

Identification of predictive factors in hepatocellular carcinoma outcome: A longitudinal study

HUIYUAN TIAN^{1*}, SHAOFENG CAO^{2*}, MINGXING HU³, YUZHU WANG³,
QIANG FU³, YANFENG PAN⁴ and TAO QIN³

¹Department of Research and Discipline Development, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, Henan 450003;

²Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450000; ³Department of Hepatobiliary and Pancreatic Surgery, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, Henan 450003;

⁴Department of Infectious Diseases, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450000, P.R. China

Received August 19, 2019; Accepted February 19, 2020

DOI: 10.3892/ol.2020.11581

Abstract. Various surgical methods impact the prognosis of patients with hepatocellular carcinoma (HCC) differently. However, clinical guidelines remain inconsistent and the relative importance of predictors of survival outcomes requires further evaluation. The present study aimed to rank the importance of predictive factors that impact the survival outcomes of patients with HCC and to compare the prognosis associated with different surgical methods based on data obtained from the Surveillance, Epidemiology and End Results database. To achieve these aims, the present study used a random forest (RF) model to detect important predictive factors associated with survival outcomes in patients with HCC. Cox regression analysis was used to compare different surgery methods. The variables included in the Cox regression model were selected based on the Gini index calculated by the RF model. Using the RF model, the present study demonstrated that surgery method, tumor size and age were the first, second and third most important factors associated with HCC prognosis,

respectively. Overall, patients who underwent local tumor destruction [hazard ratio (HR)=0.48; 95% confidence interval (CI), 0.45-0.51; P<0.001], wedge or segmental resection (HR, 0.31; 95% CI, 0.29-0.33; P<0.001), lobectomy (HR, 0.29, 95% CI, 0.27-0.31; P<0.001) or liver transplantation (HR, 0.16; 95% CI, 0.14-0.17; P<0.001) demonstrated improved overall survival time compared with those treated with surgery, with a gradual decreasing trend observed in HRs. The present study demonstrated that the surgical method used is the most important predictor of the survival outcomes of patients with HCC. Liver transplantation resulted in the best prognosis for patients with HCC, except for those with undifferentiated tumors or distant metastasis.

Introduction

Liver cancer is one of the most common types of cancers in the world, with ~841,000 newly diagnosed cases and 782,000 deaths globally per year, according to the global burden data in 2018. Hepatocellular carcinoma (HCC) constitutes 75-85% of liver cancer cases globally (1). In spite of the rapid development of clinical treatment methods in recent years, the prognosis of HCC patients remains dismal (2). Overall, the choice of surgical approach significantly affects prognosis and therefore, should be carefully considered (3).

In current clinical practice, the Barcelona Clinic Liver Cancer staging system (4) is recommended as the standard for surgical approach selection by both the European Association for the Study of the Liver and the American Association for the Study of Liver Disease (5,6). According to this standard, curative treatments are recommended as the optimal choice only for patients with very-early-stage and early-stage tumors (Barcelona Clinic Liver Cancer staging 0-A; solitary tumors or multinodular tumors with ≤3 nodules and size ≤3 cm with no vascular invasion or extrahepatic spread, Child-Turcotte-Pugh A or B, performance status 0) (7). Kutlu *et al* (8) demonstrated that radiofrequency ablation

Correspondence to: Dr Tao Qin, Department of Hepatobiliary and Pancreatic Surgery, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, 7 Weiwu Road, Zhengzhou, Henan 450003, P.R. China
E-mail: goodfreecn@163.com

Dr Yanfeng Pan, Department of Infectious Diseases, The First Affiliated Hospital of Zhengzhou University, 1 Jianshe East Road, Zhengzhou, Henan 450000, P.R. China
E-mail: panfirstfeng@163.com

*Contributed equally

Key words: carcinoma, hepatocellular, data mining, survival analysis

(RFA) is an appropriate method of treatment for patients with tumors measuring ≤ 30 mm, but that overall and cancer-specific survival (CSS) are worse for RFA compared with surgical resection or transplantation for tumors >30 mm. For patients with unresectable tumors, Shimose *et al* (9) suggested that transcatheter arterial chemoembolization (TACE) combined with RFA may prolong survival compared with TACE alone. However, some guidelines recommend surgical treatment for a broader spectrum of patients with HCC, such as the Asian Pacific Association for the Study of the Liver (10), the American Hepato-Pancreato-Biliary Association (11), the Korean Liver Cancer Study Group (12) and the Japan Society of Hepatology (13). In addition, Hyun *et al* (7) in 2019 performed a systematic review and reported that surgical treatment provides survival benefits in advanced-stage patients with HCC compared with chemoembolization. Due to these inconsistencies, further exploration is needed to establish the best surgical method for patients with HCC.

Previous studies have identified various factors that influence the prognosis of patients with HCC, such as age (14), tumor size (15,16), marital status (17), α -fetoprotein level (18), lymph node involvement, metastasis and co-infection with hepatitis B and C viruses (19), Epstein-Barr virus-induced gene 3 (20), serum interleukin-34 (IL-34) (21), however their relative importance remains unclear.

Random forest (RF) is a widely used classification machine learning method that does not require a prior hypothesis (22,23) and may provide another statistical option for researchers evaluating large datasets. RF has become a promising computational approach for determining patterns and associations based on high-dimensional datasets (24-26). The variable importance measure, a byproduct of the RF algorithm is calculated according to the predictive power of a variable and is often used to order the importance of variables, especially in genetics (27-29).

The Surveillance, Epidemiology and End Results (SEER) program collects data on cancer cases from various locations and sources throughout the USA, covering $\sim 28\%$ of the population (30). This dataset has been indicated to be valuable for predicting the prognosis of numerous types of malignant tumors, such as spinal ependymoma (31), breast carcinoma (32) and lung cancer (33).

In the present study, the RF model was used to identify the most important variables influencing the survival outcomes of patients with HCC. A longitudinal analysis was subsequently performed to determine the overall survival (OS) time and CSS outcomes of patients from SEER program with HCC (www.seer.cancer.gov) treated with different surgical methods, including no surgery, local tumor destruction, wedge or segmental resection, lobectomy and liver transplantation. The findings of the present study may provide clinicians and researchers with new evidence regarding HCC treatment selection.

Patients and methods

Population selection and data pre-processing. This cohort study was based on the newly released (1975-2016) Surveillance, Epidemiology and End Results (SEER) national database (www.seer.cancer.gov). Patients with an International

Classification of Diseases Oncology, 3rd Edition (34) code C220 and those with histology codes 8170-8175 were identified as cases with HCC. In total, 20,746 patients with HCC were identified who met the following inclusion criteria: (1) Diagnosed in 2004-2016, (2) diagnosed by histology methods and (3) aged ≥ 18 years. The exclusion criteria were no active follow-up data and missing values for surgery method, cancer-specific death, AFP, tumor size, historic stage and the T category according to the American Joint Committee on Cancer 6th edition (AJCC6_T) (<https://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/>). The detailed selection process is summarized in Fig. 1.

The data for surgery methods was categorized into no surgery (no surgery of the primary site), local tumor destruction (including photodynamic therapy, electrocautery, cryosurgery, laser, percutaneous ethanol injection, heat RFA, ultrasound and use of acetic acid), wedge or segmental resection, lobectomy (including right lobectomy, left lobectomy, lobectomy and local tumor destruction and extended lobectomy), and liver transplantation groups. Surgery of other sites (distant lymph nodes or other tissue(s)/organ(s) beyond the primary site) was divided into a surgery group and a non-surgery group. The year of diagnosis was divided into three time periods: 2004-2008, 2009-2012 and 2013-2016. The data for scope of regional lymph node surgery were divided into a non-surgical group and a surgical group. Data for marital status were dichotomously divided into a married group and an unmarried group, which included those who were divorced, separated, single, unmarried, had a domestic partner or were widowed. In the present study, data regarding insurance as uninsured or insured were recoded. The insured group included those with Medicaid, Indian or a public health service or private insurance, and those >65 years (as they were eligible for Medicare) (35).

Presentation of data and statistics. Missing values consisted of 2.38% of the total data extracted; missing continuous variables were replaced with the mean and missing categorical variables were recorded as unknown. In Table I, continuous data are presented as mean \pm SD and categorical data are presented as frequencies (%). One-way ANOVA, χ^2 test and the Kruskal-Wallis H test were used to compare differences among the treatment groups for continuous, categorical and ordinal variables, respectively. All the post hoc tests were adjusted using Bonferroni's method. All the statistical analyses were performed using the R version 3.4.3 (<https://cran.r-project.org/>). Two-sided P-values <0.05 were considered to indicate a statistically significant difference.

Variable selection and importance ranking using RF analysis. For the generation of the RF model, the data included in the present study was first trained using general demographic information [(age, insurance, education level (36), family income (36), race (37), marital status (17), sex (38), region)] and factors that may be associated with survival outcomes according to previous studies and clinical practice as follows: Surgery method (39-41), tumor size (15), diagnosis year (42), historic stage, grade (43), AJCC_Tumor (T), AJCC_Node (N), AJCC_Metastasis (M) (44), scope of regional lymph node surgery, AFP (18), surgery of other sites, number of benign

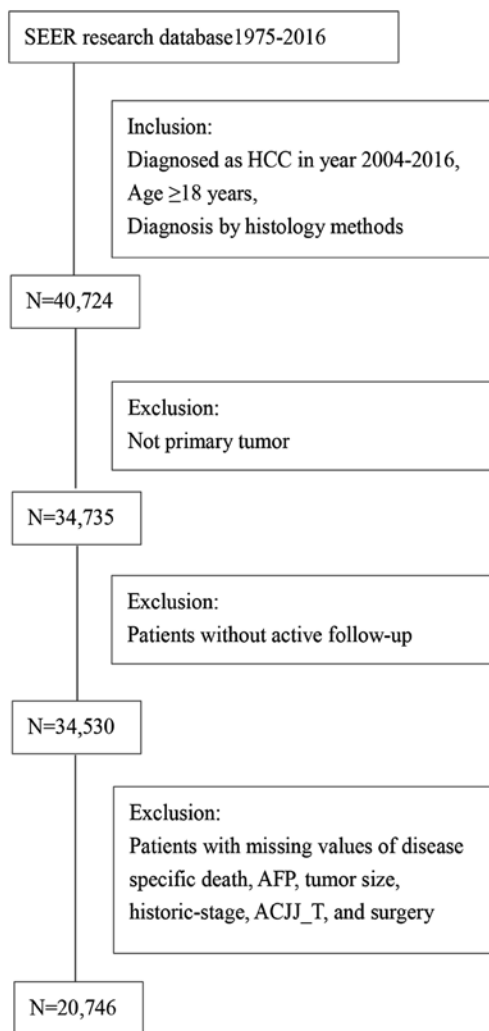


Figure 1. Patient selection criteria. A total of 20,746 patients with HCC were evaluated from the SEER national database (www.seer.cancer.gov) according to the exclusion and inclusion criteria outlined in the figure. HCC, hepatocellular carcinoma; SEER, Surveillance, Epidemiology and End Results; AFP, α fetoprotein; AJCC_T, the T category according to the American Joint Committee on Cancer 6th edition.

or borderline tumors and number of malignant tumors (45). Fibrosis score, radiation and chemotherapy were initially considered in the model, but were later excluded to preserve the power of the test as missing values accounted for over half of the population.

The RF model was employed to rank the importance of the predictors of HCC survival outcomes. This version of the RF model was performed in the R package ‘randomForest’ (v. 4.6-14) (<https://www.stat.berkeley.edu/~breiman/RandomForests/>). All data were randomly split into a training set and a validation set by a ratio of 6:4. The importance rankings of predictors of 2-year survival were obtained as a by-product of the RF model. The association between the top 10 most important predictors and 2-year survival outcome was analyzed using the logistic regression method.

Comparison of surgical approaches using Cox regression analysis. Univariate Cox regression analysis was used to compare the overall survival (OS) time and CSS of patients

with HCC treated with different surgical methods. To explore whether the association between surgical method and survival outcome is modified by other predictors, multivariate models were constructed using the 10 variables with the highest Gini index, which is a tool describing the relative contribution of each feature in predicting the outcome (46). In addition, analyses stratified by tumor size (patients were divided into two groups by median tumor size, cut-off value 54 mm), age (patients were divided into two groups by median age, cut-off value 62 years), histological stage and grade were conducted to minimize the effect of confounding factors.

Results

Predictive factors and comparison of surgical methods. A total of 20,746 patients diagnosed with HCC from 2004-2016 were included in the present study. The 10 most important factors influencing survival outcomes were surgery method, tumor size, age, AJCC_T, family income, education level, historic stage, grade, AJCC_M, and diagnosis year (Fig. 2). Overall, local tumor destruction [(hazard ratio (HR)=0.48; 95% confidence interval (CI), 0.45-0.51; $P<0.001$)], wedge or segmental resection (HR, 0.31; 95% CI, 0.29-0.33; $P<0.001$), lobectomy (HR=0.29; 95% CI, 0.27-0.31; $P<0.001$) and liver transplantation (HR=0.16; 95% CI, 0.14-0.17; $P<0.001$) demonstrated improved OS outcomes compared with no surgery, except for undifferentiated tumors and those with distant metastasis, with the HRs demonstrating a decreasing trend (Tables III, SI and SIV).

Demographic and clinical characteristics of patients with HCC. Of the total patients with HCC, 2,568 had undergone liver transplantation, 1,574 lobectomy, 2,103 wedge or segmental resection and 2,466 local tumor destruction. A total of 12,035 patients had no surgery. All the 10 most important predictors were distributed unevenly among the 5 treatment groups (all P -values <0.001). Post hoc analyses using Bonferroni's method demonstrated significant difference between the no surgery group and all 4 surgery groups (all P -values <0.05). These detailed demographic and clinical characteristics of the patients are presented in Table I.

Ranking of important predictors association with survival outcomes of patients with HCC based on the RF model. To compare the relative importance of predictive factors, an RF model using all predictors was first trained. This 21-predictor RF model achieved an accuracy of 78.05%, with $n_{tree}=1,400$ (Number of trees to grow) and $m_{try}=3$ (Number of variables randomly sampled as candidates at each split). Based on the Gini index, the 10 most important predictors associated with the survival outcomes of patients with HCC were obtained. These were surgery method, tumor size, age, AJCC_T, family income, education level, historic stage, grade, AJCC_M and diagnosis year. The importance of the ranking results are presented in Fig. 2. Next, 10 factors with the highest Gini were included in the subsequent analysis. The association between predictors and 2-year OS were further analyzed using the logistic regression method. The results indicated that more radical surgery methods, for example liver transplantation and lobectomy, higher family

Table I. Demographic and clinical characteristics of patients with hepatocellular carcinoma by categories of surgery methods.

Characteristics	No surgery	Local tumor destruction	Wedge or segmental resection	Lobectomy	Liver transplantation	Statistics	P-value
Diagnosis year, n (%)						Kruskal-Wallis $\chi^2=188.0$	<0.001
2004-2008	3,531 (29.34)	923 (37.43)	575 (27.34)	519 (32.97)	1008 (39.25)		
2009-2012	3,929 (32.65)	722 (29.28)	663 (31.53)	517 (32.85)	855 (33.29)		
2013-2015	4,575 (38.01)	821 (33.29)	865 (41.13)	538 (34.18)	705 (27.45)		
Age, mean \pm SD ^a	64.17 \pm 11.32	63.19 \pm 10.00	62.06 \pm 11.08	61.52 \pm 12.64	57.64 \pm 7.63	F=201.2	<0.001
Education level ^b , mean \pm SD	14.89 \pm 5.94	13.81 \pm 5.50	14.34 \pm 5.54	14.21 \pm 5.73	14.08 \pm 5.83	F=26.3	<0.001
Family income ^c , mean \pm SD	75.76 \pm 19.27	80.89 \pm 19.43	80.80 \pm 20.60	82.27 \pm 20.85	77.49 \pm 20.20	F=79.8	<0.001
Tumor size, mm, mean \pm SD	72.09 \pm 46.08	41.23 \pm 27.78	53.85 \pm 41.50	82.78 \pm 54.40	33.36 \pm 24.49	F=724.4	<0.001
Histological stage, n (%)						Kruskal-Wallis $\chi^2=3,269.7$	<0.001
Localized	4,899 (40.71)	1,917 (77.74)	1,720 (81.79)	1,142 (72.55)	2,022 (78.74)		
Regional	4,250 (35.31)	482 (19.55)	333 (15.83)	356 (22.62)	528 (20.56)		
Distant	2,886 (23.98)	67 (2.72)	50 (2.38)	76 (4.83)	18 (0.70)		
AJCC_T n (%) ^e						$\chi^2=2,639.3$	<0.001
T1	3,764 (31.28)	1,372 (55.64)	1,130 (53.73)	664 (42.19)	1,216 (47.35)		
T2	2,182 (18.13)	632 (25.63)	494 (23.49)	350 (22.24)	1,077 (41.94)		
T3	4,097 (34.04)	227 (9.21)	202 (9.61)	339 (21.54)	129 (5.02)		
T4	603 (5.01)	24 (0.97)	59 (2.81)	87 (5.23)	17 (0.66)		
TX	1,389 (11.54)	211 (8.56)	218 (10.37)	134 (8.51)	129 (5.02)		
AJCC_N n (%) ^f						$\chi^2=1,082.1$	<0.001
N0	8,566 (71.18)	2,134 (86.54)	1831 (87.07)	1,373 (87.23)	2,384 (92.83)		
N1	1,225 (10.18)	48 (1.95)	22 (1.05)	45 (2.86)	20 (0.78)		
NX	2,244 (18.65)	284 (11.52)	250 (11.89)	156 (9.91)	164 (6.39)		
AJCC_M n (%) ^g						$\chi^2=1,897.8$	<0.001
M0	7,946 (66.02)	2,179 (88.36)	1,834 (87.21)	1,379 (87.61)	2,414 (94.00)		
M1	2,425 (20.15)	57 (2.31)	39 (1.85)	50 (3.18)	16 (0.62)		
MX	1,664 (13.83)	230 (9.33)	230 (10.94)	145 (9.21)	138 (5.37)		
Grade, n (%) ^d						$\chi^2=2,058.9$	<0.001
I	2,299 (19.10)	624 (25.30)	396 (18.83)	246 (15.63)	557 (21.69)		
II	2,785 (23.14)	736 (29.85)	1,019 (48.45)	773 (49.11)	1,008 (39.25)		
III	1,598 (13.28)	199 (8.07)	446 (21.21)	370 (23.51)	220 (8.57)		
IV	136 (1.13)	9 (0.36)	31 (1.47)	46 (2.92)	14 (0.55)		
Unknown	5,217 (43.35)	898 (36.42)	211 (10.03)	139 (8.83)	769 (29.95)		

^aPatients age at diagnosis; ^bPercentages of county populations ($\geq 25\%$) with less than a high school education between 2013-2017; ^cMedian family income values by county, income is displayed as dollars in thousands; ^dGrade I, well differentiated; Grade II, moderately differentiated; Grade III, poorly differentiated; Grade IV, undifferentiated. ^eAJCC_T, the T category according to the American Joint Committee on Cancer 6th edition; ^fAJCC_N, the N category according to the American Joint Committee on Cancer 6th edition; ^gAJCC_M, the M category according to the American Joint Committee on Cancer 6th edition.

income and more recent diagnosis time, were protective predictive factors and that larger tumor size, older age, lower education level, metastasis, higher stage of AJCC_T and AJCC_M, less differentiated tissue were negative predictive factors (Table II).

Cox regression analysis of survival outcomes for surgery methods among patients with HCC. In univariate Cox regression analysis, patients undergoing local tumor destruction (HR, 0.34; 95% CI, 0.32-0.36; $P<0.001$), wedge or segmental resection (HR, 0.23; 95% CI, 0.22-0.25; $P<0.001$), lobectomy (HR,

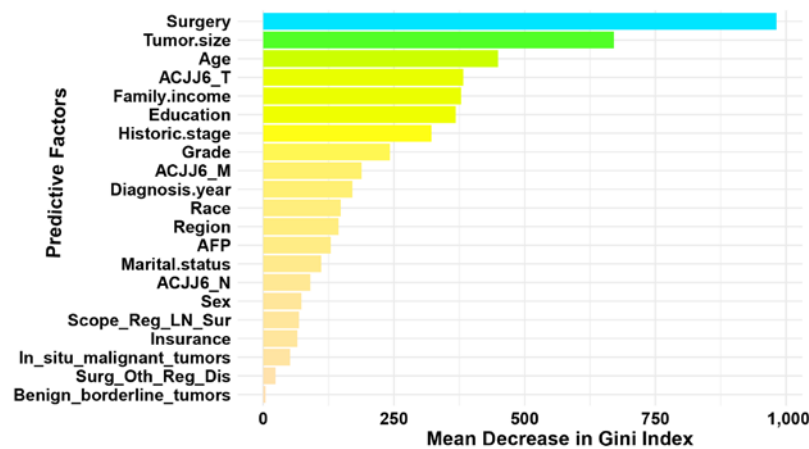


Figure 2. Order of importance of predictive factors associated with the survival outcome of patients with HCC. The relative importance was ranked according to Gini index, which is a tool describing the relative contribution of each feature in predicting the outcome. Surgery, surgery of the primary site; tumor size, the largest dimension or diameter of the primary tumor; age, patient age at diagnosis; marital status, patients' marital status at diagnosis; insurance, if patients had medical insurance at diagnosis; family income, median family income values by county; education, percentages of county populations (≥ 25 years) with less than a high school education between 2011-2015; historic stage, localized, regional and distant; grade, the degree of cell differentiation; diagnosis year, year of diagnosis; region, groups of countries at diagnosis; AFP, alpha fetoprotein; AJCC, the American Joint Committee on Cancer; T, tumor; N, node; M, metastasis; AJCC6_T, the T category according to the American Joint Committee on Cancer 6th edition; AJCC6_M, the M category according to the AJCC 6th edition; AJCC6_N, the N category according to the AJCC 6th edition; scope_reg_ln_sur, scope of regional lymph node surgery; *in situ_malignant_tumor*, total number of malignant tumors in patients; *surg_oth_reg_dis*, the surgical removal of distant lymph nodes or other tissues or organs beyond the primary site; *benign_borderline_tumor*, total number of benign or borderline tumors; HCC, hepatocellular carcinoma.

0.27; 95% CI, 0.25-0.29; $P < 0.001$) and liver transplantation (HR, 0.11; 95% CI, 0.10-0.12; $P < 0.001$) had improved overall survival time compared with patients not undergoing surgery (Table III). Following adjustment for confounding factors in the multivariate analysis, this trend was weakened but remained significant. In multivariate analyses, patients undergoing local tumor destruction (HR, 0.48; 95% CI, 0.45-0.51; $P < 0.001$), wedge or segmental resection (HR, 0.31; 95% CI, 0.29-0.33; $P < 0.001$), lobectomy (HR, 0.29; 95% CI, 0.27-0.31; $P < 0.001$), and liver transplantation (HR, 0.16; 95% CI, 0.14-0.17; $P < 0.001$) had improved OS outcomes compared with patients not undergoing surgery (Table III). Following adjustment, the HRs of local tumor destruction, wedge or segmental resection, lobectomy and liver transplantation demonstrated a decreasing trend, but there was no significant difference observed for wedge or segmental resection and lobectomy (Table III). CSS analysis demonstrated a similar trend to OS for both univariate and multivariate analyses. Survival curves showed that patients undergoing liver transplantation had the longest survival time, and patients who did not undergo surgery had the shortest survival time compared with surgical methods ($P < 0.001$, Fig. 3).

Stratified analysis. To minimize the effect of confounding factors, stratified analyses were conducted based on tumor size (median tumor size, < 54 mm or ≥ 54 mm), age (median age, < 62 years or ≥ 62 years), historic stage (localized, regional or distant), and grade (grades I-IV). The stratified survival curves are presented in Figs. S1-S4 and detailed information is provided in Tables SI-SIV. Altogether, the results demonstrated that the protective trend remained following stratification. However, there were two exceptions: In patients with distant metastasis and patients with undifferentiated tumors, no significant difference was observed for the four surgery types following stratification.

Discussion

The present longitudinal study included 20,746 patients with HCC from the SEER database and a RF model was used to predict 2-year OS and CSS outcomes. Firstly, the relative importance of predictive factors was evaluated. Subsequently, the factor that was ranked most important, surgery method was further analyzed by Cox regression analysis. The no surgery group, local tumor destruction group, wedge or segmental resection group, and lobectomy group demonstrated improved OS and CSS outcomes compared with liver transplantation group, with the HRs exhibiting a gradually decreasing trend overall. This result remained stable following stratification by tumor size, age, historic stage and grade, except for undifferentiated tumors and those with distant metastasis.

RF models are popular as they have high prediction accuracy and provide information on the relative importance of variables for classification (47). In recent years, the RF model has been commonly applied for investigating the quantitative importance of predictors for different cancers like pancreatic neuroendocrine tumors (25) and colorectal carcinoma (48); however, studies using this model on patients with HCC are very limited. Choi *et al* (23) constructed a prognostic model estimating the outcomes of 480 advanced-stage patients with HCC, all of whom were treated with sorafenib. Similarly, Kawaguchi *et al* (49) evaluated the prognosis of 247 patients with non-alcoholic fatty liver disease-HCC in Japan, indicating treatment and serum albumin level to be the two most important distinguishing factors. Our results suggested treatment was the most important prognostic factor, which was consistent with these studies. The present study had a large sample size and therefore provided a stable result.

The 10 most important predictors found in the present study were surgery method (top of the list), tumor size, age, AJCC_T, family income, education level, historic stage, grade,

Table II. Logistic regression analysis of 2-year survival and predictors in patients with hepatocellular carcinoma.

Predictive factor	Univariate analysis			Multivariate analysis ^a		
	OR	95% CI	P-value	OR	95% CI	P-value
Surgery method						
No surgery	Reference			Reference		<0.001
Local tumor destruction	0.18	0.16-0.20	<0.001	0.29	0.26-0.32	<0.001
Wedge or segmental resection	0.10	0.09-0.12	<0.001	0.15	0.13-0.16	<0.001
Lobectomy	0.15	0.13-0.16	<0.001	0.13	0.11-0.15	<0.001
Liver transplantation	0.05	0.05-0.06	<0.001	0.08	0.07-0.10	<0.001
Tumor size, mm	1.02	1.02-1.02	<0.001	1.01	1.01-1.01	<0.001
Age, years ^a	1.02	1.02-1.02	<0.001	1.01	1.01-1.02	<0.001
Education level	1.01	1.01-1.02	<0.001	0.98	0.98-0.99	<0.001
Family income	0.99	0.99-0.99	<0.001	0.99	0.99-0.99	<0.001
Historic stage						
Localized	Reference			Reference		
Regional	5.78	5.34-6.28	<0.001	2.32	1.97-2.73	<0.001
Distant	1.04	0.98-1.11	0.202	1.09	0.98-1.22	0.105
AJCC_T						
T1	Reference			Reference		
T2	1.17	1.09-1.26	<0.001	1.21	1.10-1.33	<0.001
T3	5.50	5.06-5.97	<0.001	1.77	1.58-1.98	<0.001
T4	7.00	5.78-8.54	<0.001	1.66	1.31-2.12	<0.001
Diagnosis year						
2004-2008	Reference			Reference		
2009-2012	0.88	0.82-0.94	<0.001	0.72	0.66-0.79	<0.001
2013-2015	0.57	0.54-0.61	<0.001	0.48	0.44-0.53	<0.001
Grade						
I	Reference			Reference		
II	0.97	0.90-1.06	<0.001	1.30	1.18-1.43	<0.001
III	2.22	2.01-2.45	<0.001	2.44	2.15-2.76	<0.001
IV	3.34	2.50-4.51	<0.001	3.68	2.59-5.29	<0.001
AJCC_M	2.04	1.89-2.21	<0.001			
M0	Reference			Reference		
M1	11.69	10.14-13.55	<0.001	1.74	1.33-2.27	<0.001

^aAge was analyzed as a continuous variable. OR, odd ratio; CI, confidence interval; T, Tumor; N, Node; M, Metastasis; AJCC, the American Joint Committee on Cancer.

AJCC_M and diagnosis year. Previous studies have suggested that RFA along with TACE prolongs the survival of patients with HCC (9) and that chemotherapy may achieve favorable results in patients with advanced HCC (50). However, in the present study, data on RFA therapy and chemotherapy was lacking for the majority of the study population; therefore, these methods were not investigated. No surgery and four surgery methods were discussed in the present study, including tumor destruction, wedge or segmental resection, lobectomy and transplantation. The findings of the present study demonstrated that liver transplantation was the first choice for all patients with HCC, except for those with undifferentiated tumors or distant metastasis. This result is consistent with previous studies, which have suggested that surgical resection

offers a significant survival benefit over thermal ablation (41) and transarterial chemoembolization (2). Another study based on the SEER database (2013.5-2014.1) reported similar trends; local tumor destruction, partial surgery and total surgery compared with no surgery were all significant positive prognostic factors (3). The present study included more recently released data for participants, analyzed more treatment methods and applied more reliable variable selection criteria compared with the aforementioned study providing clinicians and researchers with more evidence. Other important outcome predictors, such as tumor size, age, AJCC_T, family income, education level, historic stage, grade, AJCC_M and diagnosis year, were consistent with previous studies. The present study is novel as it identified predictors according to their relative

Table III. Cox regression and competing risk analysis for association of survival outcomes with surgery types in patients with HCC.

A, Univariate analysis

Variable	Overall survival			Cancer-specific survival		
	HR	95% CI	P-value	HR	95% CI	P-value
No surgery	Reference					
Local tumor destruction	0.34	0.32-0.36	<0.001	0.31	0.29-0.33	<0.001
Wedge or segmental resection	0.23	0.22-0.25	<0.001	0.22	0.20-0.24	<0.001
Lobectomy	0.27	0.25-0.29	<0.001	0.26	0.24-0.28	<0.001
Liver transplantation	0.11	0.10-0.12	<0.001	0.07	0.07-0.08	<0.001

B, Multivariate analysis^a

Variable	Overall survival			Cancer-specific survival		
	HR	95% CI	P-value	HR	95% CI	P-value
No surgery	Reference					
Local tumor destruction	0.48	0.45-0.51	<0.001	0.46	0.43-0.49	<0.001
Wedge or segmental resection	0.31	0.29-0.33	<0.001	0.30	0.28-0.32	<0.001
Lobectomy	0.29	0.27-0.31	<0.001	0.28	0.26-0.31	<0.001
Liver transplantation	0.16	0.14-0.17	<0.001	0.11	0.10-0.12	<0.001

^aThe 10 most important predictors were included in the multivariate models. HCC, hepatocellular carcinoma; HR, hazard ratio; CI, confidence interval; T, Tumor; N, Node; M, Metastasis; AJCC, the American Joint Committee on Cancer.

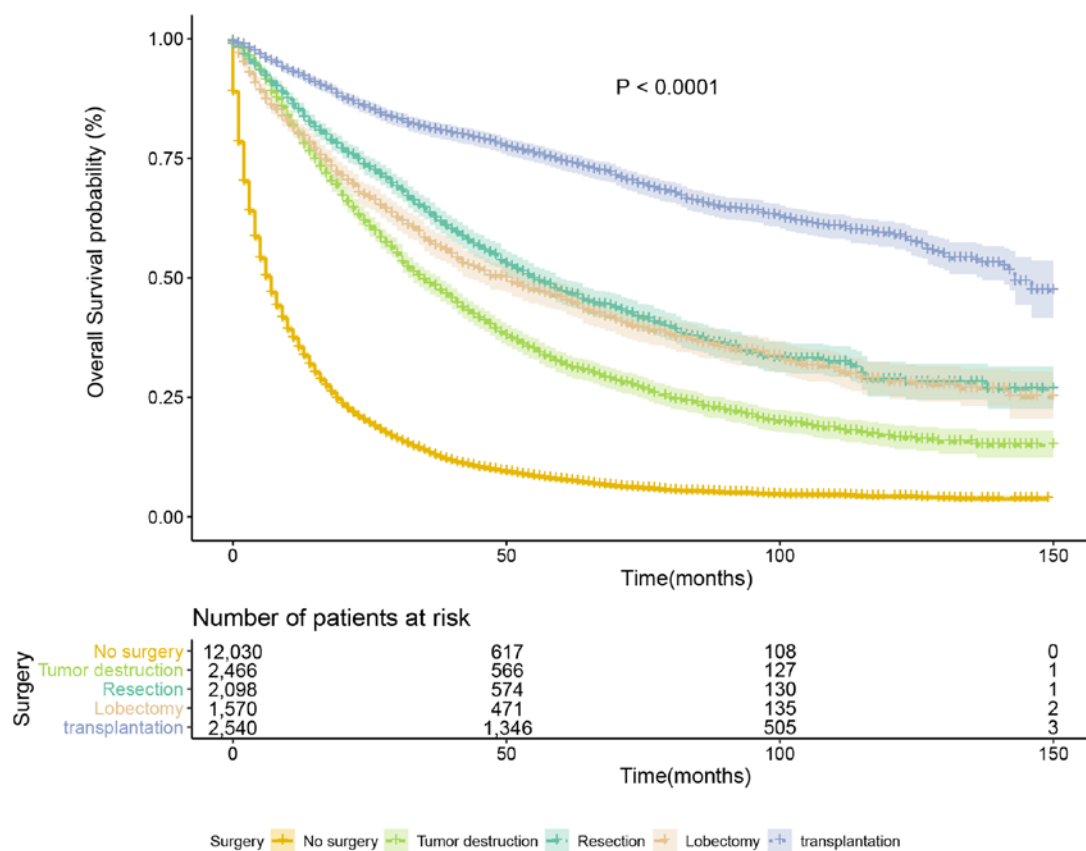


Figure 3. Survival curves of 5 surgery methods (no surgery, local tumor destruction, wedge or segmental resection, lobectomy and liver transplantation) and number of patients at risk at different survival time (P-value represented the significance of the difference among survival curves of the 5 surgery methods).

importance. Due to the limitations of the cohort studied, some important factors could not be included in the present study, such as genetic factors and cytokines, which are reported to be significantly associated with survival outcomes. For instance, EBI3 is suggested to be a cancer suppressor (20) and IL-34 may be associated with survival outcomes by regulating tumor growth and hepatic fibrosis in patients with HCC (21). Future prospective studies are required in order to investigate the relative importance of these factors.

The present study had several strengths. The RF model was used to rank the importance of variables associated with the survival outcomes of patients with HCC, providing new evidence to currently limited research. To date, studies utilizing the RF method to rank the importance of predictive clinical variables for patients with HCC are very limited (23,49). The surgical methods included in the present study are broader and more detailed compared with those in previous studies. Besides, SEER collects data with a wide temporal and spatial range and the time of diagnosis was adjusted by multifactorial regression analysis considering that the treatment modality varies year by year. In total, 20,746 patients with HCC were included in the analysis, making the models very reliable. However, there are several limitations in the present study. First, as with any observational study, the effect of residual confounding or unmeasured factors cannot be completely ruled out in spite of the attempts to account for major potential confounders. Secondly, although, prognostic models for patients with HCC were generated with a large sample size, the present study lacked external validation. There is still a need to verify the results of the present study with external results for consistency.

The three most important predictors of survival outcomes of patients with HCC were surgery method, tumor size and education level. Liver transplantation had the best prognosis for patients with HCC, except for those with undifferentiated tumors or distant metastasis.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets analyzed during the current study are available in the SEER repository (www.seer.cancer.gov).

Authors' contributions

HT and SC designed the study, analyzed and interpreted the data and drafted the manuscript. MH, YW and QF designed the study, acquired the data and revised the manuscript. YP and TQ interpreted the data and revised the manuscript. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68: 394-424, 2018.
2. Labgaa I, Demartines N and Melloul E: Surgical resection versus transarterial chemoembolization for intermediate stage hepatocellular carcinoma (BCLC-B): An unsolved question. *Hepatology* 69: 923, 2019.
3. Baddour HM Jr, Fedewa SA and Chen AY: Five- and 10-year cause-specific survival rates in carcinoma of the minor salivary gland. *JAMA Otolaryngol Head Neck Surg* 142: 67-73, 2016.
4. Lexe G, Monaco J, Doyle S, Basavanahally A, Reddy A, Seiler M, Ganesan S, Bhanot G and Madabhushi A: Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging. *Exp Biol Med* (Maywood) 234: 860-879, 2009.
5. Bruix J and Sherman M: American Association for the Study of Liver Diseases: Management of hepatocellular carcinoma: An update. *Hepatology* 53: 1020-1022, 2011.
6. European Association For The Study of The Liver and European Organisation For Research And Treatment of Cancer: EASL-EORTC clinical practice guidelines: Management of hepatocellular carcinoma. *J Hepatol* 56: 908-943, 2012.
7. Hyun MH, Lee YS, Kim JH, Lee CU, Jung YK, Seo YS, Yim HJ, Yeon JE and Byun KS: Hepatic resection compared to chemoembolization in intermediate- to advanced-stage hepatocellular carcinoma: A meta-analysis of high-quality studies. *Hepatology* 68: 977-993, 2018.
8. Kutlu OC, Chan JA, Aloia TA, Chun YS, Kaseb AO, Passot G, Yamashita S, Vauthey JN and Conrad C: Comparative effectiveness of first-line radiofrequency ablation versus surgical resection and transplantation for patients with early hepatocellular carcinoma. *Cancer* 123: 1817-1827, 2017.
9. Shimose S, Tanaka M, Iwamoto H, Niizeki T, Shirono T, Aino H, Noda Y, Kamachi N, Okamura S, Nakano M, *et al*: Prognostic impact of transcatheter arterial chemoembolization (TACE) combined with radiofrequency ablation in patients with unresectable hepatocellular carcinoma: Comparison with TACE alone using decision-tree analysis after propensity score matching. *Hepatol Res* 49: 919-928, 2019.
10. Omata M, Lesmana LA, Tateishi R, Chen PJ, Lin SM, Yoshida H, Kudo M, Lee JM, Choi BI, Poon RT, *et al*: Asian pacific association for the study of the liver consensus recommendations on hepatocellular carcinoma. *Hepatology Int* 4: 439-474, 2010.
11. Jarnagin W, Chapman WC, Curley S, D'Angelica M, Rosen C, Dixon E and Nagorney D: American Hepato-Pancreato-Biliary Association, Society of Surgical Oncology and Society for Surgery of the Alimentary Tract: Surgical treatment of hepatocellular carcinoma: Expert consensus statement. *HPB (Oxford)* 12: 302-310, 2010.
12. Korean Liver Cancer Study Group (KLCSG) and National Cancer Center, Korea (NCC): 2014 KLCSG-NCC Korea practice guideline for the management of hepatocellular carcinoma. *Gut Liver* 9: 267-317, 2015.
13. Kudo M, Matsui O, Izumi N, Iijima H, Kadoya M, Imai Y, Okusaka T, Miyayama S, Tsuchiya K, Ueshima K, *et al*: JSH consensus-based clinical practice guidelines for the management of hepatocellular carcinoma: 2014 update by the liver cancer study group of Japan. *Liver Cancer* 3: 458-468, 2014.
14. Kaibori M, Yoshii K, Yokota I, Hasegawa K, Nagashima F, Kubo S, Kon M, Izumi N, Kadoya M, Kudo M, *et al*: Impact of advanced age on survival in patients undergoing resection of hepatocellular carcinoma: Report of a Japanese nationwide survey. *Annals Surg* 269: 692-699, 2019.

15. Yang A, Xiao W, Chen D, Wei X, Huang S, Lin Y, Zhang C, Lin J, Deng F, Wu C and He X: The power of tumor sizes in predicting the survival of solitary hepatocellular carcinoma patients. *Cancer Med* 7: 6040-6050, 2018.
16. Shinkawa H, Tanaka S, Takemura S, Ishihara T, Yamamoto K and Kubo S: Tumor size drives the prognosis after hepatic resection of solitary hepatocellular carcinoma without vascular invasion. *J Gastrointest Surg*: June 13, 2019 (Epub ahead of print).
17. Wu W, Fang D, Shi D, Bian X and Li L: Effects of marital status on survival of hepatocellular carcinoma by race/ethnicity and gender. *Cancer Manag Res* 10: 23-32, 2018.
18. Bai DS, Zhang C, Chen P, Jin SJ and Jiang GQ: The prognostic correlation of AFP level at diagnosis with pathological grade, progression, and survival of patients with hepatocellular carcinoma. *Sci Rep* 7: 12870, 2017.
19. Sarveazad A, Agah S, Babahajian A, Amini N and Bahardoust M: Predictors of 5 year survival rate in hepatocellular carcinoma patients. *J Res Med Sci* 24: 86, 2019.
20. Song Q, Chen X, Hu W, Mei G, Yang X and Wu H: Downregulation of Epstein-Barr virus-induced gene 3 is associated with poor prognosis of hepatocellular carcinoma after curative resection. *Oncol Lett* 15: 7751-7759, 2018.
21. Noda Y, Kawaguchi T, Korenaga M, Yoshio S, Komukai S, Nakano M, Niizeki T, Koga H, Kawaguchi A, Kanto T and Torimura T: High serum interleukin-34 level is a predictor of poor prognosis in patients with non-viral hepatocellular carcinoma. *Hepatol Res* 49: 1046-1053, 2019.
22. Pang W, Jiang H and Li S: Sparse contribution feature selection and classifiers optimized by concave-convex variation for HCC image recognition. *Biomed Res Int* 2017: 9718386, 2017.
23. Choi GH, Han S, Shim JH, Ryu MH, Ryoo BY, Kang YK, Kim KM, Lim YS and Lee HC: Prognostic scoring models for patients undergoing sorafenib treatment for advanced stage hepatocellular carcinoma in real-life practice. *Am J Clin Oncol* 40: 167-174, 2017.
24. He Y, Wu W, Zheng HM, Li P, McDonald D, Sheng HF, Chen MX, Chen ZH, Ji GY, Zheng ZD, *et al*: Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med* 24: 1532-1535, 2018.
25. Song Y, Gao S, Tan W, Qiu Z, Zhou H and Zhao Y: Multiple machine learnings revealed similar predictive accuracy for prognosis of PNETs from the surveillance, epidemiology, and end result database. *J Cancer* 9: 3971-3978, 2018.
26. Chen X and Ishwaran H: Random forests for genomic data analysis. *Genomics* 99: 323-329, 2012.
27. Liu C, Ackerman HH and Carulli JP: A genome-wide screen of gene-gene interactions for rheumatoid arthritis susceptibility. *Hum Genet* 129: 473-485, 2011.
28. Briggs FB, Ramsay PP, Madden E, Norris JM, Holers VM, Mikuls TR, Sokka T, Seldin MF, Gregersen PK, Criswell LA and Barcellos LF: Supervised machine learning and logistic regression identifies novel epistatic risk factors with PTPN22 for rheumatoid arthritis. *Genes Immun* 11: 199-208, 2010.
29. Degenhardt F, Seifert S and Szymczak S: Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform* 20: 492-503, 2019.
30. Wu C, Ren X and Zhang Q: Incidence, risk factors, and prognosis in patients with primary hepatocellular carcinoma and lung metastasis: A population-based study. *Cancer Manag Res* 11: 2759-2768, 2019.
31. Ryu SM, Lee SH, Kim ES and Eoh W: Predicting survival of spinal ependymoma patients using machine learning algorithms with SEER database. *World Neurosurg*: Dec 28, 2018 (Epub ahead of print).
32. Shukla N, Hagenbuchner M, Win KT and Yang J: Breast cancer data analysis for survivability studies and prediction. *Comput Methods Programs Biomed* 155: 199-208, 2018.
33. Lynch CM, Abdollahi B, Fuqua JD, de Carlo AR, Bartholomai JA, Balgmann RN, van Berkel VH and Frieboes HB: Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform* 108: 1-8, 2017.
34. Adamo MB, Johnson CH, Ruhl JL and Dickie LA (eds): SEER Program Coding and Staging Manual. National Cancer Institute, NIH Publication no. 12-5581, Bethesda, MD, 2012.
35. Raghu G, Chen SY, Yeh WS, Maroni B, Li Q, Lee YC and Collard HR: Idiopathic pulmonary fibrosis in US medicare beneficiaries aged 65 years and older: Incidence, prevalence, and survival, 2001-11. *Lancet Respir Med* 2: 566-572, 2014.
36. Shen Y, Guo H, Wu T, Lu Q, Nan KJ, Lv Y and Zhang XF: Lower education and household income contribute to advanced disease, less treatment received and poorer prognosis in patients with hepatocellular carcinoma. *J Cancer* 8: 3070-3077, 2017.
37. Jones PD, Diaz C, Wang D, Gonzalez-Diaz J, Martin P and Kobetz E: The impact of race on survival after hepatocellular carcinoma in a diverse American population. *digestive diseases and sciences. Dig Dis Sci* 63: 515-528, 2018.
38. Lai MW, Chu YD, Lin CL, Chien RN, Yeh TS, Pan TL, Ke PY, Lin KH and Yeh CT: Is there a sex difference in postoperative prognosis of hepatocellular carcinoma? *BMC Cancer* 19: 250, 2019.
39. Lee GC, Ferrone CR, Vagefi PA, Upptot RN, Tanabe KK, Lillemoe KD, Blaszkowsky LS and Qadan M: Surgical resection versus ablation for early-stage hepatocellular carcinoma: A retrospective cohort analysis. *Am J Surg* 218: 157-163, 2019.
40. Yu B, Ding Y, Liao X, Wang C, Wang B and Chen X: Radiofrequency ablation versus surgical resection in elderly patients with early-stage hepatocellular carcinoma in the era of organ shortage. *Saudi J Gastroenterol* 24: 317-325, 2018.
41. Mills A, Thayer D, Noda C, Salter A, Tao Y, Xing M, Martin R, Ramaswamy R and Akinwande O: Thermal ablation versus surgical resection for localized hepatocellular carcinoma: A population study using the SEER database. *Future Oncol* 14: 631-645, 2018.
42. Yim SY, Seo YS, Jung CH, Kim TH, Lee JM, Kim ES, Keum B, Jong YK, An H, Kim JH, *et al*: The management and prognosis of patients with hepatocellular carcinoma: What has changed in 20 years? *Liver Int* 36: 445-453, 2016.
43. Shen J, Liu J, Li C, Wen T, Yan L and Yang J: The impact of tumor differentiation on the prognosis of HBV-associated solitary hepatocellular carcinoma following hepatectomy: A propensity score matching analysis. *Dig Dis Sci* 63: 1962-1969, 2018.
44. Zhang JF, Shu ZJ, Xie CY, Li Q, Jin XH, Gu W, Jiang FJ and Ling CQ: Prognosis of unresectable hepatocellular carcinoma: Comparison of seven staging systems (TNM, Okuda, BCLC, CLIP, CUPI, JIS, CIS) in a Chinese cohort. *PLoS One* 9: e88182, 2014.
45. Yoshizumi T, Ikegami T, Yoshiya S, Motomura T, Mano Y, Muto J, Ikeda T, Soejima Y, Shirabe K and Maehara Y: Impact of tumor size, number of tumors and neutrophil-to-lymphocyte ratio in liver transplantation for recurrent hepatocellular carcinoma. *Hepatol Res* 43: 709-716, 2013.
46. Jain SS, Sarkar IN, Stey PC, Anand RS, Biron DR and Chen ES: Using demographic factors and comorbidities to develop a predictive model for ICU mortality in patients with acute exacerbation COPD. *AMIA Annu Symp Proc* 2018: 1319-1328, 2018.
47. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M and van Hijum SA: Data mining in the life sciences with random forest: A walk in the park or lost in the jungle? *Brief Bioinform* 14: 315-326, 2013.
48. Liu G, Dong C, Wang X, Hou G, Zheng Y, Xu H, Zhan X and Liu L: Regulatory activity based risk model identifies survival of stage II and III colorectal carcinoma. *Oncotarget* 8: 98360-98370, 2017.
49. Kawaguchi T, Tokushige K, Hyogo H, Aikata H, Nakajima T, Ono M, Kawanaka M, Sawada K, Imajo K, Honda K, *et al*: A data mining-based prognostic algorithm for NAFLD-related hepatoma patients: A nationwide study by the Japan study group of NAFLD. *Sci Rep* 8: 10434, 2018.
50. Cheong JY, Lee KM, Cho SW, Won JH, Kim JK, Wang HJ, Hahm KB and Kim JH: Survival benefits of intra-arterial infusion chemotherapy in patients with advanced hepatocellular carcinoma with portal vein tumor thrombosis. *Hepatol Res* 32: 127-133, 2005.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.