

Prognostic Nomograms for Prediction of Recurrence and Survival After Curative Liver Resection for Hepatocellular Carcinoma

Ju Hyun Shim, MD, PhD,* Mi-Jung Jun, MD,† Seungbong Han, PhD,‡ Young-Joo Lee, MD, PhD,§ Sung-Gyu Lee, MD, PhD,§ Kang Mo Kim, MD, PhD,* Young-Suk Lim, MD, PhD,* and Han Chu Lee, MD, PhD*

Objective: To develop clinical predictive nomograms generating per-patient numerical probabilities of postoperative recurrence-free and overall survival at specific times.

Background: The prognosis after surgical resection is diverse in patients with early-stage hepatocellular carcinoma (HCC).

Methods: In a retrospective review, we evaluated data from 1085 mostly early-stage patients newly diagnosed with HCC who were subsequently treated by curative resection. We randomly divided the subjects into derivation ($n = 760$) and validation ($n = 325$) samples. Multivariate Cox proportional hazards models were developed and separately validated on the basis of pre- and postoperative clinical and pathological covariates assessed for association with 2-year recurrence and 5-year HCC-specific death. The discriminatory accuracy of the models was compared with traditional tools by analyzing receiver operating characteristic curves.

Results: The statistical nomograms built on the basis of sex, serum albumin, platelet count, microvascular invasion, and calculated tumor volume had good calibration and discriminatory abilities, with c-indices of 0.69 (2-year recurrence) and 0.66 (5-year survival), respectively. These models showed satisfactory goodness-of-fit and discrimination abilities in the independent validation cohort (c-index, 0.66 for 2-year recurrence; and 0.67 for 5-year survival). The areas under the receiver operating characteristic curve using our methods were greater than those of conventional staging systems in the validation patients, indicating better discriminatory capability (corresponding c-indices, 0.55–0.56; and 0.55–0.61, respectively).

Conclusions: Our simple user-friendly calculators, which present graphically postsurgical prognostic models for recurrence and survival outcomes in patients with curatively resectable HCC, offer useful guidance to clinicians and patients for individually planning recurrence surveillance and adjuvant therapy.

Keywords: hepatocellular carcinoma, hepatectomy, prognostic nomogram, recurrence, survival

(*Ann Surg* 2015;261:939–946)

Hepatocellular carcinoma (HCC) is a major health issue, being ranked third among leading causes of death from cancer worldwide.¹ Although the major burden of HCC is still in the Asian/Pacific Islands where hepatitis B is endemic, an increase in North America is altering the global pattern of HCC incidence.² Hepatic resection remains the best therapeutic option for potential curative outcomes, although less than a third of HCC cases are suitable for it at the time of diagnosis.³ However, a serious disadvantage of resection with regard to achieving cure and long-term survival is the high rate of recurrence, which exceeds 60% at 5 years even in patients with small tumors.⁴ The prognosis of so-called “early-stage” HCC, defined, for example, as Barcelona Clinic Liver Cancer (BCLC) stage 0 or A and considered generally curatively resectable, is far from homogenous.^{5–7} Therefore, it is reasonable to seek to subdivide this stage for purposes of counseling patients and individualizing surveillance and follow-up after surgery.

Cho et al⁸ have proposed a prognostic nomogram to estimate postoperative outcomes more accurately in patients with HCC. However, this instrument was based on data from a cohort that included patients who, according to current guidelines, are generally not robust candidates for hepatic resection (ie, with extrahepatic and portal venous invasion), along with some who did not meet the definition of R0 resection of all visible tumors. We still need well-established clinicopathological indicators with demonstrated ability to predict tumor recurrence and patient survival specifically in patients with early-stage HCC.

The purpose of this study was to construct and independently validate prognostic scoring systems for recurrence and HCC-specific death on the basis of the clinicopathologic data from a large cohort of mostly early-stage patients for whom the indication for surgery is optimal, who underwent curative partial hepatectomy for resectable HCC. In addition, we compared the accuracy of these nomograms for estimating per-patient prognosis with that achieved by traditional staging systems.

METHODS

Study Patients

In a retrospective review of the source population in the database of the Asan Liver Center, in which detailed data elements were uniformly available, we identified 1085 patients with HCC with well-preserved hepatic function (no history of decompensation) who underwent curative hepatectomy with tumor-negative resection margins (R0 resection) as a first-line antitumor treatment between January 1997 and December 2009. None of the patients had received presurgical cancer treatment or were suffering from a recurrence of HCC or from any other known malignancy, and the HCC had not invaded the portal or hepatic veins or metastasized to distant sites in any of the patients. To develop the prognostic nomograms and validate them on the basis of a split-sample approach,^{9,10} the patients were randomly divided into a derivation set ($n = 760$) and a validation ($n = 325$) set. Given the relatively large sample size of the validation set,

From the *Department of Gastroenterology, Asan Liver Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea;

†Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; ‡Department of Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; and §Department of Surgery, Asan Liver Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea.

Disclosure: The authors indicated no potential conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Ju Hyun Shim, MD, PhD, Department of Gastroenterology, Asan Liver Center, Asan Medical Center, University of Ulsan College of Medicine, 388-1 Poongnap-2dong, Songpa-gu, Seoul, 138-736, Korea. E-mail: s5854@medimail.co.kr.

Copyright © 2014 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/14/26105-0939

DOI: 10.1097/SLA.0000000000000747

with numbers of events per variable exceeding 20 for recurrence and HCC-specific death, split-sample validation was a reasonable way to estimate the performance of the models.¹⁰ Preoperative clinical and postoperative pathologic characteristics of the 2 sets are given in detail in Table 1. Median follow-up times for the derivation and validation cohorts were 59.0 months (interquartile range, 36.0–76.0) and 58.0 months (interquartile range, 37.0–74.9), respectively. Ethical approval for our research protocol was obtained from the institutional review board of our hospital.

Preoperative and Postoperative Assessment

The evaluation and management approaches employed by our surgical team before and after surgical resection were as previously described.¹¹ Briefly, the presurgical procedures included liver-protocol dynamic computed tomography and/or magnetic resonance imaging, chest computed tomography, and bone scans, in addition to measurement of serum parameters related to the etiology of liver disease and α -fetoprotein (AFP).

For all patients, postoperative routine follow-up included dynamic liver computed tomography, covering most of the chest, and laboratory tests and serum AFP analysis. These assessments were performed 1 month after resection and then generally at 3-month intervals in the first 2 years and every 3 to 6 months in subsequent years, until tumor recurrence was documented.

Statistical Analysis

In the initial phase of nomogram development, we fitted a univariate Cox proportional hazards model to the recurrence endpoints and survival endpoints directly related to recurrent HCC and the associated hepatic complications (ie, HCC-specific survival) to measure univariate associations and examine the linearity of covariates as part of the exploratory analysis. Tumor volume was calculated from the expression: $4/3 \times 3.14$ (maximum diameter of the nodule in cm)³, taking nodules to be spherical. In patients with multiple lesions, the total tumor volume was calculated as the sum of the volumes of all the nodules.^{12,13} To select prognostic factors, we employed a 2-step variable selection approach. The first step was to fit a random survival forest model to compute a variable importance score, and the second step was to compute a relative selection frequency on the basis of a bootstrap resampling method. The central elements of the random survival forest algorithm are growing the survival tree and constructing the ensemble cumulative hazard function. By assuming randomness for variable selection and resampling, the random survival forest method improves predictive ability. Random survival forests are widely used for variable selection because they produce a variable importance score.^{14,15} The main advantage of the random survival forest algorithm is that it does not use restrictive assumptions such as proportional hazards and parametric or linear effects of the variables. Therefore, it can be used flexibly for variable selection as an initial step.^{16,17} The importance score is a metric of how much the prediction error rate of a model is improved by addition of each variable (more influential factors have higher scores). Specifically, the importance score of a covariate x is the prediction error for the original ensemble trees subtracted from the prediction error for the new ensemble trees obtained using randomizing x assignments. Here, we computed the prediction error using Harrell's c-index, and the splitting rule we employed is the log-rank rule that splits tree nodes maximizing the log-rank statistics. In addition, all default options for the fitting were employed. To obtain some robustness for variable selection, we used the bootstrap resampling method. That is, we fitted an automated backward variable selection with respect to the Cox proportional hazards model and computed the relative selection frequency for 500 bootstrap samples. Here, the candidate variables for the Cox model were the variables whose relative importance scores

obtained from the random survival forest method exceeded 10%. The criterion for inclusion in the final prediction model was a relative frequency of more than 80%. Our statistician and clinical researchers determined the final variables together on the basis of the clinical knowledge and the statistical results.

Although we can predict an individual's risk from the average of the estimated cumulative hazard function obtained from the random survival forest model, it is not possible to summarize the constructed trees. This issue is related to the interpretation of the final prediction model. Because the Cox model is widely used and easy to interpret, we chose it as the base model, and the survival random forest method was used only for the first screening for variable selection. The proportional hazards assumption was also examined by assessing the log hazard ratio over time on the basis of the scaled Schoenfeld partial residuals. The R function cox.zph proposed by Therneau was used for this.¹⁸ Because the sample size is large, the formal test for assessing the correlation of time points and the scaled residuals is likely to lead to significant results. Therefore, it is important to examine the log hazard ratio plot (approximated by the sum of the scaled Schoenfeld residuals and the estimated regression coefficient). We assessed the overfitting issue by estimating slope shrinkage where values of less than 0.85 were regarded as overfitting. We also assessed multicollinearity using the variation inflation factor and used a threshold of 10 to determine multicollinearity. More importantly, we measured both discrimination and calibration abilities for the development data and the validation data on the basis of the concordance indices (c-indices) and the Hosmer-Lemeshow test ($P > 0.20$ indicating good calibration). We compared the proposed prediction model with the updated American Joint Commission on Cancer (AJCC)/Union Internationale Contre le Cancer (UICC) staging system,¹⁹ BCLC and Cancer of the Liver Italian Program (CLIP) models, and the prognostic nomogram proposed by Cho et al⁸ on the basis of receiver operating characteristic (ROC) curve analysis. The P value for the c-indices in the 2 models was computed using a bootstrapping method. All analyses were carried out with R software version 2.15.1 (<http://www.r-project.org>). For the random survival forest method and the time-dependent c-index measurements, we used the "randomSurvivalForest"²⁰ and "survivalROC" packages²¹ in R software. In addition, both the "rms"²² and the "mfp"²³ packages were used for deriving user-friendly nomograms and checking the linearity assumption of the final prediction model. Our previous publications provide further details of prediction model building.^{24,25}

RESULTS

Presurgical Characteristics of Patients

Clinical, pathological, and surgical demographics of the derivation and validation sets are summarized in Table 1. Of the 1085 patients with a mean age of 53.2 years (SD, 9.8 years), 78.2% were male, 63.3% had a body mass index of 23 or higher, and 79.7% were positive for hepatitis B. Pathological examination showed that more than half of the patients had cirrhotic livers surrounding the hepatic tumor, and 100% of the latter had Child-Pugh class A liver function whereas more than 85% were positive for hepatitis B. Mean serum AFP was 6085.9 ng/mL (SD, 38868.5 ng/mL). With regard to tumor factors, single tumors were noted in most patients, the average size of intrahepatic tumors was 5.1 cm (SD, 3.4 cm) and the average sum of calculated tumor volumes was 1529.0 ± 3849.0 cm³. Capsular infiltration, microscopic vascular invasion, and satellite nodules were recorded histologically in 166 (15.3%), 191 (17.6%), and 59 (5.4%) patients, respectively. A majority of the patients had Edmondson grade III or IV tumors. In terms of operation factors, major hepatectomy defined as resection of 3 or more liver segments was performed in 444 patients (40.9%), and resection margins of

TABLE 1. Basal Characteristics*

Characteristics	Overall Cohort (n = 1085)	Derivation Set (n = 760)	Validation Set (n = 325)	P
<i>Host factors</i>				
Age, yr	53.2 ± 9.8	53.6 ± 9.8	52.3 ± 9.8	0.09
Sex (male:female)	849 (78.2):236 (21.8)	586 (77.1):174 (22.9)	263 (80.9):62 (19.1)	0.17
BMI, kg/m ²				0.90
<23	398 (36.7)	278 (36.6)	120 (36.9)	
≥23	687 (63.3)	482 (63.4)	205 (63.1)	
<i>Etiology</i>				
HBV	865 (79.7)	601 (79.1)	264 (81.3)	
HCV	67 (6.2)	50 (6.6)	17 (5.2)	
Others	153 (14.1)	109 (14.3)	44 (13.5)	
<i>Pathologic evidence of cirrhosis</i>				
AFP, ng/mL	6085.9 ± 38868.5	5844.0 ± 38646.6	6651.5 ± 39436.7	0.69
AST, IU/L	44.2 ± 29.1	44.4 ± 30.4	43.8 ± 26.0	0.82
ALT, IU/L	41.9 ± 28.8	41.8 ± 29.1	42.1 ± 28.0	0.55
PLT, ×10 ³ /uL	172.2 ± 68.7	173.8 ± 71.0	168.3 ± 62.9	0.55
Albumin, g/dL	3.82 ± 0.45	3.82 ± 0.45	3.82 ± 0.47	0.74
Bilirubin, IU/L	0.92 ± 0.37	0.91 ± 0.36	0.94 ± 0.41	0.47
Creatinine, mg/dL	0.92 ± 0.58	0.90 ± 0.30	0.95 ± 0.94	0.40
Prothrombin time, INR	1.07 ± 0.10	1.07 ± 0.09	1.07 ± 0.11	0.65
<i>Tumor factors</i>				
Tumor volume, cm ³	1529.0 ± 3849.0	1549.0 ± 3823.0	1482.1 ± 3914.5	0.68
Tumor diameter, cm	5.1 ± 3.4	5.1 ± 3.4	5.0 ± 3.3	0.71
No. of tumors				0.12
1	1027 (94.6)	724 (95.3)	303 (93.3)	
2	51 (4.7)	32 (4.2)	19 (5.8)	
3	5 (0.5)	4 (0.5)	1 (0.3)	
≥4	2 (0.2)	0 (0)	2 (0.6)	
Microscopic vascular invasion	191 (17.6)	138 (18.2)	53 (16.3)	0.49
Capsular invasion	166 (15.3)	118 (15.5)	48 (14.8)	0.78
Satellite nodules	59 (5.4)	36 (4.7)	23 (7.1)	0.14
Edmondson grade				0.21
I or II	212 (19.5)	141 (18.6)	71 (21.8)	
III or IV	873 (80.5)	619 (81.4)	254 (78.2)	
AJCC/UICC stage				0.86
I	848 (78.2)	595 (78.3)	253 (77.8)	
II	217 (20.0)	152 (20.0)	65 (20.0)	
IIIA	20 (1.8)	13 (1.7)	7 (2.2)	
CLIP score				0.48
0	637 (58.7)	453 (59.6)	184 (56.6)	
1	327 (30.1)	221 (29.1)	106 (32.6)	
2	81 (7.5)	60 (7.9)	21 (6.5)	
3	40 (3.7)	26 (3.4)	14 (4.3)	
BLCL stage				0.12
0	146 (13.5)	105 (13.8)	41 (12.6)	
A	904 (83.3)	636 (83.7)	268 (82.5)	
B	35 (3.2)	19 (2.5)	16 (4.9)	
<i>Surgical factors</i>				
Resection type†				0.46
Major	444 (40.9)	317 (41.7)	127 (39.1)	
Minor	641 (59.1)	443 (58.3)	198 (60.9)	
Resection margin width				0.21
<10 mm	580 (53.5)	416 (54.7)	164 (50.5)	
≥10 mm	505 (46.5)	344 (45.3)	161 (49.5)	
Operation time				0.83
<5 hr	738 (68)	515 (67.8)	223 (68.6)	
≥5 hr	347 (32)	245 (32.2)	102 (31.4)	
Red cell transfusion	163 (15)	122 (16.1)	41 (12.6)	0.16

*Values are presented as no. (%) or mean ± SD.

†Major resection means more than 2 segmentectomies according to the Couinaud classification and minor resection means 2 or less segmentectomies.

AFP indicates α-fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BLCL, Barcelona-Clinic Liver Cancer; BMI, body mass index; CLIP, Cancer of the Liver Italian Program; HBV, hepatitis B virus; HCV, hepatitis C virus; INR, international normalized ratio; PLT, platelet.

less than 10 mm were observed in 580 patients (53.5%). A total of 738 (68.0%) patients had operations lasting less than 5 hours, and 163 (15.0%) required blood transfusion during the perioperative period. Patients were classified according to the conventional staging systems as 848/217/20 patients with stage I/II/IIIA by the updated AJCC/UICC staging system; 637/327/81/40 patients with scores of 0/1/2/3 by the CLIP scoring system; and 146/904/35 patients with stage 0/A/B by the BCLC staging system.

There were no significant differences between the derivation and validation groups with respect to any patient-, tumor- or operation-related covariate (Table 1).

Development of the Models From the Derivation Cohort

As shown in Figures 1A and C and Supplemental Digital Content Table 1, available at <http://links.lww.com/SLA/A573>, the median recurrence-free survival time of the derivation set was 46.1 months, with postoperative 1-, 3-, and 5-year recurrence-free survival rates of 74.1%, 54.5%, and 46.1%, respectively. The 3-, 5-, and 7-year HCC-specific survival rates were 83.5%, 72.8%, and 65.8%, respectively, with a median survival of 134 months.

Supplemental Digital Content Figure 1, available at <http://links.lww.com/SLA/A570>, presents the variable clustering plot according to the squares of the Spearman correlation coefficients. This shows that tumor volume and tumor size are perfectly correlated. Microvascular invasion status, along with aspartate aminotransferase and alanine aminotransferase levels, is also strongly correlated pair. In the

univariate Cox proportional hazards model, most of variables are quite significant except AFP, age, serum bilirubin level, capsular invasion, resection margin width, body mass index, and operation time. It seems that the directions of the relative risks were correct (see Supplemental Digital Content Table 2, available at <http://links.lww.com/SLA/A573>) and no separation problems occurred. Note, however, that we cannot select variables on the basis of the univariate analysis alone because of possible significant interactions between them. The random survival forest model automatically takes care of the possible interactions between variables and reflects them in the importance scores. As shown in Supplemental Digital Content Table 3, available at <http://links.lww.com/SLA/A573>, sex, AFP on a logarithmic scale, aspartate aminotransferase, platelet count, serum albumin level, total tumor volume on a logarithmic scale, and microvascular invasion had high relative importance (>0.1) in recurrence-free survival and HCC-specific survival. We should point out that tumor size and alanine aminotransferase were not included in the random survival forest fitting because of their strong correlations with other variables. Based on the relative selection frequencies derived from the 500 bootstrap samples (see Supplemental Digital Content Table 4, available at <http://links.lww.com/SLA/A573>), we selected sex, serum albumin level, platelet count, microvascular invasion, and total tumor volume on a logarithmic scale as prognostic factors for both recurrence-free survival and HCC-specific survival in the derivation cohort.

Nomograms incorporating each of these clinical predictors were constructed on the basis of the Cox model. Table 2 gives the estimated regression coefficients for the recurrence and the survival models in the final Cox proportional hazards model. When the

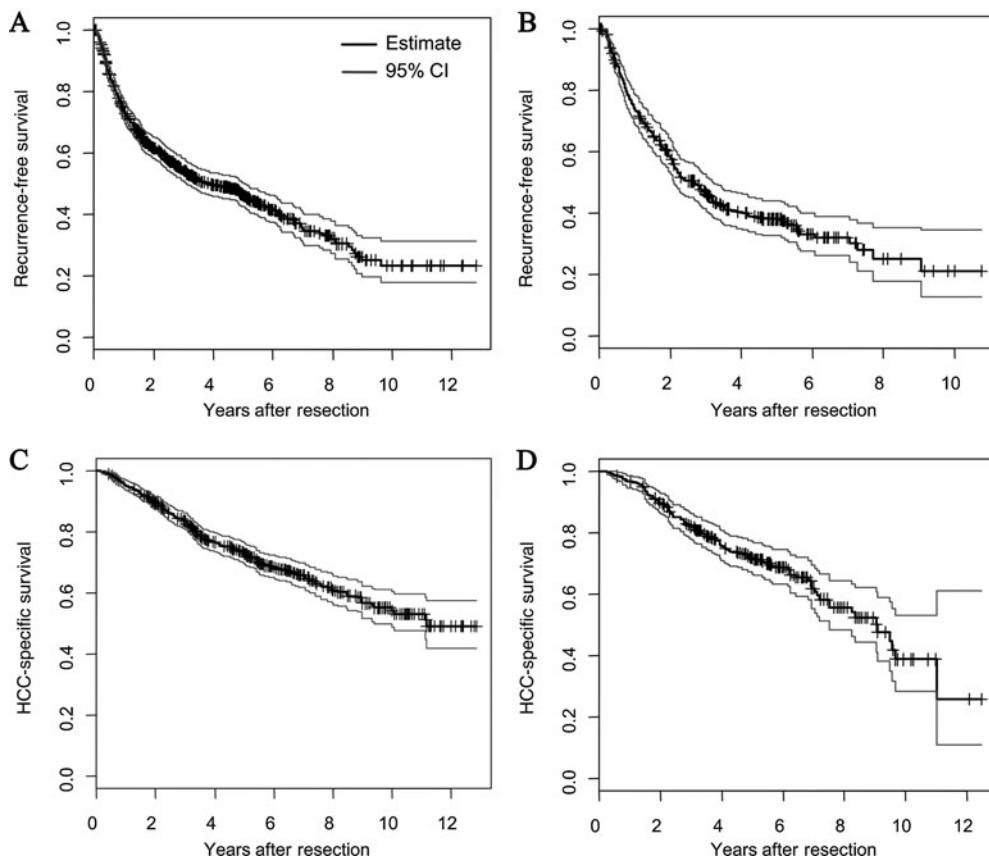


FIGURE 1. Kaplan-Meier estimates of (A and B) recurrence-free survival and (C and D) HCC-specific survival, respectively, in the derivation and validation samples.

TABLE 2. Multivariable Regression Results for Recurrence and Survival (n = 760)

Variable	2-Year Recurrence-free Survival			5-Year HCC-specific Survival		
	Estimated Coefficient	HR (95% CI)	P	Estimated Coefficient	HR (95% CI)	P
Male	0.555146	1.742 (1.334–2.275)	<0.001	0.394047	1.483 (1.048–2.098)	0.026
Platelet count	-0.00434	0.996 (0.994–0.997)	<0.001	-0.00307	0.997 (0.995–0.999)	0.002
Serum albumin	-0.62221	0.537 (0.434–0.665)	<0.001	-0.50238	0.605 (0.465–0.787)	<0.001
Log (tumor volume)	0.205353	1.228 (1.154–1.306)	<0.001	0.267817	1.307 (0.120–1.419)	<0.001
Microvascular invasion	0.430764	1.538 (1.212–1.953)	<0.001	0.533383	1.705 (1.262–2.305)	0.001

HR indicates hazard ratio.

proportionality assumption was examined, there were no serious violations because the smooth lines for recurrence-free survival and HCC-specific survival shown in Supplemental Digital Content Figure 2, available at <http://links.lww.com/SLA/A571>, and Supplemental Digital Content Figure 3, available at <http://links.lww.com/SLA/A572>, are horizontal across time. With the finally selected variables, the slope shrinkage estimates for recurrence and survival were 0.96 and 0.94, respectively, which means that overfitting was not a problem for the model. Because the variation inflation factor values of the predictors in the prediction model were less than 10, we concluded that multicollinearity did not exist. We subsequently used the derivation clinical data set to build nomogram-based scoring systems for determining likelihoods of 2-year recurrence and 5-year HCC-specific survival, both of which integrated all the aforementioned covariates (Figs. 2A, B). The c-indices for the specific models predicting 2-year recurrence probability were 0.69 [95% confidence interval (CI), 0.66–0.73] for the derivation group. For the nomogram predicting postresection 5-year survival specific for HCC, the c-index value was 0.66 (95% CI, 0.61–0.71).

Calibration plots based on the 5 predictors are shown in Figures 3A and C. There was good agreement between actual- and nomogram-predicted probabilities after resection, with a Hosmer-Lemeshow statistic of 1.86 with 3 df ($P = 0.60$) for 2-year recurrence, and 2.60 with 3 df ($P = 0.46$) for 5-year HCC-specific death. These predicted scores can be calculated automatically, as shown in Supplemental Digital Content Appendix 1, available at <http://links.lww.com/SLA/A574>.

Validation of Prognostic Accuracy

As a final method of validation, the probabilities of clinical outcomes were predicted for the separate sample of 325 patients. In this validation cohort, median recurrence-free survival and HCC-specific survival times were 32 months and 109 months, respectively (Figs. 1B and D). The c-indices of the nomograms for predicting postoperative recurrence at 2 years and HCC-specific survival at 5 years were 0.66 (95% CI, 0.59–0.72) and 0.67 (95% CI, 0.60–0.73), respectively. The calibration curves yielded good agreement between expected and observed outcomes for 2-year recurrence and 5-year HCC-specific death (Figs. 3B and D). The proposed nomograms were well-calibrated according to the Hosmer-Lemeshow goodness-of-fit test ($P = 0.64$ for recurrence-free survival and $P = 0.35$ for HCC-specific survival), indicating that both models were adequately fitted.

Comparison of Discriminatory Powers

The predictive power of the nomograms and conventional staging systems including the model of Cho et al⁸ were compared by ROC curve analysis (Figs. 4A, B). Our nomogram displayed significantly better discriminatory accuracy in predicting 2-year recur-

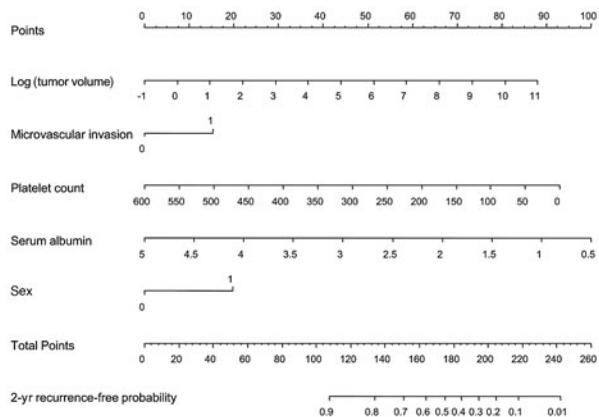
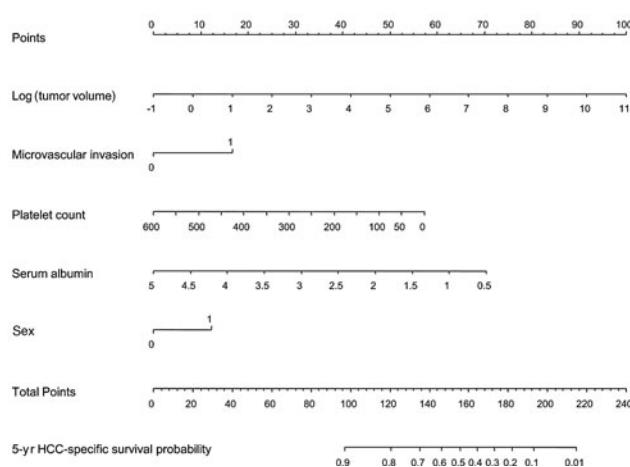
A**B**

FIGURE 2. Nomograms for predicting (A) recurrence-free probability (2 years) and (B) HCC-specific survival probability (5 years). (Sex: score 0 for female and 1 for male patients; and microvascular invasion: score 0 for absence and 1 for presence.)

rence in the validation cohort than the competing models: its c-index was 0.66 (95% CI, 0.59–0.72), substantially higher than those of the AJCC/UICC [0.56 (95% CI, 0.51–0.61); $P = 0.005$], CLIP [0.55 (95% CI, 0.48–0.61); $P = 0.01$], and BCLC [0.56 (95% CI, 0.52–0.60); $P = 0.003$] systems, and the model of Cho et al [0.58 (95% CI, 0.51–0.65); $P = 0.037$]. In terms of 5-year HCC-specific death, the

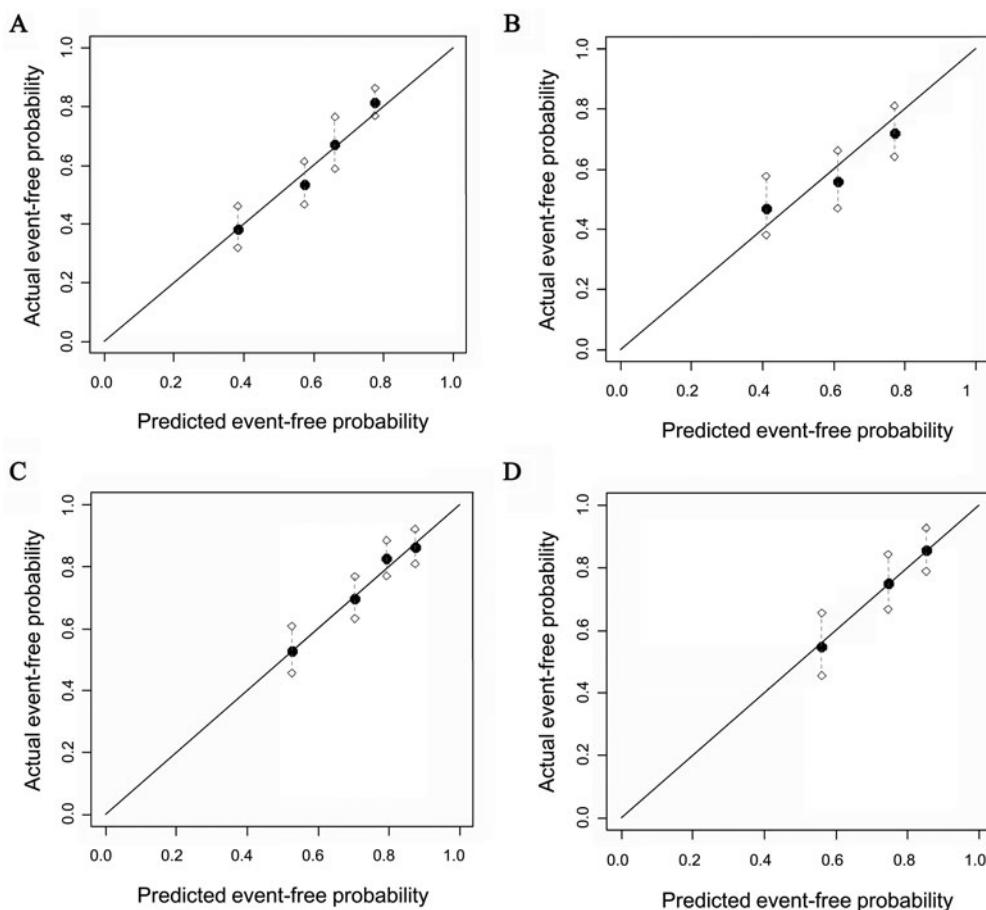


FIGURE 3. Calibration of the predictive models in the derivation and validation cohorts. The calibration curves for (A and B) 2-year recurrence and (C and D) 5-year HCC-specific survival yield good agreement between predicted and observed outcomes in the derivation set and the separate validation set [Hosmer-Lemeshow test: $P = 0.60$ (A) and 0.64 (B) for recurrence-free survival; 0.46 (C) and 0.35 (D) for HCC-specific survival, respectively].

c-index of our nomogram was 0.66 (95% CI, 0.59–0.72), significantly higher than those of the AJCC/UICC [0.55 (95% CI, 0.49–0.62); $P = 0.003$] and BCLC [0.55 (95% CI, 0.50–0.60); $P < 0.001$] systems but not significantly different from the CLIP system [0.61 (95% CI, 0.55–0.67); $P = 0.19$] and the model by Cho et al [0.64 (95% CI, 0.57–0.72); $P = 0.43$].

DISCUSSION

On the basis of our large series of patients who had undergone radical hepatectomy with negative resection margins for HCC, we have created statistical predictive nomograms tailored to the individual patient and capable of reliably generating numerical probabilities of recurrence at 2 years and HCC-specific death at 5 years. These graphical tools are simple and easy-to-use calculators, integrating 5 predictive variables that constitute the essentials of preoperative clinical evaluation and postoperative pathologic outcomes. Both predictive systems have c-indices of more than 0.65, and thus are not perfect, but they are more reliable than the traditional staging systems commonly used (c-indices, 0.55–0.61). Given the lack of consensus on the follow-up strategies for detection of recurrent HCC after resection,²⁶ these prediction models enable surgical patients to be easily monitored on an individual basis and to be appropriately allocated for participation in clinical trials of postoperative adjuvant therapy (eg,

patients with total risk scores for recurrence of ≥ 165 , in whom the probability of 2-year recurrence-free survival is $\leq 60\%$).

Even early HCC, which is generally indicated for surgical resection, is in fact quite variable in terms of tumor structure and hepatic function and eventual outcome. However, the AJCC/UICC staging system based on the TNM classification does not include underlying liver function, which has an important influence on the survival of patients with HCC, and some patients have rather poor liver functions despite having early-stage tumors. In addition, the CLIP system has less stringent criteria for tumor classification with a somewhat subjective measure of tumor burden, which may prevent accurate postoperative risk stratification in surgical patients. In addition, the comprehensive and detailed BCLC staging system, one of the systems with the greatest ability to predict survival and the most widely used of the current systems,^{27–29} was also less effective than our prognostic calculator in predicting both recurrence and survival after HCC resection. Our model is more powerful for predicting recurrence than the prognostic nomogram previously constructed by Cho et al,⁸ which was applied to a Western cohort of surgical patients with a broader spectrum of HCCs, although it had comparable efficacy with respect to survival endpoint.

Total tumor volume, estimated on the assumption that all tumors are spherical, although some tumors are irregularly shaped or infiltrative, has been shown to be a preoperative predictor of tumor

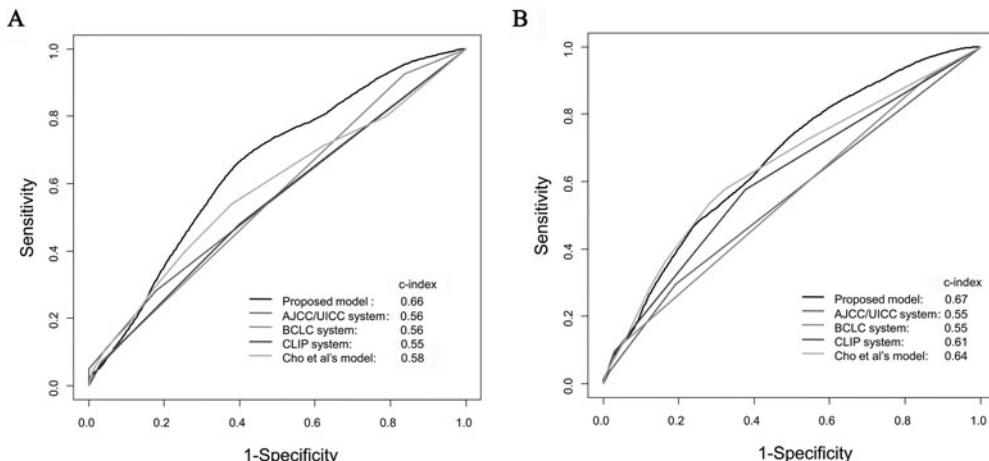


FIGURE 4. ROC analysis of (A) recurrence-free survival at 2 years and (B) HCC-specific survival at 5 years in the validation cohort using the nomograms, and the AJCC/UICC staging, BCLC classification, CLIP scoring systems, and the previous model of Cho et al.⁸ The c-index value for our model developed for recurrence-free survival was significantly higher than those for the AJCC/UICC, BCLC, and CLIP systems, and the model of Cho et al for recurrence-free survival [0.66 vs 0.56 ($P = 0.005$), 0.56 ($P = 0.003$), 0.55 ($P = 0.01$), and 0.58 ($P = 0.037$), respectively]. In terms of HCC-specific survival, the c-index of our nomogram was significantly higher than those of the AJCC/UICC and BCLC systems [0.67 vs 0.55 ($P = 0.003$) and 0.55 ($P < 0.001$), respectively] but not significantly different from the CLIP system and the model of Cho et al for HCC-specific survival [0.67 vs 0.61 ($P = 0.19$) and 0.64 ($P = 0.43$), respectively]. All P values were obtained from analyses comparing the c-indices of the nomograms and the relevant models.

grade and microvascular invasion of the HCC.¹² Importantly, a prediction model based on calculated tumor volume previously proposed by Hsu et al¹³ had superior prognostic ability to other contemporary systems in various strata of patients with HCC, constituting only about 26% of hepatectomized cases with inconsistent tumor burden and residual liver function and accordingly different treatment modalities.¹³ In keeping with previous findings, the summed volume of all tumors reflecting overall tumor burden is included in our proposed model.

The low platelet count noted in a considerable proportion of patients with cirrhosis is a well-known indicator of portal hypertension, which accounts for most of the complications of liver cirrhosis, along with hypersplenism and gastroesophageal varices.³⁰ In addition, quite different serum albumin levels often exist among patients with liver disease, all of whom are of Child-Pugh class A. The nomograms built using data from our series with well-preserved liver function, which were consistently correlated with postresection radiographic and survival outcomes, also relied on such blood tests although they excluded other components of the Child-Pugh score (ie, serum bilirubin and prothrombin time) from the calculators, probably because the values obtained were far below the lower limit of normal in almost all cases. In this regard, we believe that our scoring systems for estimating individualized prognoses after resection may be of more help to clinicians and patients for accurately predicting and more thoroughly preparing for potential recurrence than the conventional models, even those that include Child-Pugh class as a prognostic covariate (ie, CLIP and BCLC models).

Interestingly, sex is a predictive covariate related to tumor recurrence and patient survival after resection in our models, unlike preoperative serum AFP, which we did not find to be useful in predicting clinical endpoints in terms of recurrence and HCC death after resection in a test based on propensity scores.³¹ It is now well known that HCC is more prevalent in men, particularly in the Asian Pacific region, which is an endemic area of chronic hepatitis B, although the reason is still obscure.³² The male predominance of HCC recurrence

after surgery in our report is likely to be at least partially linked with this hepatocarcinogenic feature; hence, further evidence is required to confirm the causative association between sex and reoccurrence of HCC in resected patients.

The presence of microscopic venous invasion in surgical specimens is known to be an important cause of intrahepatic recurrence, and it is the most consistently reported risk factor for recurrent HCC after resection.³³ As expected, it is a key variable in our nomograms. Conventional systems based on pretreatment radiographic data, rather than pathologic estimates, cannot take this strong pathologic parameter into consideration in predicting postsurgical clinical outcomes.⁸ Moreover, AJCC/UICC staging based on pathology had poorer predictive and discriminative accuracy for both recurrence and survival after surgery than our postoperative nomograms.

This study has the limitation that it is based on a single institutional data set of patients with HCC in a hepatitis B virus-endemic area. It will certainly be necessary to validate these predictive nomograms in other geographic regions to extend our results to patients with HCC of various etiologies. Another major limitation is that our predictive tools are clinically useful for postoperative decision making rather than for preoperative decision making when careful consideration based on powerful predictive models might replace surgical intervention with nonsurgical alternatives. The anticipated future availability of molecular or genomic pathological biomarkers, especially for microvascular involvement of tumor cells, would substantially enhance the utility of our nomograms.

CONCLUSIONS

We have generated unique and consistent postoperative short- and long-term outcome prediction models from what we believe to be the largest homogeneous cohort data set yet available, which can provide specific information about individual patients treated with curative partial hepatectomy for early-phase HCC. The application of our models yielded much better estimates of both recurrence and survival than traditional systems including BCLC staging. These types

of nomograms will become increasingly important when effective adjuvant regimens for after-HCC resection are available and will further advance the date of personalized medicine for patients with HCC.

ACKNOWLEDGMENTS

Ju Hyun Shim, Mi-Jung Jun, and Seungbong Han contributed equally to this article.

REFERENCES

- Yang JD, Roberts LR. Hepatocellular carcinoma: a global view. *Nat Rev Gastroenterol Hepatol*. 2010;7:448–458.
- El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology*. 2012;142:1264–1273.e1261.
- Lau WY, Lai EC. Hepatocellular carcinoma: current management and recent advances. *Hepatobiliary Pancreat Dis Int*. 2008;7:237–257.
- Poon RT, Fan ST, Lo CM, et al. Long-term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function: implications for a strategy of salvage transplantation. *Ann Surg*. 2002;235:373–382.
- Yang XR, Xu Y, Yu B, et al. High expression levels of putative hepatic stem/progenitor cell biomarkers related to tumour angiogenesis and poor prognosis of hepatocellular carcinoma. *Gut*. 2010;59:953–962.
- Fan ST, Mau Lo C, Poon RT, et al. Continuous improvement of survival outcomes of resection of hepatocellular carcinoma: a 20-year experience. *Ann Surg*. 2011;253:745–758.
- Poon RT, Fan ST, Lo CM, et al. Long-term prognosis after resection of hepatocellular carcinoma associated with hepatitis B-related cirrhosis. *J Clin Oncol*. 2000;18:1094–1101.
- Cho CS, Gonan M, Shia J, et al. A novel prognostic nomogram is more accurate than conventional staging systems for predicting survival after resection of hepatocellular carcinoma. *J Am Coll Surg*. 2008;206:281–291.
- Picard RR, Berk KN. Data splitting. *Am Statistician*. 1990;44:140–147.
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774–781.
- Lee SG, Hwang S. How I do it: assessment of hepatic functional reserve for indication of hepatic resection. *J Hepatobiliary Pancreat Surg*. 2005;12:38–43.
- Cucchetti A, Piscaglia F, Grigioni AD, et al. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study. *J Hepatol*. 2010;52:880–888.
- Hsu CY, Huang YH, Hsia CY, et al. A new prognostic model for hepatocellular carcinoma based on total tumor volume: the Taipei Integrated Scoring System. *J Hepatol*. 2010;53:108–117.
- Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. *Ann Appl Stat*. 2008;2:841–860.
- Weichselbaum RR, Ishwaran H, Yoon T, et al. An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc Natl Acad Sci U S A*. 2008;105:18490–18495.
- Datema FR, Moya A, Krause P, et al. Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. *Head Neck*. 2012;34:50–58.
- Diller GP, Giardini A, Dimopoulos K, et al. Predictors of morbidity and mortality in contemporary Fontan patients: results from a multicenter study including cardiopulmonary exercise testing in 321 patients. *Eur Heart J*. 2010;31:3073–3083.
- Therneau TMGP. Strategies for nonproportional data. In: *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000:142–147.
- Greene FL, American Joint Committee on Cancer. Liver (including intrahepatic bile ducts). In: *AJCC Cancer Staging Manual*. 6th ed. New York, NY: Springer; 2002:131–138.
- Hemant I, Udaya BK. Package “randomSurvivalForest.” Version 3.6.4. 2013.
- Patrick JH, Paramita S-C. Package “survivalROC.” Version 1.0.3. 2013.
- Frank EHJ. Package “rms.” Version 3.6–3. 2013.
- Gareth A, Axel B. Package “mfp.” Version 1.4.9. 2013.
- Park HW, Han S, Lee JS, et al. Risk stratification for advanced proximal colon neoplasm and individualized endoscopic screening for colorectal cancer by a risk-scoring model. *Gastrointest Endosc*. 2012;76:818–828.
- Jin YJ, Lim YS, Han S, et al. Predicting survival after living and deceased donor liver transplantation in adult patients with acute liver failure. *J Gastroenterol*. 2012;47:1115–1124.
- EASL-EORTC clinical practice guidelines. Management of hepatocellular carcinoma. *J Hepatol*. 2012;56:908–943.
- Cillo U, Vitale A, Grigoletto F, et al. Prospective validation of the Barcelona Clinic Liver Cancer staging system. *J Hepatol*. 2006;44:723–731.
- Grieco A, Pompili M, Caminiti G, et al. Prognostic factors for survival in patients with early-intermediate hepatocellular carcinoma undergoing non-surgical therapy: comparison of Okuda, CLIP, and BCLC staging systems in a single Italian centre. *Gut*. 2005;54:411–418.
- Guglielmi A, Ruzzennete A, Pachera S, et al. Comparison of seven staging systems in cirrhotic patients with hepatocellular carcinoma in a cohort of patients who underwent radiofrequency ablation with complete response. *Am J Gastroenterol*. 2008;103:597–604.
- de Franchis R, Baveno VF. Revising consensus in portal hypertension: report of the Baveno V consensus workshop on methodology of diagnosis and therapy in portal hypertension. *J Hepatol*. 2010;53:762–768.
- Shim JH, Yoon DL, Han S, et al. Is serum alpha-fetoprotein useful for predicting recurrence and mortality specific to hepatocellular carcinoma after hepatectomy? A test based on propensity scores and competing risks analysis. *Ann Surg Oncol*. 2012;19:3687–3696.
- Wands J. Hepatocellular carcinoma and sex. *N Engl J Med*. 2007;357:1974–1976.
- Tung-Ping Poon R, Fan ST, Wong J. Risk factors, prevention, and management of postoperative recurrence after resection of hepatocellular carcinoma. *Ann Surg*. 2000;232:10–24.