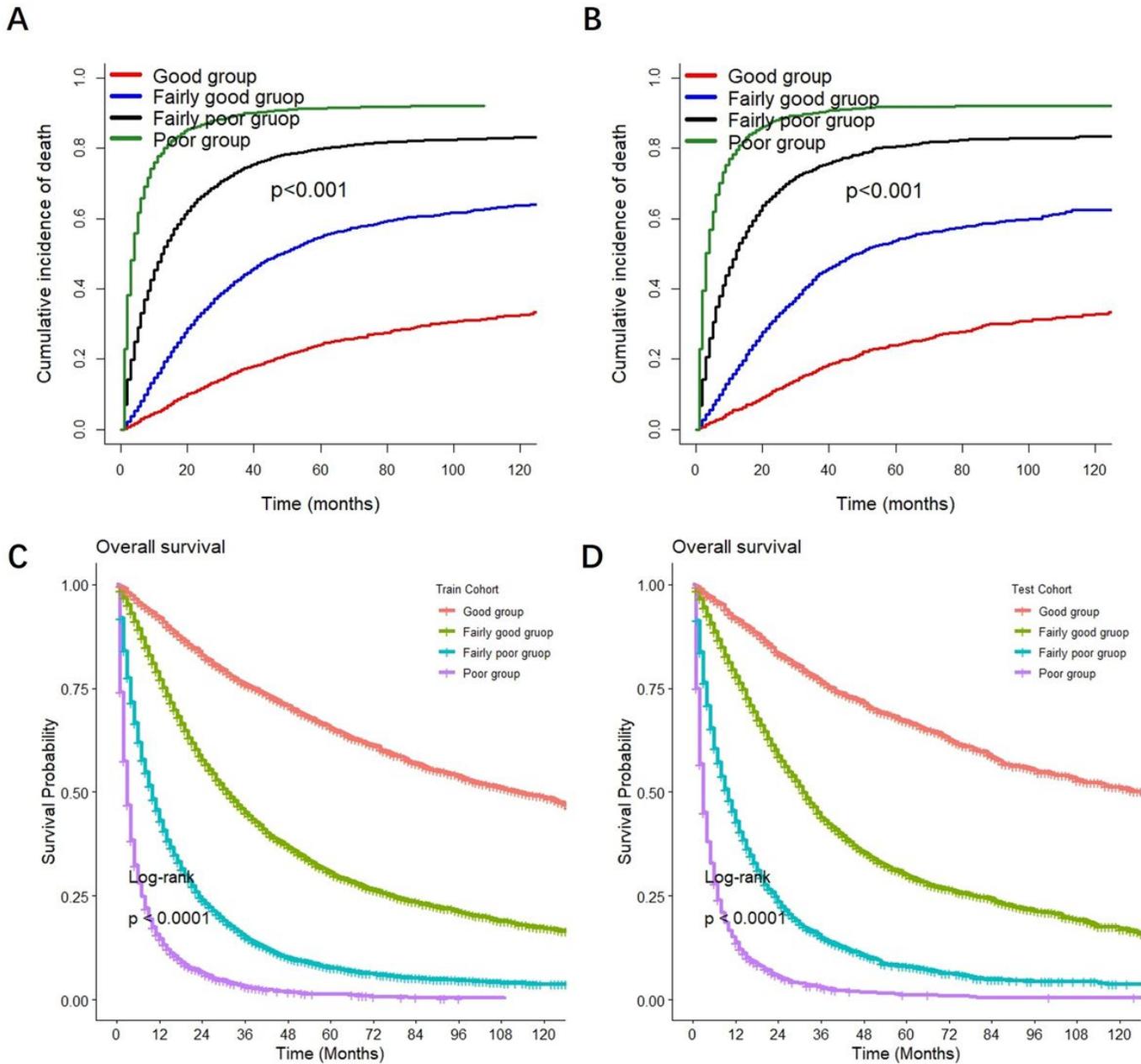
**Figure 3**

Nomograms for prediction of CSM and OS in patients with HCC. (A) CSM; (B) OS. CSM, cancer-specific mortality; OS, overall survival.



**Figure 4**

The calibration curves for predicting the 5-year CSM and OS in the training and validation sets. (A) CSM, training set; (B) CSM, validation set; (C) OS, training set; (D) OS, validation set. CSM, cancer-specific mortality; OS, overall survival.



**Figure 5**

Cumulative incidence function for CSM of risk groups defined by the nomogram model. (A) training set; (B) validation set; Kaplan-Meier plots for OS of risk groups defined by the nomogram model. (C) training set; (B) validation set. CSM, cancer-specific mortality; OS, overall survival.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [revisedSupplementaryinformation.docx](#)