**Name of Journal:** *World Journal of Gastrointestinal Oncology*

**Manuscript NO:** 83815

**Manuscript Type:** ORIGINAL ARTICLE

***Observational Study***

**Development and application of hepatocellular carcinoma risk prediction model based on clinical characteristics and liver related indexes**

Liu ZJ *et al*. HCC risk prediction model

Zhi-Jie Liu, Yue Xu, Wen-Xuan Wang, Bin Guo, Guo-Yuan Zhang, Guang-Cheng Luo, Qiang Wang

**Zhi-Jie Liu,** Department of Clinical Transfusion, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, Sichuan Province, China

**Yue Xu, Bin Guo, Guo-Yuan Zhang, Guang-Cheng Luo, Qiang Wang,** Department of Clinical Laboratory, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, Sichuan Province, China

**Wen-Xuan Wang,** Department of Radiology, Nanchong Central Hospital, Nanchong 637000, Sichuan Province, China

**Author contributions:** Luo GC and Wang Q contributed to the conceptualisation and design of this study; Liu ZJ, Xu Y, Wang WX, Guo B, and Zhang GY contributed to the clinical data collection and analysis; Liu ZJ, Xu Y, and Wang WX prepared and wrote the first draft of this manuscript; Wang Q revised the manuscript.

**Corresponding author: Qiang Wang, PhD, Research Assistant Professor,** Department of Clinical Laboratory, Affiliated Hospital of North Sichuan Medical College, No. 1 Maoyuan South Road, Shunqing District, Nanchong 637000, Sichuan Province, China. wqiang\_1981@126.com

**Received:** February 10, 2023

**Revised:** May 28, 2023

**Accepted:** June 25, 2023

**Published online:**

**Abstract**

BACKGROUND

Hepatocellular carcinoma (HCC) is difficult to diagnose with poor therapeutic effect, high recurrence rate and has a low survival rate. The survival of patients with HCC is closely related to the stage of diagnosis. At present, no specific serological indicator or method to predict HCC, early diagnosis of HCC remains a challenge, especially in China, where the situation is more severe.

AIM

To identify risk factors associated with HCC and establish a risk prediction model based on clinical characteristics and liver-related indicators.

METHODS

The clinical data of patients in the Affiliated Hospital of North Sichuan Medical College from 2016 to 2020 were collected, using a retrospective study method. The results of needle biopsy or surgical pathology were used as the grouping criteria for the experimental group and the control group in this study. Based on the time of admission, the cases were divided into training cohort (*n* = 1739) and validation cohort (*n* = 467). Using HCC as a dependent variable, the research indicators were incorporated into logistic univariate and multivariate analysis. An HCC risk prediction model, which was called NSMC-HCC model, was then established in training cohort and verified in validation cohort.

RESULTS

Logistic univariate analysis showed that, gender, age, alpha-fetoprotein, and protein induced by vitamin K absence or antagonist-II, [gamma-glutamyl transferase](http://www.baidu.com/Link?url=wYikzgzQkbnEpG_8VtGWOEisIcT1vdDLockQkwg1MPpx481nGsQUoEXFOf7oP6LzrIDaDkYrOcBlc9zypzQ7a8jIZwzUtPh1ZN5dupfdlgLlBKEOqAihUwwrFDZNNoAJ), aspartate aminotransferase and hepatitis B surface antigen were risk factors for HCC, alanine aminotransferase, total bilirubin and total bile acid were protective factors for HCC. When the cut-off value of the NSMC-HCC model joint prediction was 0.22, the area under receiver operating characteristic curve (AUC) of NSMC-HCC model in HCC diagnosis was 0.960, with sensitivity 94.40% and specificity 95.35% in training cohort, and AUC was 0.966, with sensitivity 90.00% and specificity 94.20% in validation cohort. In early-stage HCC diagnosis, the AUC of NSMC-HCC model was 0.946, with sensitivity 85.93% and specificity 93.62% in training cohort, and AUC was 0.947, with sensitivity 89.10% and specificity 98.49% in validation cohort.

CONCLUSION

The newly NSMC-HCC model was an effective risk prediction model in HCC and early-stage HCC diagnosis.

**Key Words:** Hepatocellular carcinoma; Risk prediction model; Logistic regression model; Tumour markers; Metabolic markers; Clinical characteristics

Liu ZJ, Xu Y, Wang WX, Guo B, Zhang GY, Luo GC, Wang Q. Development and application of hepatocellular carcinoma risk prediction model based on clinical characteristics and liver related indexes. *World J Gastrointest Oncol* 2023; In press

**Core Tip:** This study identified the risk factors associated with hepatocellular carcinoma (HCC) and further established a risk prediction model based on the clinical characteristics and liver indicators. By evaluating in the training cohort and confirming with the validation cohort, we proved that the proposed model has good sensitivity and specificity in high-risk populations with HCC, with a high accuracy in early-stage HCC diagnosis. In addition, we recommend a risk prediction scale (low to very high risk). This will help clinicians to diagnose HCC earlier and thus improve the prognosis of patients.

**INTRODUCTION**

Hepatocellular carcinoma (HCC), a common and highly malignant tumour globally, ranks fourth among the most common malignant tumours in China[[1](#_ENREF_1" \o "Zhou, 2019 #1)]. Approximately 86% of HCC is caused by hepatitis B virus (HBV) infection[[2](#_ENREF_2" \o "Wang, 2017 #2)]. HCC is characterized by high malignancy and rapid progression. In developing countries, only 30% of patients with HCC are diagnosed at an early stage and thus receive effective treatment[[3](#_ENREF_3" \o "Zhang, 2018 #43)]. As such, early diagnosis of HCC is difficult, therapeutic effect and prognosis is poor, and recurrence rate is high. Early diagnosis and early treatment significantly improve prognosis and prolong survival among patients with HCC.

Serological markers and imaging are the most important methods for monitoring HCC in the early stages. In general, serology is the first choice for clinical detection as it is simple, repeatable, and cheap. Alpha-fetoprotein (AFP) is the most widely used tumour marker. However, AFP is negative in about one-third of patients, limiting the diagnosis of HCC[[4](#_ENREF_4" \o "Matsuda, 2011 #3)]. Protein induced by vitamin K absence or antagonist-II (PIVKA-II) is an abnormal thrombin that lacks clotting activity, and is a commonly used diagnostic marker of HCC in clinics. Previous study show that the sensitivity and specificity of PIVKA-II in diagnosing HCC is higher than that of AFP[[5](#_ENREF_5" \o "Baek, 2009 #18)]. However, the sensitivity of PIVKA-II with a single tumour diameter smaller than 2 cm is 30%-53%, while that of AFP is only 13%[[6](#_ENREF_6" \o "Li, 2021 #40)]. Thus, the tumour markers, PIVKA-II and AFP have limitations in diagnosing HCC.

As a unique method of diagnosis, combined diagnosis can make up for the shortcomings of individual items and significantly improve their diagnostic efficiency. A single-centre cohort study showed that when the critical value of AFP was 400 μg/L, the sensitivity of AFP combined with PIVKA-II for HCC diagnosis increased from 28.7% to 54.9%[[7](#_ENREF_7" \o "Ertle, 2013 #5)]. Another study showed that the sensitivity of combined detection of HCC, AFP, and PIVKA-II, with a diameter smaller than 2 cm was 57%, and the sensitivity of combined diagnosis with HCC and PIVKA-II for a diameter greater than 3 cm was 84%[[8](#_ENREF_8" \o "Nie, 2013 #41)]. Hanahan and Weinberg[[9](#_ENREF_9" \o "Hanahan, 2011 #20)] proposed the abnormal energy metabolism of tumour cells, and suggested that the structural and functional changes of some genes in tumour cells lead to a series of metabolic changes characterised by the Warburg effect that help tumour cells adapt to the microenvironment of local hypoxia. Therefore, tumor markers combined with metabolic markers in the diagnosis of HCC has become a new idea. The study by Park *et al*[[10](#_ENREF_10" \o "Park, 2017 #21)] showed that the sensitivity of a diagnostic test based on PIVKA-II combined with AFP or AFP-L3 was the highest at 84.81%, and the specificity was 51.95%. Compared with the diagnosis using only two indicators, the area under receiver operating characteristic curve (AUC) was 0.684. A meta-analysis of 9597 patients in 11 studies showed that the AUC of AFP + AFP-L3 + des-gamma-carboxy prothrombin (DCP) for HCC was 0.91. The sensitivity and specificity were 88% and 79%, respectively, which were higher than those of AFP alone[[11](#_ENREF_11" \o "Wang, 2020 #12)].

The GALAD model was established in 2014, based on sex, age, and three tumour markers: AFP, AFP-L3%, and PIVKA-II. The AUC of this model in diagnosing HCC was 0.91, and the sensitivity and specificity were 68% and 95%, respectively[[12](#_ENREF_12" \o "Best, 2020 #35)]. Moreover, the ASAP model is based on sex, age, AFP, and PIVKA-II. The AUC of the ASAP model for diagnosis of HCC is 0.941, and the sensitivity and specificity are 88.3% and 85.1%, respectively. This model has a good diagnostic efficacy in Chinese patients with HCC secondary to HBV, even better than the GALAD model[[13](#_ENREF_13" \o "Yang, 2019 #28)]. These two models have good sensitivities and specificities in diagnosing HCC; however, they do not include liver-related metabolic markers. There is a lack of a single test with high sensitivity and specificity for the early diagnosis of HCC and prediction of high-risk groups. Thus, this study aimed to use logistic regression analysis to screen risk factors related to HCC, to build a risk prediction model, and to provide screening methods for high-risk groups of HCC.

**MATERIALS AND METHODS**

***Study objects***

Data of 2206 patients were collected from the Affiliated Hospital of North Sichuan Medical College between January 2016 and December 2020. Patients admitted from January 2016 to December 2019 were included in training cohort, which included 496 patients with HCC and 1243 patients with benign liver diseases. Patients admitted from January 2020 to December 2020 were included in the validation cohort, which included 156 patients with HCC and 311 patients with benign liver diseases.

The selection criteria of HCC group were: (1) Patients with HCC diagnosed for the first time in our hospital; (2) Patients with HCC diagnosed by pathological biopsy or intraoperative pathological biopsy; (3) Anti-tumour behaviours such as no radical operation, no transcatheter arterial chemoembolisation and radiotherapy, and chemotherapy; and (4) Patients with complete medical records. The diagnostic criteria of HCC followed the standard for diagnosis and treatment of primary liver cancer (2019 edition) issued by CSCO in 2020[[14](#_ENREF_14)].

The selection criteria for the early HCC group were: (1) Stage I HCC diagnosed by pathology (liver puncture) biopsy or surgical pathology and in accordance with the Chinese staging of liver cancer programme: Single tumour, diameter ≤ 5 cm, no vascular invasion and extrahepatic metastasis, and liver function grade Child-Pugh A/B[15]; (2) No anti-tumour treatment; and (3) Complete clinical information and examination indicators.

The exclusion criteria of HCC group were: (1) Patients with HCC who were not diagnosed for the first time in our hospital; (2) Complicated with other serious diseases or conditions or a history of surgery, radiotherapy and chemotherapy, and serious diseases or major injuries and burns occurring seven days before sampling; (3) Patients who recently consumed vitamin K or vitamin K antagonists such as warfarin; (4) Reproductive and embryonic tumours or other tumours with liver metastasis; and (5) Incomplete clinical data.

Benign liver diseases included hepatic hemangioma, hepatic cyst, hepatic abscess, hepatic hemangioma, liver cirrhosis, chronic hepatitis B, cholecystolithiasis, cholecystitis, and hepatolithiasis. These patients did not report development of HCC after at least six months follow-up.

***Experimental method***

Blood samples were collected from the patients enrolled in the study within three days after admission. Approximately 3-5 mL fasting venous blood was collected in heparin anticoagulant and anticoagulant-free serum tubes. After collection, samples were mixed, coagulated, and centrifuged at 3500 rpm for 5 min. The serum level of PIVKA-II was detected by chemiluminescence microparticle immunoassay (Archtect i1000, ABBOTT, United States). The serum level of AFP was detected by electrochemiluminescence assay (Cobas e602, Roche, Inc., Germany). The serum level of total bilirubin (TBIL) was detected by vanadate oxidation method (ADVIA-2400, SIEMENS, Germany). The serum levels of [gamma-glutamyl transferase](http://www.baidu.com/Link?url=wYikzgzQkbnEpG_8VtGWOEisIcT1vdDLockQkwg1MPpx481nGsQUoEXFOf7oP6LzrIDaDkYrOcBlc9zypzQ7a8jIZwzUtPh1ZN5dupfdlgLlBKEOqAihUwwrFDZNNoAJ) (GGT), alanine amino transferase (ALT) , and [aspartate transaminase](http://www.baidu.com/Link?url=8n1MswXxMeHwczZ9KOgXvur001zysKIc0yApRzykn9Vi2AnTNfbd0rdaTKM0FtRHSOG8QV1f4RK6_gsbQUZFfcHUpDCLh-KXfCx-WugYfCJ8fZcqa2q6FN4y0SBjaX3nfYmnBlR5nHJYb4Iu2Qrl7a) (AST) were detected by rate method (ADVIA-2400, SIEMENS, Germany). The serum level of total bile acid (TBA) was detected by enzyme cycle method (ADVIA-2400, SIEMENS, Germany). The serum level of albumin (ALB) was detected by albumin-bromocresol green method (ADVIA-2400, SIEMENS, Germany).

***Statistical analysis***

Descriptive statistics were used to summarise the characteristics of all participants. The metrological data were expressed by the median (interquartile interval), and each group was tested using normality and variance homogeneity tests before analysing. The differences between the two groups were compared using independent samples *t*-test. If the results were not normally distributed, the differences between groups were compared by nonparametric rank sum test (Mann-Whitney *U* test). The categorical data were expressed as percentage (%), and a chi-square test was used for comparison between the two groups. Statistically significant factors were included in logistic multivariate regression analysis. Taking HCC as the dependent variable, the odds ratio (OR) and 95% confidence interval (CI) of each factor were calculated. The risk factors related to HCC were screened out, and the distribution was represented using forest map. The risk prediction model, which was called NSMC-HCC model, was constructed based on the results of logistic multi-factor analysis (drawn by the R3.5 software “rms” package). The AUC, sensitivity, specificity, accuracy, and other related indexes under the receiver operating curve were used to test and evaluate the NSMC-HCC model. Statistical Package for Social Sciences (SPSS), version 26.0 (IBM, United States) and R 3.5 (MathSoft, United States) were used for statistical analysis. The difference was statistically significant at the level of *P* < 0.05.

**RESULTS**

***Patient characteristics***

Of the 2206 patients, 1739 were included in the training cohort to establish the risk prediction model and 467 were included in the validation cohort to evaluate the prediction effect of the model. The age of patients with HCC in the training cohort and validation cohort was significantly higher than that of patients without HCC (*P* < 0.001). Although, in terms of sex, men were dominant in both the training and validation cohorts, the sex composition did not vary significantly between the groups (*P* > 0.05). In the training and validation cohorts, the levels of tumour markers AFP and PIVKA-II were significantly different between patients with and without HCC. The serum levels of patients with HCC were significantly higher than those of patients without HCC (*P* < 0.001), as shown in Table 1.

***HCC risk prediction model***

For data conversion, logarithmic (lg) conversion was performed on all variables except sex, age, hepatitis B surface antigen (HBsAg), and Child-Pugh scores, as lg (AFP), lg (PIVKA-II), lg (TBIL), lg (GGT), lg (AST), lg (ALT), lg (TBA), and lg (ALB).

Taking HCC as the dependent variable and the above mentioned research indices as independent variables, the binary logistic regression analysis was carried out, where the binary independent variables were assigned as follows: Male = 1, female = 0; HBsAg positive = 1, HBsAg negative = 0; Child-Pugh Class A = 1, Child-Pugh Class B = 2, and Child-Pugh Class C = 3. Multivariate logistic regression analysis showed that sex, age, HBsAg, AFP, PIVKA-II, GGT, and AST were risk factors for HCC, while TBIL, ALT and TBA were protective factors for HCC (Figure 1).

Based on the multi-factor risk prediction model of HCC, the line diagram (NSMC-HCC model) established by the training cohort data was used to predict the risk of HCC (Figure 2). The nomogram was calculated as follows: ln (P/1-P) = -7.115 + 1.879 × lg (PIVKA-II) + 1.422 × lg (AFP) + 1.537 × HBsAg + 1.115 × lg (GGT) + 1.133 × lg (ALT) + 0.627 × age + 0.051 × sex - 0.840 × lg (TBA) - 1.464 × lg (ALT) - 2.836 × lg (TBIL).

***AFP and PIVKA-II alone or both in HCC diagnosis***

In the training cohort, compared with 1243 patients with benign liver diseases, the AUC of AFP for HCC was 0.812, and the sensitivity and specificity were 49.19% and 93.00%, respectively. The AUC of PIVKA-II for HCC was 0.882, and the sensitivity and specificity were 63.91% and 94.93%, respectively. The AUC of AFP combined with PIVKA-II for HCC was 0.896, and the sensitivity and specificity were 69.35% and 95.82%, respectively. The AUC of AFP for early-stage HCC was 0.860, and the sensitivity and specificity were 51.06% and 97.49%, respectively. The AUC of PIVKA-II for early-stage HCC was 0.863, and the sensitivity and specificity were 40.43% and 96.98%, respectively. The AUC of AFP combined with PIVKA-II for early-stage HCC was 0.923, and the sensitivity and specificity were 63.83% and 96.98%, respectively. However, with the addition of Child Pugh classification for liver function comparison and HBsAg grouping comparison, the results may be more complete.

In the validation cohort, compared with 311 patients with benign liver diseases, the AUC of AFP for detection of HCC was 0.845, and the sensitivity and specificity were 52.56% and 91.00%, respectively. The AUC of PIVKA-II for detection of HCC was 0.878, and the sensitivity and specificity were 64.74% and 94.53%, respectively. The AUC of AFP combined with PIVKA-II for detection of HCC was 0.888, and the sensitivity and specificity were 67.95% and 95.17%, respectively. The AUC of AFP for early-stage HCC was 0.714, the sensitivity and specificity were 14.55% and 98.49%, respectively. The AUC of PIVKA-II for early-stage HCC was 0.868, and the sensitivity and specificity were 52.73% and 96.48%, respectively. The AUC of AFP combined with PIVKA-II for early-stage HCC was 0.896, and the sensitivity and specificity were 50.91% and 95.98%, respectively. The sensitivity of AFP alone diagnosis and combination diagnosis was much lower than that of the training cohort, which may be due to the fact that fewer cases of early-stage HCC were enrolled (Table 2).

***Ability of NSMC-HCC model in HCC diagnosis***

In the training cohort, when the diagnostic threshold for predicting risk was set at 0.22, the AUC of NSMC-HCC model in HCC diagnosis was 0.960 (95%CI: 0.950-0.971) (Figure 3A), with a sensitivity of 94.40%, specificity of 95.35%, and accuracy of 94.67%. The AUC of NSMC-HCC model in early-stage HCC diagnosis was 0.946 (95%CI: 0.901-0.991) (Figure 3C), with a sensitivity of 85.93%, specificity of 93.62%, and accuracy of 87.40%.

The data of the validation cohort were used to verify the NSMC-HCC model. The results showed that the AUC of NSMC-HCC model in HCC diagnosed was 0.966 (95%CI: 0.945-0.986) (Figure 3B). There was no significant difference between training cohort and validation cohort (*P* > 0.05). When the diagnostic threshold for predicting risk is set at 0.22, the sensitivity, specificity, and accuracy of NSMC-HCC model in the validation cohort were 90.00%, 94.20%, and 93.58%, respectively. The AUC of NSMC-HCC model in early-stage HCC diagnosis was 0.947 (95%CI: 0.901-0.994) (Figure 3D), with a sensitivity of 89.10%, specificity of 98.49%, and accuracy of 96.46% (Table 3). The AUC of NSMC-HCC model in HCC and early-stage HCC diagnosis were all higher than that of AFP combined with PIVKA-II in training cohort and validation cohort (all *P* < 0.001).

***Proposed risk scale***

According to the data from the training and validation cohorts, we proposed a simple standard scale of risk prediction probability based on the NSMC-HCC model for clinicians to evaluate the risk level of HCC (Table 4). This mainly follows the principles: (1) The maximum risk prediction probability whose negative predictive value (NPV) ≥ 99.00% is defined as low risk; (2) The risk prediction probability between the minimum risk prediction probability whose NPV < 99.00% and cut-off is defined as medium risk; (3) The risk prediction probability between cut-off and the maximum risk prediction probability whose positive predictive value (PPV) < 99.00% is defined as high risk; and (4) The minimum risk prediction probability whose PPV ≥ 99.00% is defined as the highest risk (most likely HCC).

**DISCUSSION**

The morbidity and mortality of HCC is ranked among the top five causes, globally[16]. Most patients are diagnosed at middle and advanced stages, and thus lose essential time for optimal treatment. The prognosis of patients with HCC largely depends on the stage of the diagnostic time. From 2012 to 2015, the 5-year survival rate of liver cancer in China is only 12.1%-18.0%[17]. In contrast, the prognosis of patients with early diagnosis is more than 70%. For example, Lim *et al*[18] analysed the clinical data of 100 patients with early-stage HCC and reported that the 5-year survival rate after surgery was as high as 90%. Therefore, appropriate early screening of high-risk HCC groups is crucial to improve the prognosis. In this study, a risk prediction model for HCC was established by combining sex, age, tumour markers of AFP and PIVKA-II, metabolic markers of TBIL, GGT, AST, ALT, TBA, and infection index HBsAg. By evaluating in the training cohort and confirming with the validation cohort, we proved that the proposed model has good sensitivity and specificity in high-risk populations with HCC, with a high accuracy in early-stage HCC diagnosis.

AFP and PIVKA-II are the two most widely used tumour markers in diagnosing HCC; however, their sensitivity and specificity for diagnosing HCC are not high, hence their limited utility. Combining markers can improve the sensitivity of diagnosis. A risk prediction model, ASAP model from 11 medical centres in China which included age, sex, AFP, and PIVKA-II, was used to predict the risk of HCC in patients with HBV infection. The model has a good clinical value for predicting HBV-HCC (AUC is 0.941). The diagnostic sensitivity and specificity are 85.3% and 90.4%, respectively[13]. However, other risk factors crucial for HCC development were not included in the model. Among the currently available prediction tools, the line chart model has high accuracy and good discrimination in terms of predicting results and is easy to use[19]. The nomogram proposed in this study contains ten comprehensive and easily available patient variables. The AUC in the training and validation cohort was 0.960 and 0.966, respectively. The diagnostic sensitivity and specificity for the training cohort were 94.40% and 95.35%, respectively, while that for the validation cohort was 90.00% and 94.20%, respectively. Those results showed high value of AUC, low value of standard error, and good diagnostic efficiency of the NSMC-HCC model in HCC diagnosis, which was better than AFP and PIVKA-II alone or combination.

The metabolic markers included in this study were AST, ALT, TBIL, GGT, TBA, and ALB. Univariate analysis showed that GGT and AST were risk factors for HCC, while ALT, TBIL, and TBA were protective factors for HCC. Although the occurrence of HCC associated with AST and GGT has not been reported, the study of Yang *et al*[20] showed that the combined detection of AFP and GGT/AST plays an important role in the differential diagnosis of benign liver disease and HCC. They further mentioned that GGT and AST are risk factors for the prognosis of HCC treatment[21]. Hernaez *et al*[22] reported that among men in Taiwan, without cancer, elevations in ALT, AST, and GGT are associated with future all-cause death, all cancer, and HCC mortality. These studies support our findings that GGT and AST are high-risk factors associated with HCC.

Furthermore, we have quantified the possibility of HCC risk, and this can evaluate the risk of HCC according to the risk prediction probability. According to the risk stratification of HCC, abdominal ultrasound and serum AFP are recommended as routine screening, and multimodal liver magnetic resonance imaging and/or computed tomography are recommended for enhanced screening[23]. For low risk patients, we recommend routine screening once a year; for moderate risk patients, we recommend routine screening every 6 mo; for high-risk patients, we recommend routine screening every 3-6 mo and intensive screening every 6-12 mo; for very high-risk patients, we recommend routine screening every 3 mo and intensive screening every 6 mo[24]. Therefore, the HCC risk prediction model constructed in this study can help clinicians make early-stage HCC diagnosis and improve the early detection rate of HCC.

The newly established NSMC-HCC model can reliably predict the occurrence of HCC and has a strong accuracy for the early detection of HCC and the NSMC-HCC model performs well in the training and validation cohorts. This will contribute to the risk prediction and estimation of the high-risk population of HCC. Recent study has shown that the purpose of risk prediction models is not just to classify patients into simple high or low risk groups, but to view pathogenic risk as a continuum, interpreted in the clinical context of each patient, which can be constructed by grouping risk factors[25]. All the subjects were included in subgroup analysis (0 risk factor,1 risk factor, ≥ 2 risk factors), and the results of different risk groups were observed and predicted to make the risk prediction model more individualized and reduce unnecessary testing and treatment for healthy people.

Therefore, in the follow-up study, we stratified the risk factors and improved upon the shortcomings of this study, such as incomplete risk factors and lack of HBV DNA. The basic clinical information of the study participants was incomplete, including the family history and ethnic history of HCC patients, pathogenic factors of patients, and imaging indicators of patients’ tumours. As the sample size was small and the patients were from the same medical institution, a sampling bias may have occurred. All participants in this study were of Asian ethnicity, and our prediction model was applicable to most Asian populations due to the genetic and environmental differences between different ethnic groups.

**CONCLUSION**

In summary, our study identified the risk factors associated with HCC and further established a risk prediction model based on the clinical characteristics and liver indicators. The broader aim of this study is to aid early detection of HCC to improve the prognosis among patients. We believe that our study makes a significant contribution to the literature as it provides robust evidence of differences in sensitivity/specificity and accuracy among single and combination of diagnostic tests that will help in early detection of HCC.

**ARTICLE HIGHLIGHTS**

***Research background***

Hepatocellular carcinoma (HCC) is the most common primary liver cancer, which currently faces difficulties in early diagnosis, high recurrence rate, and low overall survival rate. Early detection and diagnosis are main way to reduce the incidence rate and mortality of HCC.

***Research motivation***

Using logistic regression models to identify high-risk factors related to HCC, and combining clinical features and liver related indicators to establish a predictive model for HCC.

***Research objectives***

This study aims to establish a model that can predict HCC and can be applied in clinical practice.

***Research methods***

Patients were divided into a modeling group and a validation group based on the results of puncture biopsy or surgical pathological diagnosis. HCC was used as the dependent variable, and the research indicators were included in logistic univariate and multivariate analysis to establish a HCC risk prediction model.

***Research results***

Logistic univariate analysis showed that, gender, age, alpha-fetoprotein (AFP), and protein induced by vitamin K absence or antagonist-II (PIVKA-II), [gamma-glutamyl transferase](http://www.baidu.com/Link?url=wYikzgzQkbnEpG_8VtGWOEisIcT1vdDLockQkwg1MPpx481nGsQUoEXFOf7oP6LzrIDaDkYrOcBlc9zypzQ7a8jIZwzUtPh1ZN5dupfdlgLlBKEOqAihUwwrFDZNNoAJ) (GGT), [aspartate transaminase](http://www.baidu.com/Link?url=8n1MswXxMeHwczZ9KOgXvur001zysKIc0yApRzykn9Vi2AnTNfbd0rdaTKM0FtRHSOG8QV1f4RK6_gsbQUZFfcHUpDCLh-KXfCx-WugYfCJ8fZcqa2q6FN4y0SBjaX3nfYmnBlR5nHJYb4Iu2Qrl7a) (AST), hepatitis B surface antigen (HBsAg) were risk factors for HCC, and in the training cohort and confirming with the validation cohort, the NSMC-HCC model has good sensitivity and specificity in high-risk populations with HCC, with a high accuracy in early-stage HCC diagnosis.

***Research conclusions***

We have established a relatively effective HCC risk prediction model that includes gender, age, AFP, PIVKA-I, total bilirubin, GGT, AST, alanine amino transferase, total bile acid, and HBsAg, and this model has high accuracy in the diagnosis of early HCC.

***Research perspectives***

This study is an observational study that included samples from the same medical institution, which may have sampling bias. Further validation of multicenter, large sample studies is needed in the future.

**REFERENCES**

1 **Zhou M**, Wang H, Zeng X, Yin P, Zhu J, Chen W, Li X, Wang L, Wang L, Liu Y, Liu J, Zhang M, Qi J, Yu S, Afshin A, Gakidou E, Glenn S, Krish VS, Miller-Petrie MK, Mountjoy-Venning WC, Mullany EC, Redford SB, Liu H, Naghavi M, Hay SI, Wang L, Murray CJL, Liang X. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019; **394**: 1145-1158 [PMID: 31248666 DOI: 10.1016/S0140-6736(19)30427-1]

2 **Wang M**, Wang Y, Feng X, Wang R, Wang Y, Zeng H, Qi J, Zhao H, Li N, Cai J, Qu C. Contribution of hepatitis B virus and hepatitis C virus to liver cancer in China north areas: Experience of the Chinese National Cancer Center. *Int J Infect Dis* 2017; **65**: 15-21 [PMID: 28935244 DOI: 10.1016/j.ijid.2017.09.003]

3 **Zhang XX**. [Effects of phosphatidylinositol proteoglycan-3(GPC-3) on the growth and invasion of hepatocellular carcinoma]. Fujian Medical University 2018

4 **Matsuda M**, Asakawa M, Amemiya H, Fujii H. Lens culinaris agglutinin-reactive fraction of AFP is a useful prognostic biomarker for survival after repeat hepatic resection for HCC. *J Gastroenterol Hepatol* 2011; **26**: 731-738 [PMID: 21155886 DOI: 10.1111/j.1440-1746.2010.06532.x]

5 **Baek YH**, Lee JH, Jang JS, Lee SW, Han JY, Jeong JS, Choi JC, Kim HY, Han SY. Diagnostic role and correlation with staging systems of PIVKA-II compared with AFP. *Hepatogastroenterology* 2009; **56**: 763-767 [PMID: 19621698]

6 **Li S**. [Study on the value of GALAD model and GAAP model based on AFP and DCP in hepatocellular carcinoma screening]. Shandong University 2021

7 **Ertle JM**, Heider D, Wichert M, Keller B, Kueper R, Hilgard P, Gerken G, Schlaak JF. A combination of α-fetoprotein and des-γ-carboxy prothrombin is superior in detection of hepatocellular carcinoma. *Digestion* 2013; **87**: 121-131 [PMID: 23406785 DOI: 10.1159/000346080]

8 **Nie DJ**. [Application and research progress of tumor markers in common tumors]. Chongqing Medical University 2013

9 **Hanahan D**, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646-674 [PMID: 21376230 DOI: 10.1016/j.cell.2011.02.013]

10 **Park SJ**, Jang JY, Jeong SW, Cho YK, Lee SH, Kim SG, Cha SW, Kim YS, Cho YD, Kim HS, Kim BS, Park S, Bang HI. Usefulness of AFP, AFP-L3, and PIVKA-II, and their combinations in diagnosing hepatocellular carcinoma. *Medicine (Baltimore)* 2017; **96**: e5811 [PMID: 28296720 DOI: 10.1097/MD.0000000000005811]

11 **Wang X**, Zhang Y, Yang N, He H, Tao X, Kou C, Jiang J. Evaluation of the Combined Application of AFP, AFP-L3%, and DCP for Hepatocellular Carcinoma Diagnosis: A Meta-analysis. *Biomed Res Int* 2020; **2020**: 5087643 [PMID: 33015170 DOI: 10.1155/2020/5087643]

12 **Best J**, Bechmann LP, Sowa JP, Sydor S, Dechêne A, Pflanz K, Bedreli S, Schotten C, Geier A, Berg T, Fischer J, Vogel A, Bantel H, Weinmann A, Schattenberg JM, Huber Y, Wege H, von Felden J, Schulze K, Bettinger D, Thimme R, Sinner F, Schütte K, Weiss KH, Toyoda H, Yasuda S, Kumada T, Berhane S, Wichert M, Heider D, Gerken G, Johnson P, Canbay A. GALAD Score Detects Early Hepatocellular Carcinoma in an International Cohort of Patients With Nonalcoholic Steatohepatitis. *Clin Gastroenterol Hepatol* 2020; **18**: 728-735.e4 [PMID: 31712073 DOI: 10.1016/j.cgh.2019.11.012]

13 **Yang T**, Xing H, Wang G, Wang N, Liu M, Yan C, Li H, Wei L, Li S, Fan Z, Shi M, Chen W, Cai S, Pawlik TM, Soh A, Beshiri A, Lau WY, Wu M, Zheng Y, Shen F. A Novel Online Calculator Based on Serum Biomarkers to Detect Hepatocellular Carcinoma among Patients with Hepatitis B. *Clin Chem* 2019; **65**: 1543-1553 [PMID: 31672853 DOI: 10.1373/clinchem.2019.308965]

14 **Department of Medical Administration, National Health and Health Commission of the People's Republic of China**. [Guidelines for diagnosis and treatment of primary liver cancer in China (2019 edition)]. *Zhonghua Gan Zang Bing Za Zhi* 2020; **28**: 112-128 [PMID: 32164061 DOI: 10.3760/cma.j.issn.1007-3418.2020.02.004]

15 **Anthony PP**, Ishak KG, Nayak NC, Poulsen HE, Scheuer PJ, Sobin LH. The morphology of cirrhosis: definition, nomenclature, and classification. *Bull World Health Organ* 1977; **55**: 521-540 [PMID: 304393]

16 **Zheng RS**, Sun KX, Zhang SW, Zeng HM, Zou XN, Chen R, Gu XY, Wei WW, He J. [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi* 2019; **41**: 19-28 [PMID: 30678413 DOI: 10.3760/cma.j.issn.0253-3766.2019.01.005]

17 **Zeng H**, Chen W, Zheng R, Zhang S, Ji JS, Zou X, Xia C, Sun K, Yang Z, Li H, Wang N, Han R, Liu S, Li H, Mu H, He Y, Xu Y, Fu Z, Zhou Y, Jiang J, Yang Y, Chen J, Wei K, Fan D, Wang J, Fu F, Zhao D, Song G, Chen J, Jiang C, Zhou X, Gu X, Jin F, Li Q, Li Y, Wu T, Yan C, Dong J, Hua Z, Baade P, Bray F, Jemal A, Yu XQ, He J. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health* 2018; **6**: e555-e567 [PMID: 29653628 DOI: 10.1016/S2214-109X(18)30127-X]

18 **Lim C**, Bhangui P, Salloum C, Gómez-Gavara C, Lahat E, Luciani A, Compagnon P, Calderaro J, Feray C, Azoulay D. Impact of time to surgery in the outcome of patients with liver resection for BCLC 0-A stage hepatocellular carcinoma. *J Hepatol* 2017 [PMID: 28989094 DOI: 10.1016/j.jhep.2017.09.017]

19 **Lei Z**, Li J, Wu D, Xia Y, Wang Q, Si A, Wang K, Wan X, Lau WY, Wu M, Shen F. Nomogram for Preoperative Estimation of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma Within the Milan Criteria. *JAMA Surg* 2016; **151**: 356-363 [PMID: 26579636 DOI: 10.1001/jamasurg.2015.4257]

20 **Yang JG**, He XF, Huang B, Zhang HA, He YK. Rule of changes in serum GGT levels and GGT/ALT and AST/ALT ratios in primary hepatic carcinoma patients with different AFP levels. *Cancer Biomark* 2018; **21**: 743-746 [PMID: 29286913 DOI: 10.3233/CBM-170088]

21 **Hernaez R**, Yeh HC, Lazo M, Chung HM, Hamilton JP, Koteish A, Potter JJ, Brancati FL, Clark JM. Elevated ALT and GGT predict all-cause mortality and hepatocellular carcinoma in Taiwanese male: a case-cohort study. *Hepatol Int* 2013; **7**: 1040-1049 [PMID: 26202033 DOI: 10.1007/s12072-013-9476-6]

22 **Wong DK**, Yuen MF, Poon RT, Yuen JC, Fung J, Lai CL. Quantification of hepatitis B virus covalently closed circular DNA in patients with hepatocellular carcinoma. *J Hepatol* 2006; **45**: 553-559 [PMID: 16904225 DOI: 10.1016/j.jhep.2006.05.014]

23 **Hung YC**, Lin CL, Liu CJ, Hung H, Lin SM, Lee SD, Chen PJ, Chuang SC, Yu MW. Development of risk scoring system for stratifying population for hepatocellular carcinoma screening. *Hepatology* 2015; **61**: 1934-1944 [PMID: 25418332 DOI: 10.1002/hep.27610]

24 **Chinese Society of Hepatology, Chinese Medical Association**. [Consensus on secondary prevention of primary liver cancer (2021 version)]. *Zhonghua Gan Zang Bing Za Zhi* 2021; **29**: 216-226 [PMID: 33902188 DOI: 10.3760/cma.j.cn501113-20210210-00082]

25 **O'Mahony C**, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, Biagini E, Gimeno JR, Limongelli G, McKenna WJ, Omar RZ, Elliott PM; Hypertrophic Cardiomyopathy Outcomes Investigators. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J* 2014; **35**: 2010-2020 [PMID: 24126876 DOI: 10.1093/eurheartj/eht439]

**Footnotes**

**Institutional review board statement:** This study was approved by the Ethics Committee of Affiliated Hospital of North Sichuan Medical College and conducted in accordance with the declaration of Helsinki Principles.

**Informed consent statement:** The informed consent of this study was exempted by the Ethics Committee of Affiliated Hospital of North Sichuan Medical College.

**Conflict-of-interest statement:** All the authors report no relevant conflicts of interest for this article.

**Data sharing statement:** Technical appendix, statistical code, and dataset available from the corresponding author at email address. Participants gave informed consent for data sharing.

**STROBE statement:** The authors have read the STROBE Statement-checklist of items, and the manuscript was prepared and revised according to the STROBE Statement-checklist of items.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

**Provenance and peer review:** Unsolicited article; Externally peer reviewed.

**Peer-review model:** Single blind

**Peer-review started:** February 10, 2023

**First decision:** May 19, 2023

**Article in press:**

**Specialty type:** Oncology

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B

Grade C (Good): C, C

Grade D (Fair): 0

Grade E (Poor): E

**P-Reviewer:** Cainap C, Romania; Costache RS, Romania; Gad EH, Egypt **S-Editor:** Wang JJ **L-Editor:** A **P-Editor:** Wang JJ

**Figure Legends**



**Figure 1 Forest plot of variables in the diagnosis of hepatocellular carcinoma.** OR: Odd ratio; CI: Confidence interval; HBsAg: Hepatitis B surface antigen; TBA: Total bile acid; TBIL: Total bilirubin; GGT: Gamma-glutamyl transferase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; AFP: Alpha-fetoprotein; PIVKA-II: Protein induced by vitamin K absence or antagonist-II.

 

**Figure 2 Nomogram to predict the presence of hepatocellular carcinoma.** HBsAg: Hepatitis B surface antigen; TBA: Total bile acid; TBIL: Total bilirubin; GGT: Gamma-glutamyl transferase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; AFP: Alpha-fetoprotein; PIVKA-II: Protein induced by vitamin K absence or antagonist-II; HCC: Hepatocellular carcinoma.



**Figure 3 Diagnostic values of NSMC-hepatocellular carcinoma model, vitamin K absence or antagonist-II and alpha-fetoprotein in hepatocellular carcinoma patients and early-stage hepatocellular carcinoma patients.** A and B: Comparison of the area under the receiver operating characteristic curve between NSMC-hepatocellular carcinoma (HCC) model and alpha-fetoprotein (AFP), protein induced by vitamin K absence or antagonist-II (PIVKA-II) alone or both for HCC diagnosis in training cohort (A) and validation cohort (B); C and D: Compare the area under the operating characteristic curve of subjects using NSMC-HCC model and AFP, PIVKA-II alone or both for HCC diagnosis in the early-stage HCC in training cohort (C) and validation cohort (D). CI: Confidence interval; AFP: Alpha-fetoprotein; PIVKA-II: Protein induced by vitamin K absence or antagonist-II; HCC: Hepatocellular carcinoma; AUC: Area under receiver operating characteristic curve.

**Table 1 Characteristics of the study population (*n* = 2206)**

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Training cohort** | **Validation cohort** |
| **HCC (*n* = 496)** | **Non-HCC (*n* = 1243)** | **HCC (*n* = 156)** | **Non-HCC (*n* = 311)** |
| Age (yr) | 57.86 ± 11.89a | 53.84 ± 14.53 | 58.47 ± 11.87b | 54.33 ± 13.53 |
| Gender |  |  |  |  |
| Male, *n* (%) | 411 (82.9)a | 759 (61.1) | 131 (84.0)b | 169 (54.3%) |
| Female, *n* (%) | 85 (17.1)a | 484 (38.9) | 25 (16.0)b | 142 (45.7%) |
| PIVKA-II, mAU/mL | 1321.03 (117.91-9792.97)a | 24.04 (18.21-37.02) | 1337.95 (92.76-11380.07)b | 22.69 (16.92-32.41) |
| AFP, ng/mL | 178.65 (8.33-6474.00)a | 3.60 (1.80-11.75) | 145.45 (6.25-2439.10)b | 3.70 (1.90-8.75) |
| TBIL, μmol/L | 21.20 (15.60-31.85)a | 29.80 (15.70-127.50) | 21.20 (14.43-32.73)b | 27.40 (15.45-94.05) |
| GGT, IU/L | 130.00 (59.75-279.75)a | 78.20 (29.00-197.00) | 145.40 (60.75-272.75)b | 63.00 (25.00-168.35) |
| AST, U/L | 61.10 (38.00-110.00)a | 56.00 (28.00-178.00) | 60.00 (37.00-102.83)b | 49.00 (25.00-147.80) |
| ALT, U/L | 42.00 (26.00-70.00)a | 44.00 (21.00-210.85) | 43.00 (25.57-68.00)b | 37.00 (19.50-149.00) |
| TBA, μmol/L | 11.10 (4.50-27.45)a | 18.80 (4.10-123.95) | 10.20 (4.27-22.10)b | 16.20 (4.00-97.45) |
| ALB, g/L | 37.10 (32.30-41.40)a | 36.10 (30.00-42.20) | 37.75 (34.00-41.42)b | 36.80 (30.85-41.90) |
| HBsAg |  |  |  |  |
| Positive, *n* (%) | 388 (78.2)a | 555 (44.7) | 123 (78.8)b | 131 (42.1) |
| Negative, *n* (%) | 108 (21.8)a | 688 (55.3) | 33 (21.2)b | 180 (57.9) |
| Child Pugh Class |  |  |  |  |
| Class A, *n* (%) | 317 (63.9)a | 583 (46.9) | 103 (66.0)b | 155 (49.8) |
| Class B, *n* (%) | 151 (30.4)a | 430 (34.6) | 45 (28.8)b | 103 (33.1) |
| Class C, *n* (%) | 28 (5.6)a | 230 (18.5) | 8 (5.1)b | 53 (17.0) |

a*P* < 0.05, compared with non-hepatocellular carcinoma group in training cohort.

b*P* < 0.05, compared with non-hepatocellular carcinoma group in validation cohort.

Data are presented as mean standard deviation, median (interquartile range), or numbers (%). HBsAg: Hepatitis B surface antigen; TBA: Total bile acid; TBIL: Total bilirubin; GGT: Gamma-glutamyl transferase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; AFP: Alpha-fetoprotein; PIVKA-II: Protein induced by vitamin K absence or antagonist-II; HCC: Hepatocellular carcinoma; ALB: Albumin.

**Table 2 Alpha-fetoprotein, protein induced by vitamin K absence or antagonist-II, or both in the diagnosis of hepatocellular carcinoma**

|  |  |  |
| --- | --- | --- |
|  | **Training cohort** | **Validation cohort** |
| **AUC (95%CI)** | **SEN (%)** | **SPE (%)** | **PPV (%)** | **NPV (%)** | **AUC (95%CI)** | **SEN (%)** | **SPE (%)** | **PPV (%)** | **NPV (%)** |
| HCC |  |  |  |  |  |  |  |  |  |  |
| AFP | 0.812 (0.788-0.815) | 49.19 | 93.00 | 42.21 | 94.63 | 0.845 (0.786-0.880) | 52.56 | 91.00 | 49.04 | 92.09 |
| PIVKA-II | 0.882 (0.865-0.903) | 63.91 | 94.93 | 51.21 | 96.93 | 0.878 (0.839-0.916) | 64.74 | 94.53 | 57.35 | 95.93 |
| AFP + PIVKA-II | 0.896 (0.877-0.915) | 69.35 | 95.82 | 55.51 | 97.65 | 0.888 (0.851-0.924) | 67.95 | 95.17 | 59.83 | 96.56 |
| Early-stage HCC |  |  |  |  |  |  |  |  |  |  |
| AFP | 0.860 (0.792-0.928) | 51.06 | 97.49 | 82.77 | 89.40 | 0.714 (0.642-0.787) | 14.55 | 98.49 | 72.70 | 80.66 |
| PIVKA-II | 0.863 (0.790-0.936) | 40.43 | 96.98 | 75.97 | 87.33 | 0.868 (0.802-0.935) | 52.73 | 96.48 | 80.55 | 88.07 |
| AFP + PIVKA-II | 0.923 (0.869-0.978) | 63.83 | 96.98 | 83.31 | 91.90 | 0.869 (0.804-0.933) | 50.91 | 95.98 | 77.78 | 87.61 |

SEN: Sensitivity; SPE: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; CI: Confidence interval; PIVKA-II: Protein induced by vitamin K absence or antagonist-II; HCC: Hepatocellular carcinoma; AUC: Area under receiver operating characteristic curve; AFP: Alpha-fetoprotein.

**Table 3 Diagnostic performance of hepatocellular carcinoma, early hepatocellular carcinoma, and different subgroups of hepatocellular carcinoma in the NSMC-hepatocellular carcinoma model**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cut-off value** | **Training cohort** | **Validation cohort** |
| **AUC (95%CI)** | **SEN (%)** | **SPE (%)** | **PPV (%)** | **NPV (%)** | **AUC (95%CI)** | **SEN (%)** | **SPE (%)** | **PPV (%)** | **NPV (%)** |
| HCC | 0.22 | 0.960 (0.950-0.971) | 94.40 | 95.35 | 87.25 | 98.10 | 0.966 (0.945-0.986) | 90.00 | 94.20 | 89.87 | 95.47 |
| Early HCC |  | 0.946 (0.901-0.991) | 85.93 | 93.62 | 61.11 | 98.28 | 0.947 (0.901-0.994) | 89.10 | 98.49 | 94.23 | 97.02 |

SEN: Sensitivity; SPE: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; CI: Confidence interval; HCC: Hepatocellular carcinoma; AUC: Area under receiver operating characteristic curve; AFP: Alpha-fetoprotein.

**Table 4 Proposed risk scale and corresponding probability of predictive risk of hepatocellular carcinoma**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk level** | **Probability of risk** | **PPV (%)** | **NPV (%)** |
| Low risk | 0.000-0.007 | NA | ≥ 99.00 |
| Moderate risk | 0.008-0.220 | NA | < 99.00 |
| High risk | 0.221-0.940 | < 99.00 | NA |
| Most likely HCC | 0.941-1.000 | ≥ 99.00 | NA |

NA: Not applicable; PPV: Positive predictive value; NPV: Negative predictive value; HCC: Hepatocellular carcinoma.