International Journal of Epidemiology, 2020, 1562–1571
doi: 10.1093/ije/dyaa089
Advance Access Publication Date: 28 July 2020
Original Article

Cancer

Risk of hepatocellular carcinoma in individuals without traditional risk factors: development and validation of a novel risk score

Dong Hyun Sinn,[1†] Danbee Kang,[2†] Soo Jin Cho,[3] Seung Woon Paik,[1] Eliseo Guallar,[4,5] Juhee Cho[2,4,5†] and Geum-Youn Gwak [1*]

[1]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea, [2]Department of Clinical Research Design and Evaluation, SAIHST, Sungkyunkwan University, Seoul, South Korea, [3]Center for Health Promotion, Samsung Medical Center, Seoul, South Korea, [4]Center for Clinical Epidemiology, Samsung Medical Center, Sungkyunkwan University, Seoul, South Korea and [5]Departments of Epidemiology and Medicine and Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Medical Institutions, Baltimore, MD, USA

*Corresponding author. Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-Gu, 06351 Seoul, South Korea. E-mail: gy.gwak@samsung.com
[†]These authors contributed equally to this study.

Editorial decision 21 April 2020; Accepted 30 April 2020

Abstract

Background: Although hepatocellular carcinoma (HCC) occurs mostly in patients with chronic hepatitis B virus (HBV) or hepatitis C virus (HCV) infection or heavy alcohol use or cirrhosis, some patients develop HCC without these risk factors. Our objective in this study was to develop and validate a new HCC risk score that could stratify HCC risk in patients who develop HCC without known risk factors.

Methods: A new HCC risk score was developed using a nationwide, population-based cohort among individuals without chronic HBV infection, chronic HCV infection, heavy alcohol use or cirrhosis ($n = 467\ 206$, derivation cohort). The performance of the HCC risk score was validated using an independent Samsung Medical Center Health Promotion Center cohort ($n = 91\ 357$, validation cohort).

Results: Multivariable Cox regression analysis identified six independent risk factors: age, sex, smoking, diabetes, total cholesterol level and serum alanine aminotransferase level. A 19-point scale for HCC risk score was developed, with 10-year risk of HCC ranging from 0.0% to 6.16% for the lowest and highest risk scores, respectively. The area under the receiver operating characteristics curve values (AUROCs) to predict HCC development were 0.83 [95% confidence interval (CI): 0.77, 0.88)] and 0.92 (95% CI: 0.89, 0.95) at 10 years in the derivation and validation cohorts, respectively. Predicted risk was well correlated with the Kaplan-Meier observed HCC risk.

Conclusions: A simple-to-use, novel HCC risk score was developed for predicting HCC development in individuals without alleged risk factors. It can be used to assess the risk

of HCC in this population so that decisions about their clinical management, including risk reduction interventions, can be subsequently made.

Key words: Alanine aminotransferase, diabetes, hepatocellular carcinoma, national cohort, risk score

Key Messages

- Hepatocellular carcinoma (HCC) occurrence among individuals without alleged risk factors for HCC (chronic hepatitis B or hepatitis C virus infection or heavy alcohol use or cirrhosis) is rare at population level.
- Age, sex, smoking, diabetes, total cholesterol and alanine aminotransferase level are independent risk factors associated with HCC development among individuals without alleged risk factors for HCC.
- HCC development in individuals without traditional risk factors can be predicted by a simple-to-use, novel HCC risk score using these six variables.

Introduction

Hepatocellular carcinoma (HCC) is a unique cancer that develops mostly in patients with chronic hepatitis B virus (HBV) infection, chronic hepatitis C virus (HCV) infection, heavy alcohol use or liver cirrhosis.[1–4] In a study using Globoscan 2012 data, HBV and HCV potentially account for 94% of incident HCC in the world.[5] Until now, HCC risk assessment has been mainly focused on people with chronic HBV infection, chronic HCV infection, heavy alcohol use or cirrhosis, as HCC occurs mostly in people with these risk factors. However, the epidemiology of HCC is evolving. Prevalence of chronic HBV infection is decreasing with universal vaccination of newborns for HBV in many countries.[6,7] Effective antiviral therapies for chronic HBV infection and chronic HCV infection are available. They can effectively suppress viral replication or eradicate virus.[8,9] Hence, HBV- or HCV-related HCC is expected to decrease in a large part of the world. In contrast, non-alcoholic fatty liver disease (NAFLD) is rapidly increasing in many countries.[10,11] It is becoming the most rapidly growing risk factor for HCC in some countries.[12,13] In addition, several reports have suggested that HCC can develop in individuals with NAFLD in the absence of cirrhosis.[12,14] With an evolving epidemiology of HCC, the magnitude of HCC risk for the population without traditional risk factors is yet to be determined. Screening strategies for this population have not been established either.[15,16]

As the risk of HCC in people without traditional risk factors is low,[7] a tool to identify and stratify HCC risk is needed for this population. To date, there is no reliable tool to assess HCC risk in this population. Thus, the objective of the present study was to evaluate the risk of HCC development among individuals without chronic HBV infection, chronic HCV infection, heavy alcohol use or cirrhosis, using a nationwide population-based cohort. We developed and validated a new HCC risk score that could be used to stratify HCC risk in this population. It can also be used to make decisions about their further clinical management.

Methods

Ethical approval

The Institutional Review Board of the Samsung Medical Center approved this study and waived the requirement for informed consent as we used only de-identified data routinely collected during health screening visits.

Study population and design

To produce a robust tool to predict the risk of HCC development, data of two parallel populations were collected: National Health Insurance Service (NHIS)-National Sample Cohort (NSC) as derivation cohort, and Samsung Medical Center Health Promotion Center Cohort (SMC-HPCC) as external validation cohort.

The NHIS-NSC is a population-based retrospective cohort consisting of a representative sample of 2.2% of Korean citizens enrolled in the NHIS.[17] Sampling procedures and representativeness of the cohort have been described in detail previously.[17] In Korea, the NHIS also provides annual or biennial health screening examinations free of charge to all insured subjects.[18] For the derivation

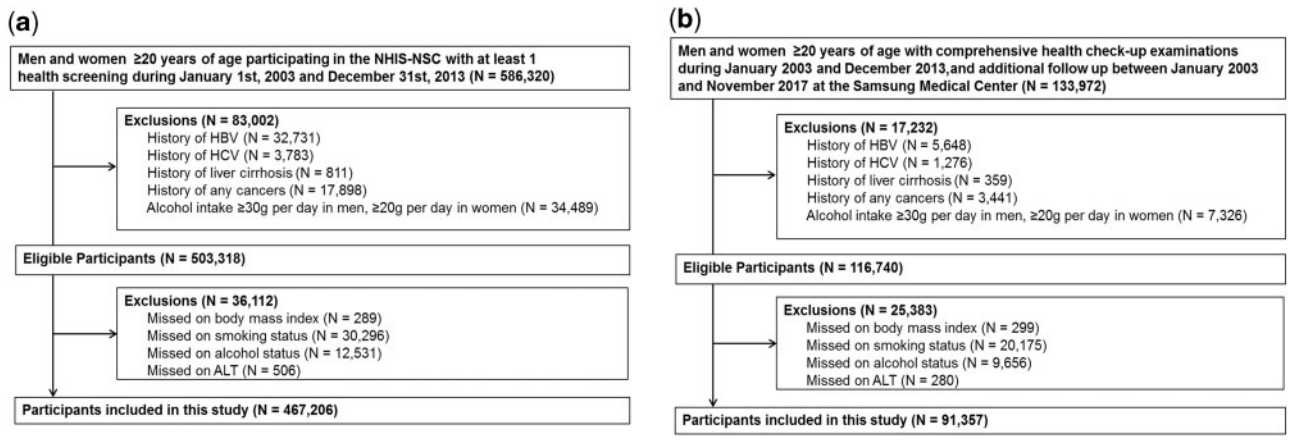

Figure 1 Flow chart for the derivation cohort (a) and the validation cohort (b). NHIS-NSC, National Health Insurance Service-National Sample Cohort; HBV, hepatitis B virus; HCV, hepatitis C virus; ALT, alanine aminotransferase.

cohort, we used person-level longitudinal NHIS-NSC registration, claim and health screening examination data recorded between 1 January 2002 and 31 December 2013.[17] Our study population included all men and women ≥20 years of age participating in the NHIS-NSC cohort with at least one health screening between 1 January 2003 and 31 December 2013 ($N = 586\,320$). We then excluded participants with HBV (ICD-10 codes: B18.0, B18.1, Z22.5) ($N = 32\,731$), HCV (ICD-10 codes: B18.2) ($N = 3783$), liver cirrhosis (ICD-10 codes: K74) ($N = 811$), any cancer (ICD-10 codes: including any C code) ($N = 17\,898$) and heavy drinking which was defined as alcohol intake ≥30 g per day in men and ≥20 g per day in women ($N = 34\,489$). After excluding participants who had missing data for body mass index (BMI) ($N = 289$), smoking status ($N = 30\,296$), alcohol status ($N = 12\,531$) and alanine aminotransferase (ALT) ($N = 506$), the final sample size was 467 206 (218 707 men and 248 499 women; Figure 1a).

The SMC-HPCC consisted of men and women ≥20 years of age who underwent comprehensive health check-up examinations between January 2003 and December 2013, and additional follow-up between January 2003 and November 2017 at Samsung Medical Center, Seoul, Korea ($N = 133\,972$). Among these participants, 17 232 were excluded according to the same exclusion criteria as the derivation cohort. A further 25 383 participants were additionally excluded because of missing values. After exclusions, the final sample size for the external validation cohort was 91 357 (43 178 men and 48 179 women; Figure 1b).

Data collection, definitions and endpoints

Databases of the NHIS-NSC cohort have been described in detail previously.[17] The primary outcome was HCC development during follow-up. HCC was defined as three or more outpatient clinic visits with associated C22.0 or C22.9 codes within a year, or one inpatient hospitalization with the same C code. In addition, a 1-year look-back window was used to exclude patients with a previous diagnosis of cancer (C code).[19]

To identify potential risk factors, we included the following variables: age,[20–22] sex,[20–22] BMI,[23] smoking status,[24] alcohol intake, exercise, hypertension, diabetes,[25] total cholesterol, dyslipidaemia and ALT level.[16] Data on smoking habits, exercise, history of diabetes and medication use were collected by self-administered questionnaires. Smoking status was categorized into never or past smoker and current smoker. Current alcohol consumption was categorized into none or modest (<30 g/day in men and <20 g/day in women). Height, weight and blood pressure were measured. BMI was calculated as weight in kilograms divided by height in metres squared.

BMI was classified according to Asian-specific criteria (underweight, BMI of <18.5 kg/m^2; normal weight, BMI of 18.5 to 22.9 kg/m^2; overweight, BMI of 23 to 24.9 kg/m^2; and obese, BMI ≥25 kg/m^2).[26] Glucose, total cholesterol and ALT levels were measured in fasting samples. Pre-hypertension was defined as a systolic blood pressure ≥130–<140 mmHg or a diastolic blood pressure ≥85–<90 mmHg at the baseline screening. Hypertension was defined as the presence of at least one I10–I13 or I15 code during the year preceding the screening, or a systolic blood pressure ≥140 mmHg or a diastolic blood pressure ≥90 mmHg at the baseline screening. Pre-diabetes was defined as a fasting glucose level of ≥100–<126 mg/dL at the baseline screening. Diabetes was defined as the presence of at least one E11–E14 code or a fasting glucose level of ≥126 mg/dL at the baseline screening. Dyslipidaemia was defined as the presence of an E78 code or a total cholesterol level of >240 mg/dL at the baseline screening. ALT

level was classified into three groups: low (<30 U/L for males and <20 U/L for females), mildly elevated (30–89 U/L for males and 20–59 U/L for females) and elevated (≥90 U/L for males and ≥ 60 U/L for females). The cutoff point of ALT was determined after receiver operation characteristics (ROC) curve analysis.

For the validation cohort, HCC development during follow-up period was identified by reviewing electronical medical records at Samsung Medical Center. HCC was diagnosed according to regional HCC guidelines.[27] Patients were censored at the latest follow-up date of any outpatient clinic visit or health checkup visit at Samsung Medical Center (reference date: 17 November 2017). The same variables (age, sex, BMI, smoking status, alcohol intake, exercise, hypertension, diabetes, total cholesterol, dyslipidaemia and ALT level) were collected in the validation set as well. Smoking status, alcohol consumption, exercise, medical history and medication use were collected through standardized, self-administered questionnaires. Height, weight and sitting blood pressure were measured by trained nurses. Pre-hypertension was defined as a systolic blood pressure ≥130–<140 mmHg or a diastolic blood pressure ≥85–<90 mmHg at the baseline screening examination. Hypertension was defined as a systolic blood pressure ≥140 mm Hg or a diastolic blood pressure ≥90 mmHg or current use of antihypertensive medications. Serum glucose was measured using the hexokinase/glucose-6-phosphate dehydrogenase method. Pre-diabetes was defined as a fasting glucose ≥100–<126 mg/dL at the baseline screening. Diabetes was defined as a fasting serum glucose ≥126 mg/dL or self-reported use of insulin or antidiabetic medications. Dyslipidaemia was defined as triglyceride level ≥150 mg/dl, high-density lipoprotein (HDL)-cholesterol level <40 mg/dl or the use of medication for dyslipidaemia. ALT was measured following the International Federation of Clinical Chemistry method. All derived categories variables, including BMI, smoking status, alcohol status and ALT, in the validation set were classified according to the same criteria used for the training set.

Statistical analysis

Person-years of follow-up were calculated from the enrolment date to the diagnosis date of HCC, date of death, or the latest date of follow-up. Cox proportional hazards models were used to estimate crude and multivariable-adjusted hazard ratios (HR) with 95% confidence intervals (CI) for risk predictors of HCC.

To develop risk prediction models for HCC, the regression coefficient of each risk predictor from the multivariable Cox proportional hazards model was divided by the regression coefficient of the lowest value and the resulting number was rounded to an integer value to generate each score.

To evaluate predictive accuracies of risk prediction models, area under the receiver operating characteristics curve values (AUROCs) were calculated at 10 years using flexible parametric survival models. To evaluate the discriminatory ability of each risk model, the observed cumulative HCC risk of three groups with very low, low and intermediate risk scores in the validation set were compared. For assessment of calibration, we compared observed HCC risk with the Kaplan-Meier method and predicted HCC risk with the modelled risk score. Kaplan-Meier estimates were plotted against the mean predicted risk in the group to form a calibration chart. Survival curves were generated by the Kaplan-Meier product-limit method and compared by log-rank test. We examined proportional hazards assumption using plots of log (-log) survival function and Schoenfeld residuals. All analyses were performed using STATA version 14 (StataCorp LP, College Station, TX, USA).

Results

Baseline characteristics of the derivation cohort are summarized in Table 1. During a median follow-up of 8.0 years (range, 0.2–11.0 years), 236 (0.1%) patients developed HCC. Cumulative HCC incidence rates at 3, 5 and 10 years were 0.02%, 0.03% and 0.07%, respectively. Age, sex, BMI, smoking, hypertension, diabetes, total cholesterol and ALT levels were associated with HCC occurrence during follow-up. In multivariable-adjusted analysis, age, sex, smoking, diabetes, total cholesterol and ALT levels were independent factors associated with HCC development. Using these six variables, we developed a 19-point scale HCC risk score (Table 2). The lowest score (0 points) showed a 10-year HCC incidence rate of 0%, and the highest score (18 points) showed a 10-year incidence rate of 6.16% (Table 3). The AUROC for developing HCC at 10 years was 0.83 (95% CI: 0.77, 0.88) (Figure 2a) and the C index was 0.80 (95% CI: 0.74, 0.87).

To assess calibration, patients were divided into three groups: very low risk group (score 0–6, less than 0.01% of cumulative incidence of HCC at 10 years), low risk group (score 7–15, cumulative incidence of HCC between 0.01% and 0.5% at 10 years) and intermediate risk group (score ≥16, cumulative incidence of HCC exceeding 0.5% at 10 years). Incidence rates at 5 and 10 years were 0.00% and 0.00% for a risk score of 0–6, 0.02% and 0.03% for a risk score of 7–15, and 0.28% and 1.57% for a risk score ≥16, respectively (Figure 3a; P <0.001 by log-rank test).

Table 1. Baseline characteristics according to hepatocellular carcinoma (HCC) development in the derivation cohort ($N = 467\,206$)

Characteristics	HCC development		*P*-value
	No ($N = 466\,970$)	Yes ($N = 236$)	
Age (years)			<0.001
<50	303 261 (64.9)	28 (11.9)	
50–59	82 192 (17.6)	40 (17.0)	
60–69	52 744 (11.3)	95 (40.3)	
≥70	28 773 (6.2)	73 (30.9)	
Sex			<0.001
Male	218 536 (46.8)	171 (72.5)	
Female	248 434 (53.2)	65 (27.5)	
Body mass index (kg/m^2)			0.001
Underweight (<18.5)	23 042 (4.9)	13 (5.5)	
Normal (18.5–<23)	198 018 (42.4)	74 (31.4)	
Overweight (23–<25)	108 201 (23.2)	53 (22.5)	
Obese (≥25)	137 709 (29.5)	96 (40.7)	
Smoking status			<0.001
Never	321 986 (69.0)	135 (57.2)	
Past	26 456 (5.7)	18 (7.6)	
Current	118 528 (25.4)	83 (35.2)	
Alcohol intake			0.052
None	337 605 (72.3)	184 (78.0)	
Modest	129 365 (27.7)	52 (22.0)	
Vigorous exercise (yes)	184 168 (39.4)	84 (35.6)	0.33
Hypertension			<0.001
No hypertension	278 928 (59.7)	70 (30.0)	
Pre-hypertension	77 643 (16.6)	47 (19.9)	
Hypertension	110 399 (23.6)	119 (50.4)	
Diabetes			<0.001
No diabetes	338 118 (72.4)	108 (45.8)	
Pre-diabetes	92 913 (19.9)	63 (26.7)	
Diabetes	35 939 (7.7)	65 (27.5)	
Total cholesterol (≥200 mg/dL)	181 649 (38.9)	52 (22.0)	<0.001
Dyslipidaemia	63 691 (13.6)	25 (10.6)	0.17
Alanine aminotransferase[a]			<0.001
Low	332 307 (71.2)	96 (40.7)	
Mildly elevated	127 125 (27.2)	111 (47.0)	
Elevated	7538 (1.7)	29 (12.3)	

Values in the table are number (%).

[a]Low: <30 U/L for males and <20 U/L for females; mildly elevated: 30–89 U/L for males and 20–59 U/L for females; elevated: ≥90 U/L for males and ≥60 U/L for females.

Baseline characteristics of the validation cohort are summarized in Supplementary Table S1, available as Supplementary data at *IJE* online. During a median follow-up of 4.6 years (range, 0.0–14.9 years), 35 (0.04%) patients developed HCC. In the validation cohort, the AUROC of the risk score for developing HCC at 10 years was 0.92 (95% CI: 0.89, 0.95) (Figure 2b). The HCC incidence rate increased with increasing risk scores (Table 3). Incidence rates at 5 and 10 years were 0% and 0.01% for risk score 0–6, 0.08% and 0.22% for risk score of 7–15, and 4.0% and 23.20% for risk score ≥16, respectively (Figure 3b; $P < 0.001$ by log-rank test).

We also plotted a calibration chart for predicted HCC risk and observed risk, showing good correlations in both derivation (Figure 4a) and validation cohorts (Figure 4b). The correlation coefficient was 0.9994 in the derivation cohort and 0.9983 in the validation cohort.

Discussion

We found that HCC occurrence among individuals without alleged risk factors for HCC (chronic HBV infection, chronic HCV infection, heavy alcohol use or cirrhosis) was rare at population level. The 10-year cumulative incidence

Table 2. Hepatocellular carcinoma (HCC) risk scoring system

	Hazard ratio (95% CI)	Beta coefficient	Risk score
Age			
<50	Reference		0
50–59	5.68 (3.47, 9.3)	1.737	3
60–69	20.65 (13.2, 32.3)	3.028	6
≥70	38.62 (23.9, 62.42)	3.654	7
Sex			
Female	Reference		0
Male	3.44 (2.46, 4.81)	1.236	2
Body mass index			
Underweight	1.78 (0.98, 3.22)	0.575	0
Normal	Reference		0
Overweight	0.95 (0.67, 1.36)	−0.047	0
Obesity	1.12 (0.81, 1.55)	0.115	0
Modest alcohol use (yes)	0.82 (0.58, 1.16)	−0.197	0
Smoking			
Never	Reference		0
Past	1.36 (0.81, 2.28)	0.306	0
Current	1.65 (1.20, 2.26)	0.498	1
Vigorous exercise (yes)	0.87 (0.67, 1.15)	−0.134	0
Hypertension			
No hypertension	Reference		0
Pre-hypertension	1.32 (0.91, 1.91)	0.275	0
Hypertension	1.28 (0.94, 1.75)	0.251	0
Diabetes			
No diabetes	Reference		0
Pre-diabetes	1.32 (0.96, 1.81)	0.278	0
Diabetes	1.97 (1.42, 2.72)	0.676	1
Total cholesterol (mg/dL)			
<200	3.27 (2.39, 4.46)	1.184	2
≥200	Reference		0
Dyslipidaemia (yes)	0.9 (0.58, 1.39)	−0.105	0
Alanine aminotransferase[a]			
Low	Reference		0
Mildly elevated	3.38 (2.54, 4.5)	1.219	2
Elevated	17.29 (11.26, 26.57)	2.85	6

[a]Low: <30 U/L for males and <20 U/L for females; mildly elevated: 30–89 U/L for males and 20–59 U/L for females; elevated: ≥90 U/L for males and ≥60 U/L for females.

rate of HCC was 0.07%. Age, sex, smoking, diabetes, total cholesterol and ALT level were independent risk factors associated with developing HCC. The performance of the HCC risk scoring system developed using these six variables was excellent, showing high AUROC values in both the derivation cohort and the validation cohort. According to this scoring system, individuals without risk factors for HCC could be categorized into very low risk (cumulative incidence rate of HCC <0.01% at 10 years), low risk (cumulative incidence rate of HCC between 0.01–0.5%) and intermediate risk (cumulative incidence rate of HCC ≥0.5%) groups. A new, simple-to-use HCC risk score can be used to estimate the risk of HCC development among individuals without traditional risk factors. This can be used in the primary care setting to screen and identify people at risk for HCC.

In this study, older age, male sex, current smoking, diabetes, low cholesterol and elevated ALT levels were independent risk factors associated with HCC. Several previous studies have shown that age and male sex are associated with increased HCC risk.[20–22] Smoking has been associated with HCC risk in a meta-analysis of 38 cohort and 58 case-control studies.[24] Diabetes was associated with increased HCC risk in a meta-analysis of 25 cohort studies.[25] Hypercholesterolaemia was associated with lower HCC risk in this study. Similarly, in a study of 400 318 adults, a high cholesterol level was associated with reduced risk of HCC.[28] ALT level has also been

shown to be correlated with HCC risk.[29] However, it is also noteworthy that HCC risk is low in this population. Thus, one cannot rely on a single risk factor. The cumulative incidence rate of HCC was less than 0.5% at 10 years for those with risk score of less than 15. The highest risk score by a single risk factor was age, which was 7 points for people aged more than 70 years. Hence, combinations of risk factors are needed to identify people at increased risk. People with multiple risk factors may warrant careful evaluation for their increased HCC risk. Also, they may be a target population for early intervention to decrease the risk of HCC. Among these six identified risk factors, smoking, diabetes, total cholesterol and ALT levels are modifiable risk factors.

Table 3. Risk estimation for hepatocellular carcinoma (HCC) development according to risk score

Risk score	Derivation cohort		Validation cohort	
	5 years	10 years	5 years	10 years
0	0.00%	0.00%	0.00%	0.00%
1	0.00%	0.00%	0.00%	0.00%
2	0.01%	0.01%	0.00%	0.00%
3	0.00%	0.00%	0.00%	0.00%
4	0.00%	0.00%	0.00%	0.00%
5	0.00%	0.00%	0.00%	0.00%
6	0.00%	0.00%	0.00%	0.07%
7	0.01%	0.01%	0.03%	0.03%
8	0.00%	0.01%	0.00%	0.00%
9	0.01%	0.03%	0.00%	0.37%
10	0.02%	0.04%	0.05%	0.41%
11	0.02%	0.03%	0.05%	0.13%
12	0.03%	0.08%	0.38%	0.54%
13	0.08%	0.22%	0.38%	0.38%
14	0.09%	0.19%	1.43%	1.43%
15	0.00%	0.23%	0.93%	3.77%
16	0.00%	0.52%	0.00%	20.00%
17	0.00%	0.50%	0.00%	0.00%
18	0.71%	6.16%	20.00%	
19	0.50%	0.50%		

In terms of HCC screening strategies, our data could not clarify whether those with high risk scores needed regular HCC surveillance. HCC surveillance requires multiple considerations, including the incidence of HCC in a given population, currently available resources, cost-effectiveness, performance of screening modalities and utilization of medical resources.[15] Given the limited screening resources, a threshold of HCC incidence rate of 1.5%/year has been suggested for HCC surveillance as at this point the screening practice becomes cost-effective.[30] In this study, HCC incidence rate did not reach this threshold even in the population with the highest risk score (about 0.6%/year). Hence, further stratification by adding new biomarkers is needed to ensure cost-effective screening for HCC in this population. Nevertheless, this risk score is useful in identifying subset of patients who are at relatively higher risk of HCC than others. Thus, it can be used as an initial step to identify people with HCC risk.

Our study has some other limitations. First, as the derivation cohort was based primarily on claims data from the NHIS database, we did not have detailed information on several important factors that could be associated with HCC risk, such as dietary patterns, abdominal circumference, insulin sensitivity, diabetes control and NAFLD. NAFLD is an important cause of HCC in individuals without chronic HBV or HCV infection, heavy alcohol use or cirrhosis.[31] For large epidemiological studies where biopsy or ultrasonography are not available, several serum biomarkers are often used as alternative tools to detect fatty liver.[32] Thus, we used ALT as a biomarker for NAFLD, as has been done in several previous studies.[33–35]

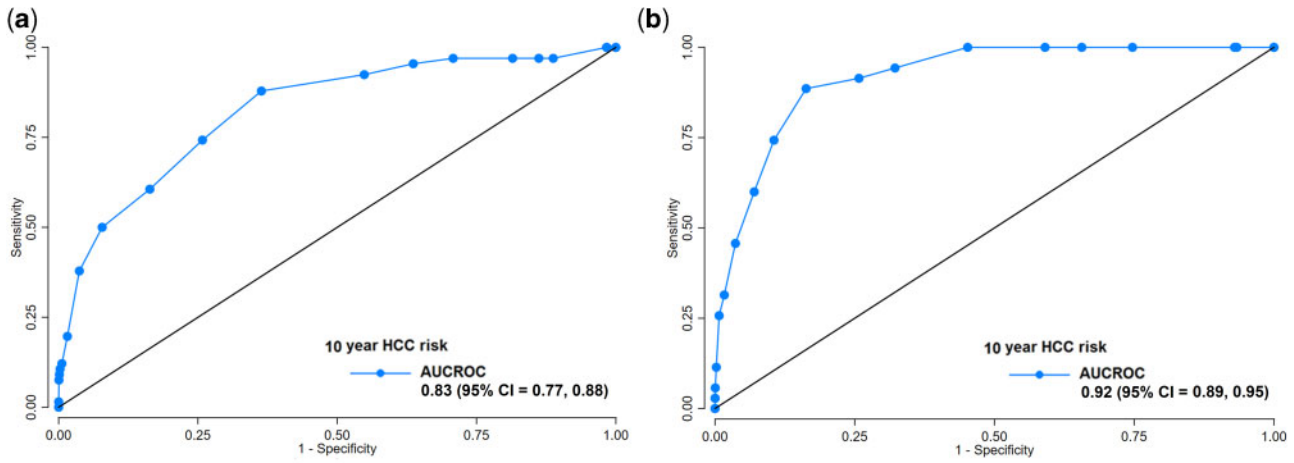

Figure 2 Area under the curve (AUC) of the risk score for developing hepatocellular carcinoma (HCC) at 10 years for the derivation cohort (a) and the validation cohort (b).

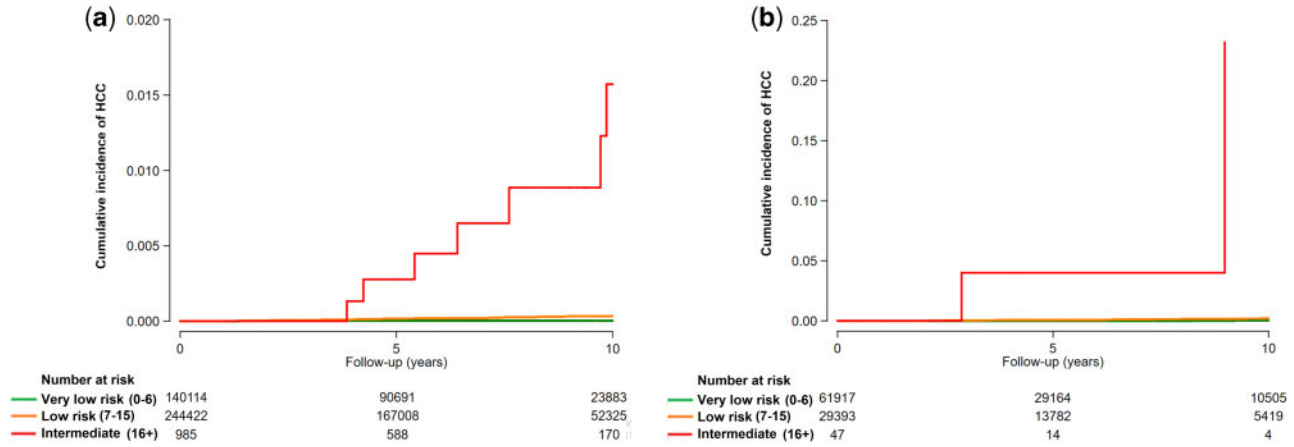

Figure 3 Kaplan-Meier curve for the development of hepatocellular carcinoma (HCC) in the derivation cohort (a) and the validation cohort (b).

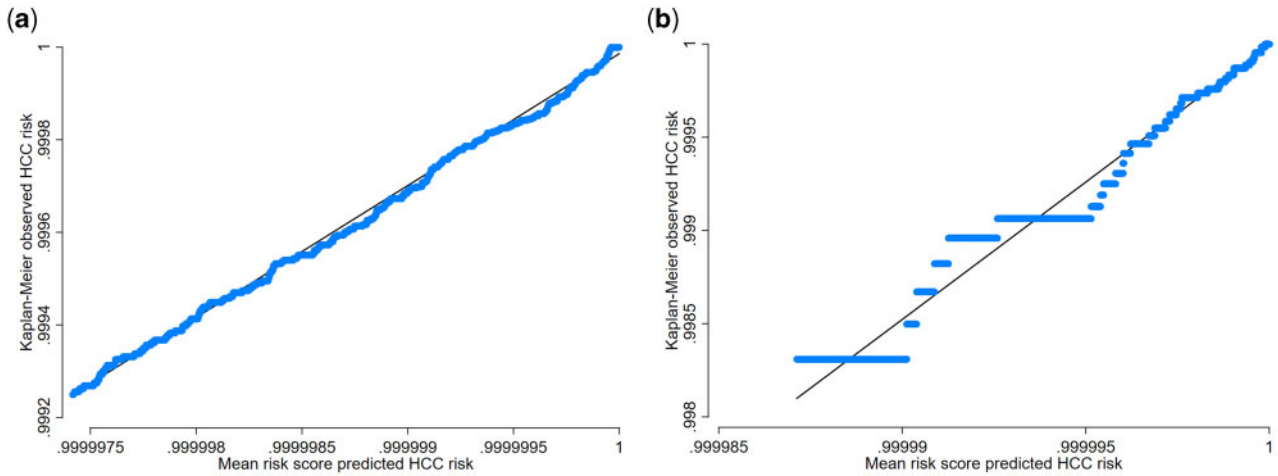

Figure 4 Calibration curve for the derivation cohort (a) and the validation cohort (b).

Second, the validation cohort was from a health screening cohort. Health screening cohorts are composed of those motivated by concern for their health. Hence, the scoring system needs further external validation using different cohorts from community as well. Third, BMI, smoking, and alcohol status can change over the follow-up period, and this may affect HCC risk. Fourth, for those who developed HCC, we lacked detailed information for possible causes of their chronic liver disease. Fifth, some important factors associated with HCC development, such as family history of HCC or previous HBV infection, were not fully captured in our risk model. Sixth, we used ICD codes to define HCC, hypertension, diabetes and dyslipidaemia in the derivation cohort, and these are prone to misclassification error. Last, this study is based on a Korean population. Korea is a rapidly urbanizing country with a high incidence and prevalence of NAFLD.[36] Obesity and diabetes are also emerging causes of liver cancer in Korea.[4] Still, significant differences exist between Eastern and Western populations in HCCa etiology.[37] For instance, hepatitis B core immunoglobuin G is positive in many Korean HCC patients without chronic HBV or HCV infection or heavy alcohol use,[38] whereas this is not a usual finding in Western patients.[39] In addition, the association between genetic variation and HCC risk may differ by ethnicity. Gene coding for the patatin-like phospholipase domain-containing 3 (PNPLA3) polymorphism is associated with HCC risk in Caucasians, but not in Asians.[40] Hence, the generalizability of our findings to other ethnicities needs to be assessed.

In summary, we developed a novel HCC risk scoring system that could be used for individuals without traditional risk factors (chronic HBV infection, chronic HCV infection, heavy alcohol use or cirrhosis). In this study, the occurrence of HCC in the absence of traditional risk factors was very low, but not null. Risk stratification is an initial step to identify at-risk individuals who may benefit from HCC surveillance using more expensive or invasive tests. Our risk score model is based on simple, easy-to-

obtain variables, and would be useful for primary screening purposes in daily practice.

Data sharing

Requests for access to data from the study should be addressed to the corresponding author at [gy.gwak@samsung.com]. All proposals requesting data access will need to specify how it is planned to use the data.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

D.H.S. and G.Y.G.designed the study. D.H.S, D.K., S.J.C. and S.W.P. collected the data. D.K., E.G. and J.C. performed the data analysis. D.H.S., D.K., S.W.P., E.G., J.C. and G.Y.G. wrote the final report. All authors contributed to critical revision of the final report. J.C. and G.Y.G. are guarantors. All the authors had full access to all of the data and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of interest

None declared.

References

1. Wallace MC, Preen D, Jeffrey GP, Adams LA. The evolving epidemiology of hepatocellular carcinoma: a global perspective. *Expert Rev Gastroenterol Hepatol* 2015;**9**:765–79.
2. Tiribelli C, Melato M, Croce LS, Giarelli L, Okuda K, Ohnishi K. Prevalence of hepatocellular carcinoma and relation to cirrhosis: comparison of two different cities of the world - Trieste, Italy, and Chiba, Japan. *Hepatology* 1989;**10**:998–1002.
3. Yang JD, Kim WR, Coelho R *et al.* Cirrhosis is present in most patients with hepatitis B and hepatocellular carcinoma. *Clin Gastroenterol Hepatol* 2011;**9**:64–70.
4. Kim BH, Park JW. Epidemiology of liver cancer in South Korea. *Clin Mol Hepatol* 2018;**24**:1–9.
5. Sartorius K, Sartorius B, Aldous C, Govender PS, Madiba TE. Global and country underestimation of hepatocellular carcinoma (HCC) in 2012 and its implications. *Cancer Epidemiol* 2015;**39**:284–90.
6. Cho EJ, Kim SE, Suk KT *et al.* Current status and strategies for hepatitis B control in Korea. *Clin Mol Hepatol* 2017;**23**:205–11.
7. Kulik L, El-Serag HB. Epidemiology and management of hepatocellular carcinoma. *Gastroenterology* 2019;**156**:477–91.e1.
8. Korean Association for the Study of the Liver (KASL). 2017 KASL clinical practice guidelines: management of hepatitis C: treatment of chronic hepatitis C. *Clin Mol Hepatol* 2018;**24**:169–229.
9. Korean Association for the Study of the Liver. KASL clinical practice guidelines: management of chronic hepatitis B. *Clin Mol Hepatol* 2016;**22**:18–75.
10. Korean Association for the Study of the Liver (KASL). KASL clinical practice guidelines: management of nonalcoholic fatty liver disease. *Clin Mol Hepatol* 2013;**19**:325–48.
11. Younossi ZM. Non-alcoholic fatty liver disease—a global public health perspective. *J Hepatol* 2019;**70**:531–44.
12. Mittal S, El-Serag HB, Sada YH *et al.* Hepatocellular carcinoma in the absence of cirrhosis in United States veterans is associated with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2016;**14**:124–31.e1.
13. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology* 2014;**59**:2188–95.
14. Piscaglia F, Svegliati-Baroni G, Barchetti A *et al.*; HCC-NAFLD Italian Study Group. Clinical patterns of hepatocellular carcinoma in nonalcoholic fatty liver disease: a multicenter prospective study. *Hepatology* 2016;**63**:827–38.
15. Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* 2018;**68**:526–49.
16. Singal AG, El-Serag HB. Hepatocellular carcinoma from epidemiology to prevention: translating knowledge into practice. *Clin Gastroenterol Hepatol* 2015;**13**:2140–51.
17. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. *Int J Epidemiol* 2017;**46**:e15.
18. National Health Insurance Service (NHIS). *National Health Examination Statistical Yearbook*. Seoul: National Health Insurance Service, 2014.
19. Sinn DH, Kang D, Kang M *et al.* Late presentation of hepatitis B among patients with newly diagnosed hepatocellular carcinoma: a national cohort study. *BMC Cancer* 2019;**19**:286.
20. Yang HI, Yuen MF, Chan HL *et al.* Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. *Lancet Oncol* 2011;**12**:568–74.
21. Yang HI, Sherman M, Su J *et al.* Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *J Clin Oncol* 2010;**28**:2437–44.
22. Wen CP, Lin J, Yang YC *et al.* Hepatocellular carcinoma risk prediction model for the general population: the predictive power of transaminases. *J Natl Cancer Inst* 2012;**104**:1599–611.
23. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* 2014;**384**:755–65.
24. Lee YC, Cohet C, Yang YC, Stayner L, Hashibe M, Straif K. Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. *Int J Epidemiol* 2009;**38**:1497–511.
25. Wang C, Wang X, Gong G *et al.* Increased risk of hepatocellular carcinoma in patients with diabetes mellitus: a systematic review and meta-analysis of cohort studies. *Int J Cancer* 2012;**130**:1639–48.

26. Kim MK, Lee WY, Kang JH *et al.* Clinical practice guidelines for overweight and obesity in Korea. *Endocrinol Metab* 2014;**29**: 405–09.
27. Korean Liver Cancer Study Group, National Cancer Center Korea. KLCSG-NCC Korea practice guideline for the management of hepatocellular carcinoma. *Gut Liver* 2014;**9**:267–317. 2015;
28. Yi SW, Kim SH, Han KJ, Yi JJ, Ohrr H. Higher cholesterol levels, not statin use, are associated with a lower risk of hepatocellular carcinoma. *Br J Cancer* 2020;**122**:630–33.
29. Hung YC, Lin CL, Liu CJ *et al.* Development of risk scoring system for stratifying population for hepatocellular carcinoma screening. *Hepatology* 2015;**61**:1934–44.
30. Bruix J, Sherman M; Practice Guidelines Committee AAftSoLD. Management of hepatocellular carcinoma. *Hepatology* 2005;**42**: 1208–36.
31. Heimbach JK, Kulik LM, Finn RS *et al.* AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 2018;**67**: 358–80.
32. European Association for the Study of the Liver, European Association for the Study of Diabetes, European Association for the Study of Oncology. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016;**64**:1388–402.
33. Fraser A, Longnecker MP, Lawlor DA. Prevalence of elevated alanine aminotransferase among US adolescents and associated factors: NHANES 1999-2004. *Gastroenterology* 2007;**133**: 1814–20.
34. Dunn W, Xu R, Schwimmer JB. Modest wine drinking and decreased prevalence of suspected nonalcoholic fatty liver disease. *Hepatology* 2008;**47**:1947–54.
35. Dunn W, Xu R, Wingard DL *et al.* Suspected nonalcoholic fatty liver disease and mortality risk in a population-based cohort study. *Am J Gastroenterol* 2008;**103**:2263–71.
36. Park SH, Plank LD, Suk KT *et al.* Trends in the prevalence of chronic liver disease in the Korean adult population, 1998-2017. *Clin Mol Hepatol* 2020;**26**:209–15.
37. Choo SP, Tan WL, Goh BKP, Tai WM, Zhu AX. Comparison of hepatocellular carcinoma in Eastern versus Western populations. *Cancer* 2016;**122**:3430–46.
38. Kim J, Kang W, Sinn DH *et al.* Potential etiology, prevalence of cirrhosis, and mode of detection among patients with non-B non-C hepatocellular carcinoma in Korea. *Korean J Intern Med* 2020;**35**:65–78.
39. Lok AS, Everhart JE, Di Bisceglie AM *et al.*; HALT-C Trial Group. Occult and previous hepatitis B virus infection are not associated with hepatocellular carcinoma in United States patients with chronic hepatitis C. *Hepatology* 2011;**54**:434–42.
40. Li JF, Zheng EQ, Xie M. Association between rs738409 polymorphism in patatin-like phospholipase domain-containing protein 3 (PNPLA3) gene and hepatocellular carcinoma susceptibility: Evidence from case-control studies. *Gene* 2019;**685**:143–48.