ARTICLE

Development and Validation of Insulin-like Growth Factor-1 Score to Assess Hepatic Reserve in Hepatocellular Carcinoma

Ahmed O. Kaseb, Lianchun Xiao, Manal M. Hassan, Young Kwang Chae, Ju-Seog Lee, Jean-Nicolas Vauthey, Sunil Krishnan, Sheree Cheung, Hesham M. Hassabo, Thomas Aloia, Claudius Conrad, Steven A. Curley, John M. Vierling, Prasun Jalal, Kanwal Raghav, Michael Wallace, Asif Rashid, James L. Abbruzzese, Robert A. Wolff, Jeffrey S. Morris

Correspondence to: Ahmed O. Kaseb, MD, Department of Gastrointestinal Medical Oncology, Unit 426, University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030 (e-mail: akaseb@mdanderson.org).

Background   Child-Turcotte-Pugh (CTP) score is the standard tool to assess hepatic reserve in hepatocellular carcinoma (HCC), and CTP-A is the classic group for active therapy. However, CTP stratification accuracy has been questioned. We hypothesized that plasma insulin-like growth factor 1 (IGF-1) is a valid surrogate for hepatic reserve to replace the subjective parameters in CTP score to improve its prognostic accuracy.

Methods   We retrospectively tested plasma IGF-1 levels in the training set (n = 310) from MD Anderson Cancer Center. Recursive partitioning identified three optimal IGF-1 ranges that correlated with overall survival (OS): greater than 50 ng/mL = 1 point; 26 to 50 ng/mL = 2 points; and less than 26 ng/mL = 3 points. We modified the CTP score by replacing ascites and encephalopathy grading with plasma IGF-1 value (IGF-CTP) and subjected both scores to log-rank analysis. Harrell's C-index and U-statistics were used to compare the prognostic performance of both scores in both the training and validation cohorts (n = 155). All statistical tests were two-sided.

Results   Patients' stratification was statistically significantly stronger for IGF-CTP than CTP score for the training (P = .003) and the validation cohort (P = .005). Patients reclassified by IGF-CTP relative to their original CTP score were better stratified by their new risk groups. Most important, patients classified as A by CTP but B by IGF-CTP had statistically significantly worse OS than those who remained under class A by IGF-CTP in both cohorts (P = .03 and P < .001, respectively, from Cox regression models). AB patients had a worse OS than AA patients in both the training and validation set (hazard ratio [HR] = 1.45, 95% confidence interval [CI] = 1.03 to 2.04, P = .03; HR = 2.83, 95% CI = 1.65 to 4.85, P < .001, respectively).

Conclusions   The IGF-CTP score is simple, blood-based, and cost-effective, stratified HCC better than CTP score, and validated well on two independent cohorts. International validation studies are warranted.

Functional liver reserve is an important predictor of outcome in hepatic diseases. Therefore, a comprehensive and accurate evaluation of the liver reserve is crucial to predicting patients' survival and treatment outcome. An early attempt to develop a system for evaluating liver reserve was the Child-Turcotte score in 1964 (1), which involved two objective variables (serum bilirubin and albumin) and three subjective variables (severity of ascites, encephalopathy, and nutritional status). In 1973, Child and Pugh modified the score by replacing the most subjective variable (nutritional status) with an objective test (prothrombin time). The resulting Child-Turcotte-Pugh (CTP) score was originally intended as an assessment of life expectancy in the setting of transection of the esophagus for portal hypertensive variceal bleeding in cirrhotic patients (2). Eventually, however, the CTP score became the standard method for evaluating hepatic reserve and predicting life expectancy in patients with

cirrhosis. Because cirrhosis underlies most cases of hepatocellular carcinoma (HCC) and advanced cirrhosis can, in fact, affect patients' survival to a greater degree than the carcinoma itself (3,4), the CTP score has become the standard prognostic tool for predicting survival and for assessing hepatic reserve to guide initial or subsequent therapy decisions by predicting risk of liver failure and death after local and systemic therapies and for categorizing patients under HCC staging systems for trial entry (5).

The five CTP variables are each scored on a scale of 1 to 3 points; thus, the minimum score is 5 and the maximum score is 15 (Table 1). The lowest scores (scores 5 and 6) are considered CTP class A, which carries the best prognosis; the middle scores (scores 7–9) are class B, and the highest scores (scores 10–15) are class C. Because survival rates are universally low for CTP classes B and C compared with class A, multiple expert panels have reached the consensus that patients with HCC should have

a CTP score of A to be considered for aggressive therapies to facilitate assessment of the effect of treatment without the confounding issues of liver failure and death as a result of underlying poor hepatic reserve (3,4). However, it is now recognized that clinical outcome can vary considerably among patients within the same CTP class. Furthermore, CTP is partially based on subjective assessment of empiric dynamic clinical parameters (hepatic encephalopathy and ascites) with arbitrary cutoff ranges that are difficult to grade subjectively and may vary in severity according to nutritional status, comorbidities, and in response to medical management (6–8). Therefore, the CTP score's reliability for survival prediction and clinical decision-making was questioned, and more objective liver scores were introduced (9,10), including the objective Model for End-Stage Liver Disease (MELD) score, which replaced CTP to stratify patients for prioritization for orthotopic liver transplantation. However, none was developed for patients with HCC. Other limitations of the CTP score include the disproportionate number of patients under class A and its need for more accurate objective markers that reflect hepatic reserve (6).

The insulin-like growth factor (IGF) family comprises two ligands (IGF-1 and IGF-2), two receptors (the IGF-1 and IGF-2 receptors), and six IGF binding proteins (11,12). IGF-1 exerts its function after birth, whereas IGF-2 is prevalent in the fetal period and is a potent mitogen that regulates cell growth and differentiation (12). Notably, the IGF family has recently been linked to the pathogenesis of several cancer types (12). Results from early studies suggested higher plasma IGF-1 in patients with prostate cancer (13), breast cancer (14), colon cancers (15), and lung cancer (16). However, circulating levels of IGF-1 decrease sharply in patients with chronic liver diseases and HCC (17–22) because the liver synthesizes most of the circulating IGF-1. Subsequently, low IGF-1 levels lead to bone loss and other metabolic changes in patients with cirrhosis (23–25). Additionally, circulating IGF-1 has been found to correlate with the status of liver disease, histologic grade of fibrosis, and liver reserve scores such as CTP and MELD (26–28,29,30), which is used to predict 3-month mortality risk to determine liver organ allocation priorities. Our recent studies (31,32) showed that baseline plasma IGF-1 level was statistically significantly associated with patients' survival, synthetic function of the liver, and tumor parameters. However, no study to date has assessed the potential role of integrating plasma IGF-1 into CTP parameters on patients' prognostication. Because HCC tumors act as space-occupying lesions that decrease the synthetic function of the liver and decrease hepatic IGF-1 production, we hypothesized that plasma IGF-1 could be used as a surrogate marker for hepatic reserve that can be substituted for the subjective variables (ascites and hepatic encephalopathy) in the CTP score to create a novel liver score based exclusively on objective variables with increased prognostic accuracy. The study aims were 1) to retrospectively replace ascites and encephalopathy grading within CTP with plasma IGF-1 level in 310 patents (training cohort) and compare the performance of the IGF and the CTP scores, and 2) to prospectively validate the performance of the IGF score as compared with CTP score in an independent cohort of 155 HCC patients.

Methods

Patients with HCC enrolled at MD Anderson Cancer Center (MDACC) from January 2000 to May 2008 were used as a discovery or training cohort, and patients enrolled from June 2008 to January 2011 were used as an independent prospective validation cohort, as proposed in National Cancer Institute grant R21CA170035. Patients' blood samples and epidemiologic and clinical data were prospectively collected, and plasma samples were analyzed retrospectively for IGF-1 in the training cohort but analyzed prospectively in the validation cohort. This study was approved by the institutional review boards at MDACC, and patients signed written informed consent. For the two HCC cohorts, the study involved pathologically or radiologically confirmed HCC; the diagnosis was based on the criteria set forth by the 2005 guidelines of the American Association for the Study of Liver Diseases (33) for patients enrolled after 2005 who did not have biopsy samples available.

Baseline Plasma IGF-1 Level

For all patients, peripheral venous blood specimens (3–5 mL) were collected, anticoagulated by ethylenediaminetetraacetic acid, and subjected to centrifugation for 15 minutes at 3000 rpm. The plasma was then removed, aliquoted, and snap-frozen at –20°C until analyzed.

Table 1. The original Child-Turcotte-Pugh scoring system and a proposed new scoring system (Kaseb-Morris Score) that incorporates plasma level of insulin-like growth factor 1*

Parameter	Original CTP score† (points)			IGF-CTP score‡ (points)		
	1	2	3	1	2	3
Encephalopathy	None	Mild (1–2)	Severe (3–4)	—	—	—
Ascites	None	Mild/moderate	Severe/refractory	—	—	—
Albumin, g/dL	>3.5	2.8–3.5	<2.8		Same as for CTP score	
PT prolongation, sec	<4	4–6	>6		Same as for CTP score	
Bilirubin, mg/dL§	<2	2–3	>3		Same as for CTP score	
IGF-1, ng/mL	—	—	—	>50	26–50	<26

* CTP = Child-Turcotte-Pugh; IGF-1 = insulin-like growth factor 1; PT = prothrombin time.

† CTP score classes: A (5–6 points), B (7–9 points), C (>9 points).

‡ IGF-CTP (Kaseb-Morris) score classes: A (4–5 points), B (6–7 points), C (>7 points).

§ In primary biliary cirrhosis and primary sclerosing cholangitis, the upper limit of bilirubin for 1 point is 4 mg/dL and the upper limit for 2 points is 10 mg/dL.

MDACC Training Cohort. IGF-1 was tested by enzyme-linked immunosorbent assay performed according to the manufacturer's directions (Quantikine Human IGF-1 ELISA Kit; R&D Systems, Minneapolis, MN).

MDACC Validation Cohort Testing. Plasma IGF-1 levels were measured at a Clinical Laboratory Improvement Amendments–certified facility that uses Luminex microsphere technology by Myriad Laboratories (Austin, TX).

Statistical Analysis

The overall survival (OS) was computed as the time period from the date of the blood draw when IGF-1 was measured to the date of death or last follow-up visit when patients were alive. The patients

Table 2. Baseline characteristics of patients in the training and validation cohorts*

Patient characteristic	Parameter	Training cohort, No. (%) (n = 310)	Validation cohort, No. (%) (n = 155)	P†
Age, y	≤60	136 (43.9)	67 (43.2)	.92
	>60	174 (56.1)	88 (56.8)	
Sex	Male	218 (70.3)	113 (72.9)	.59
	Female	92 (29.7)	42 (27.1)	
Viral hepatitis infection	HCV, HBV, or both	139 (44.8)	78 (50.3)	.28
	None	171 (55.2)	77 (49.7)	
Serum α-FP, ng/mL	<400	207 (66.8)	99 (63.9)	.15
	≥400	97 (31.3)	56 (36.1)	
	Missing	6 (1.9)	None	
Tumor differentiation	Well	122 (39.4)	49 (31.6)	.05
	Moderate	100 (32.3)	42 (27.1)	
	Poor	52 (16.7)	38 (24.5)	
	Missing (no biopsy)	36 (11.6)	26 (16.8)	
Tumor nodularity	Uninodular	108 (36)	47 (30·3)	.20
	Multinodular	192 (64)	108 (69·7)	
	Missing	10	0	
Tumor size, proportion of liver	≤50%	203 (66.3)	113 (73.4)	.14
	>50%	103 (33.7)	41 (26.6)	
	Missing	4	1	
Vascular invasion	Yes	88 (28.5)	54 (34.8)	.30
	No	221 (71.5)	100 (64·5)	
	Missing	1	1	
Lymph node spread	Yes	126 (41)	90 (58.1)	<.001
	No	181 (59)	65 (41.9)	
	Missing	3	None	
Extrahepatic metastasis	No	242 (78.1)	70 (45.2)	<.001
	Yes	66 (21.3)	85 (54.8)	
	Missing	2 (0.6)	None	
ALT, U/L	≤40	137 (44.5)	67 (43.2)	.84
	>40	171 (55.5)	88 (56.8)	
	Missing	2	None	
AST, U/L	≤45	93 (32.6)	26 (16.8)	<.001
	>45	192 (67.4)	129 (83.2)	
	Missing	25	None	
Cirrhosis	No	116 (37.4)	52 (36.4)	.83
	Yes	194 (62.6)	93 (63.6)	
CTP score class	A	221 (71.8)	126 (81.3)	.06
	B	79 (25.6)	25 (16.1)	
	C	8 (2.6)	4 (2.6)	
	Missing	2	None	
Barcelona Clinic liver cancer stage	0	20 (6.8)	2 (1.3)	.009
	A	27 (9.1)	13 (8.4)	
	B	30 (10.1)	17 (11)	
	C	196 (66.2)	119 (76.8)	
	D	23 (7.6)	4 (2.5)	
	Missing	14	None	
Treatment history	Combined therapy, systemic/local	143 (46.1)	89 (57.4)	<.001
	Local therapy only	29 (9.4)	40 (25.8)	
	Combined therapy, local/surgery	94 (30.3)	15 (9.7)	
	No treatment	44 (14.2)	11 (7.1)	

* α-FP = α-fetoprotein; ALT = alanine transaminase; AST = aspartate transaminase; CTP = Child-Turcotte-Pugh; HBV = hepatitis B virus; HCV = hepatitis C virus.

† Fisher's exact test was used, and P values are two-sided.

alive were censored for analyses. Optimal IGF-1 cutoff points were identified by the recursive partitioning method, as previously described (31). Briefly, to identify an optimal IGF-1 cut point, we first split our training data (n = 310) randomly into training (two-thirds) and validation (test; one-third) sets. We applied recursive partitioning methodology for censored data (using R function rpart) in the training set to find the optimal cut point maximizing the survival difference between the low and high IGF-1 groups and then validated the cut point by using the log-rank test to the dichotomized IGF-1 value on the test data. We repeated this methodology using 20 different random splits. We applied this routine to identify the first cut point of 26. Then we used the data with IGF greater than 26 with the same analysis routine to identify the second cut point of 50. The training cohort alone was used for model building. The log-rank test was used to compare OS between plasma IGF-1 concentrations within each of the three traditional CTP classes. A univariable Cox model was used to compare OS among subgroups of patients. The method by Grambsch and Therneau (34) was used to test the proportional hazards assumption. Recursive partitioning identified optimal IGF-1 ranges within the training set and established three

Table 3. Log-rank and Cox model results for overall survival of the training and validation cohorts based on insulin-like growth factor 1 level*

Variable	Level	Training cohort (n = 310)				Validation cohort (n = 155)			
		No. (%)	Death event	Median OS, months (95% CI)	P†	No. (%)	Death event	Median OS, months (95% CI)	P†
All patients	—	310	238	13.22 (11.4 to 16.6)	—	155	71	15.7 (12.2 to 19.9)	—
IGF-1 level, ng/mL	1 (>50)	133 (42.9)	92	22.6 (15.1 to 28.8)	<.001	61 (39.4)	24	23.7 (18.4 to NA)	<.001
	2 (26–50)	109 (35.2)	85	13.6 (8.5 to 19.3)	—	17 (10.9)	5	9.5 (7.6 to NA)	—
	3 (<26)	68 (21.9)	61	5.0 (4.01 to 11.9)	—	77 (49.8)	42	9.4 (6.4 to 14.74)	—
				HR (95% CI)	P			HR (95% CI)	P
	1 (>50)	—	—	1.00 (referent)	—	—	—	1.00 (referent)	--
	2 (26–50)	—	—	1.45 (1.08 to 1.95)	.01	—	—	1.26 (0.48 to 3.33)	.64
	3 (<26)	—	—	2.50 (1.80 to 3.47)	<.001	—	—	2.81 (1.68 to 4.68)	<.001
		—	—			—	—		
	2 (26–50)	—	—	1.00 (referent)	—	—	—	1.00 (referent)	—
	3 (<26)	—	—	1.72 (1.24 to 2.40)	.001	—	—	2.28 (0.9 to 5.76)	.08

* CI = confidence interval; HR = hazard ratio; IGF-I, insulin-like growth factor 1; NA = not applicable "reached"; OS = overall survival.

† The log rank test and univariable Cox model were used to calculate the P values, and the P values were two-sided.

Table 4. Log-rank and Cox model results for overall survival of the training and validation cohorts by Child-Turcotte-Pugh and insulin-like growth factor classes*

Scoring system	Class	Training cohort (n = 310)				Validation cohort (n = 155)			
		No.†	Death event	Median OS, months (95% CI)	P‡	No.	Death event	Median OS, months (95%CI)	P‡
IGF score	A (4–5)	186	132	20.5 (15.0 to 26.7)	<.001	70	25	25.9 (18.4 to NA)	<.001
	B (6–7)	87	72	11.5 (8.3 to 16.3)		70	38	9.5 (7.7 to 16.2)	
	C (>7)	26	25	2.5 (2.2 to 5.7)		15	8	5.1 (2.1 to NA)	
CTP score	A (5–6)	221	163	17.1 (12.9 to 1.4)	<.001	126	56	16.9 (13.2 to 25.1)	<.001
	B (7–9)	79	67	6.5 (4.8 to 13.1)		25	12	8.8 (7.1 to NA)	
	C (>10)	8	7	2.6 (0.7 to NA)		4	3	2.1 (0.5 to NA)	
				HR (95% CI)	P			HR (95% CI)	P
IGF score	A	—	—	1.00 (referent)	—	—	—	1.00 (referent)	—
	B	—	—	1.61 (1.20 to 2.15)	.001	—	—	2.67 (1.61 to 4.44)	<.001
	C	—	—	6.34 (4.03 to 9.97)	<.001	—	—	7.70 (3.34 to 17.74)	<.001
		—	—			—	—		
	B	—	—	1.00 (referent)	—	—	—	1.00 (referent)	—
	C	—	—	3.94 (2.47 to 6.31)	<.001	—	—	2.88 (1.31 to 6.34)	.008
CTP score	A	—	—	1.00 (referent)	—	—	—	1.00 (referent)	—
	B	—	—	1.57 (1.18 to 2.09)	.002	—	—	1.94 (1.03 to 3.06)	.04
	C	—	—	4.63 (2.15 to 9.94)	<.001	—	—	17.12 (4.94 to 59.31)	<.001
		—	—			—	—		
	B	—	—	1.00 (referent)	—	—	—	1.00 (referent)	—
	C	—	—	2.94 (1.31 to 6.45)	.007	—	—	8.84 (2.35 to 33.24)	.001

* CI = confidence interval; CTP = Child-Turcotte-Pugh; IGF = insulin-like growth factor; NA = not applicable; OR = odds ratio; OS = overall survival.

† Parameters from 11 and two patients were missing to calculate their IGF-CTP score and CTP score, respectively.

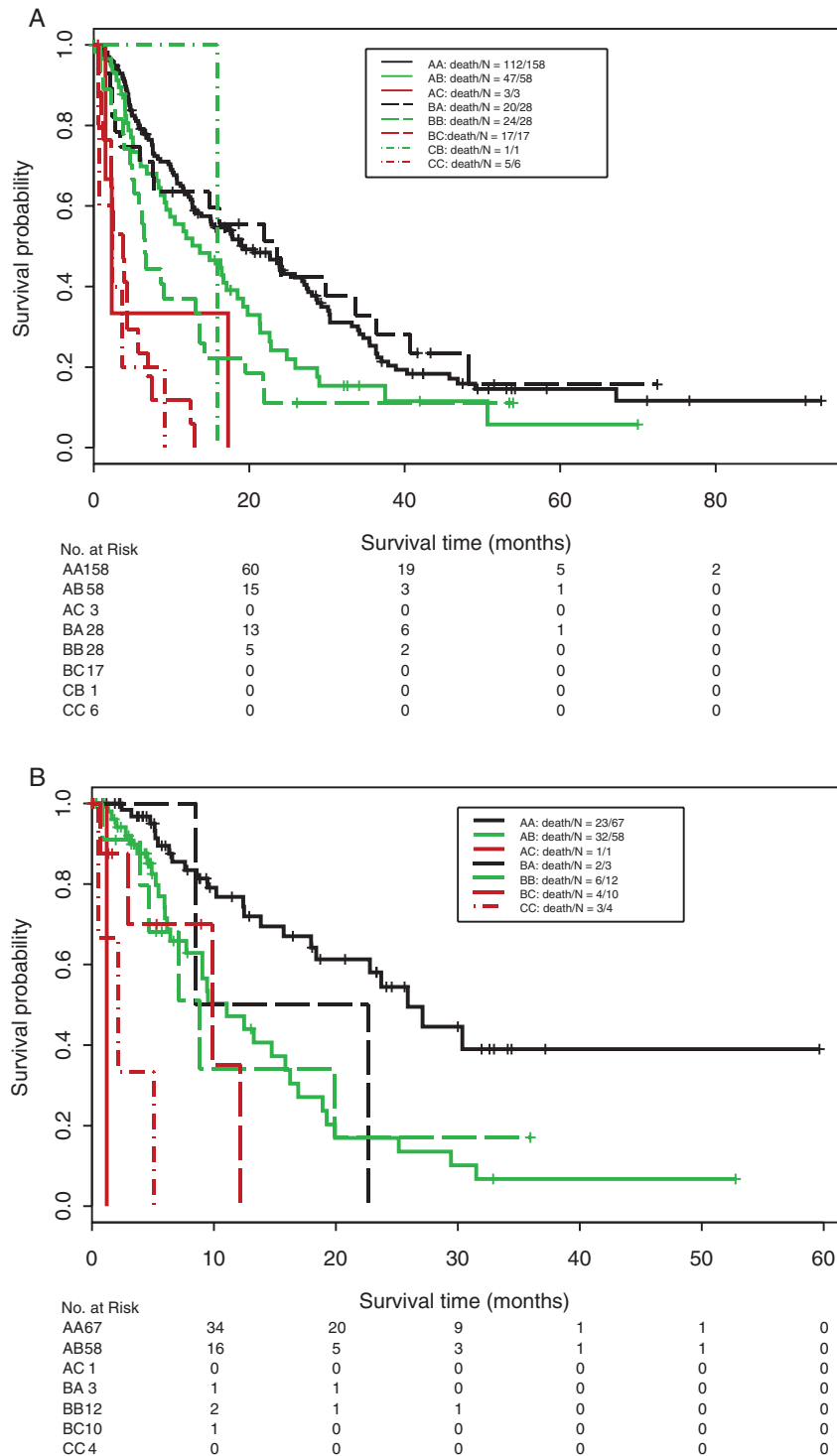‡ Log-rank test and univariable Cox model were used to calculate P values, and all P values were two-sided.

A

No. at Risk				
AA158	60	19	5	2
AB 58	15	3	1	0
AC 3	0	0	0	0
BA 28	13	6	1	0
BB 28	5	2	0	0
BC 17	0	0	0	0
CB 1	0	0	0	0
CC 6	0	0	0	0

B

No. at Risk						
AA67	34	20	9	1	1	0
AB58	16	5	3	1	1	0
AC 1	0	0	0	0	0	0
BA 3	1	1	0	0	0	0
BB12	2	1	1	0	0	0
BC10	1	0	0	0	0	0
CC 4	0	0	0	0	0	0

Figure 1. Kaplan–Meier estimates of overall survival of patients split by both scores (old/new) in the training cohort (**A**) and validation cohort (**B**). Tables of the numbers of patients at risk at various time points are below each graph.

distinct groups that associated with survival duration: greater than 50 ng/mL = 1 point; 26 to 50 ng/mL = 2 points; and less than 26 ng/mL = 3 points. We then constructed the new IGF score (IGF-CTP) by replacing ascites and encephalopathy grading with plasma IGF-1 value to construct an exclusively blood-based score (Table 1). Based on our training cohort OS and IGF-1 results, 22%, 35%, and 43% of patients had low, intermediate, and high IGF-1 levels, respectively.

Therefore, we determined that a prospective cohort of 155 patients (expected 22% low, 35% intermediate, 43% high) would have at least 80% power to detect a 7.2-month increase in median OS for intermediate or high IGF patients relative to low IGF patients, assuming a two-sided statistical significance level of .05. Harrell's C-index and U-statistics were used to compare the prognostic performance of both scores in both the training and validation cohorts.

Results

Patient Characteristics

Training Cohort. Between January 2000 and May 2008, 420 patients were enrolled, 310 (73.8%) of whom had available plasma samples for IGF-1 testing. This included 288 patients from a recently published study (31).

Prospective Validation Cohort. Between June 2008 and September 2011, 197 HCC patients were enrolled, 155 (78.6%) of whom had available plasma samples.

Comparison of Training and Validation Cohorts. We found no statistically significant differences in demographic, clinical, and epidemiological features between patients who had available blood samples and those who did not, including incidence of cirrhosis, CTP classification, HCC stage, age, sex, and tumor parameters (all *P* > .05). The reason for some missing samples was mainly related to insufficient time to obtain plasma samples during initial assessment in clinic. The median follow-up times were 43.3 months (95% confidence interval [CI] = 41 to 53.5 months) for the training cohort and 16.5 months (95% CI = 9.7 to 24.1 months) for the MDACC prospective validation cohort. Most patient characteristics were similar (Table 2), although the validation cohort had more patients with poor tumor differentiation, lymph node spread, metastases, advanced Barcelona Clinic Liver Cancer stage, higher aspartate transaminase, and more use of local therapy.

Comparison of OS Duration and Prognostic Accuracy by CTP Score vs IGF-CTP Score

The median OS was 13.2 months (95% CI = 11.4 to 16.6) in the training cohort and 15.7 months (95% CI = 12.2 to 19.9 months) in the validation cohort. Table 3 summarizes OS by plasma IGF-1 level. Patients with high IGF had statistically significantly better prognosis than those with low IGF (*P* < .001 for both cohorts). Patients with low IGF had worse prognosis than intermediate IGF in the training cohort (*P* = .001), but in the validation cohort, low IGF had a worse prognosis than intermediate IGF, but this difference was not statistically significant (hazard ratio [HR] =2.3; 95% CI = 0.9 to 5.8; *P* = .08). Note that only 17 of 155 (11.0%) of the patients in the validation cohort were IGF intermediate, far fewer than those who were classified as intermediate in the training cohort (n = 109 of 310; 35.2%), which limited our power to find statistical significance for even a large hazard ratio such as 2.3.

Table 4 and Figure 1 summarize OS by CTP and IGF-CTP scores. Both scores stratified patients into low- (A), intermediate- (B), and high-risk (C) groups that differed in OS (*P* < .001). A C-index analysis demonstrated that the prognostic stratification provided by IGF-CTP was statistically significantly better than CTP in both the training cohort (*P* = .003) and validation cohort (*P* = .005) (see Table 5). This improvement can be seen more clearly by focusing on the patients who are classified into different risk groups by the two scoring systems.

Reassignment of Patients Under CTP Classes to New IGF Score Classes

We found that 61.9% of the training cohort and 53.5% of the validation cohort were classified in the same risk groups by both scoring systems (Table 6). As seen in Kaplan–Meier plots in Figure 1 (and accompanying statistical comparisons in Table 7), patients reclassified by the IGF-CTP scoring system were better stratified by their new risk groups. For example, in the training cohort, 158 of 219 (72.1%) of CTP-A patients were classified as IGF-CTP-A and had median OS of 19.3 months (95% CI = 14.9 to 27.0 months), whereas 58 of 219 (26.5%) were reclassified as intermediate risk (IGF-CTP-B) and had shorter median OS of 13.6 months (95% CI = 9.1 to 19.7 months). This subset of CTP-A

Table 5. Ranking of scoring systems by C-index*

Patient cohort	Scoring system	C-index (95% CI)	*P*†
Training cohort (n = 310)	IGF-CTP score	0.608 (0.606 to 0.610)	.003
	CTP score	0.573 (0.71 to 0.575)	
Validation cohort (n = 155)	IGF-CTP score	0.672 (0.666 to 0.677)	.005
	CTP score	0.579 (0.576 to 0.583)	

* CI, confidence interval; CTP = Child-Turcotte-Pugh; IGF = insulin-like growth factor.

† U-statistics were used to calculate *P* values, and all *P* values were two-sided.

Table 6. Patient distribution in the training and validation cohorts for insulin-like growth factor 1 score class by Child-Turcotte-Pugh score class*

| Patient cohort | IGF-CTP score | CTP score, No. (%) | | |
		A (5–6)	B (7–9)	C (>10)
Training cohort (n = 310)	A (4–5)	158 (72.1)	28 (38.4)	0
	B (6–7)	58 (26.5)	28 (38.4)	1 (14.3)
	C (>7)	3 (1.4)	17 (23.2)	6 (85.7)
Validation cohort (n = 155)	A (4–5)	67 (53.2)	3 (12.0)	0
	B (6–7)	58 (46.0)	12 (48.0)	0
	C (>7)	1 (0.8)	10 (40.0)	4

* Notably, a substantial number of patients categorized under original Child-Turcotte-Pugh (CTP) class A moved to class B under the new insulin-like growth factor (IGF) scoring system: 58 of 219 patients (26.5%) in the training cohort, 58 of 126 patients (46%) in the validation cohort.

Table 7. Overall survival of the training and validation cohorts by Child-Turcotte-Pugh and Insulin-like Growth Factor–Child-Turcotte-Pugh score classes*

Variable	Training cohort (n = 310)				Validation cohort (n = 155)			
	No.	Death events	Median OS, months (95% CI)	P†	No.	Death events	Median OS, months (95% CI)	P†
Original A to new A (AA)	158	112	19.3 (14.9 to 27.0)	<.001†	67	23	25.9 (18.4 to NA)	<.001†
Original A to new B (AB)	58	47	13.6 (9.1 to 19.7)		58	32	11.0 (7.7 to 16.9)	
Original A to new C (AC)	3	3	2.3 (1.5 to NA)		1	1	1.2 (NA to NA)	
Original B to new A (BA)	28	20	23.5 (7.6 to 40.6)		3	2	15.6 (8.5 to NA)	
Original B to new B (BB)	28	24	6.5 (5.1 to 13.6)		12	6	8.8 (4.6 to NA)	
Original B to new C (BC)	17	17	3.7 (2.2 to 6.9)		10	4	9.8 (2.9 to NA)	
Original C to new A (CA)	0	—	—		0	—	—	
Original C to new B (CB)	1	1	15.9 (NA to NA)		0	—	—	
Original C to new C (CC)	6	5	2.3 (0.7 to NA)		4	3	2.1 (0.5 to NA)	
			HR (95% CI)	P			HR (95% CI)	P
AA			1.00 (referent)				1.00 (referent)	
AB			1.45 (1.03 to 2.04)	.03			2.83 (1.65 to 4.85)	<.001
AC			4.05 (1.28 to 12.83)	.02			NA	NA
BA			0.95 (0.59 to 1.53)	.84			2.63 (0.62 to 11.23)	.19
BB			2.0 (1.28 to 3.11)	.002			2.87 (1.17 to 7.08)	.02
BC			6.45 (3.78 to 11.01)	<.001			4.66 (1.57 to 13.86)	.006
CB			1.69 (0.24 to 112.16)	.60			NA	NA
CC			8.94 (3.59 to 22.23)	<.001			32.39 (8.86 to 118.49)	<.001
BB			1.00 (referent)				1.00 (referent)	
AB			0.73 (0.44 to 1.19)	.20			0.98 (0.41 to 2.36)	.97
AC			2.03 (0.61 to 6.77)	.25			NA	NA
BA			0.48 (0.26 to 0.87)	.03			0.92 (0.18 to 4.57)	.91
BC			3.23 (1.72 to 6.09)	<.001			1.62 (0.45 to 5.86)	.46
CB			0.85 (0.11 to 6.28)	.87			NA	NA
CC			4.48 (1.69 to 11.8)	.003			11.27 (2.64 to 48.14)	.001
CC			1.00 (referent)				1.00 (referent)	
AB			0.11 (0.05 to 0.28)	<.001			0.09 (0.02 to 0.31)	<.001
AC			0.45 (0.11 to 1.91)	.28			NA	NA
BA			0.11 (0.04 to 0.29)	<.001			0.08 (0.01 to 0.52)	.008
BC			0.72 (0.27 to 1.96)	.52			0.14 (0.03 to 0.68)	.01
CB			0.19 (0.02 to 1.63)	.13			NA	NA

* CTP = Child-Turcotte-Pugh; HR = hazards ratio; IGF = insulin-like growth factor; NA = not applicable; OS = overall survival.

† P value compares across all groups.

‡ The log-rank test and univariable Cox models were used to calculate the two-sided P values.

patients who were reclassified as IGF-CTP-B had statistically significantly worse prognosis than other CTP-A patients classified as IGF-CTP-A (HR = 1.45; 95% CI = 1.03 to 2.04; P = .03) (Table 7). Of the 73 patients classified as CTP-B in the training cohort, 28 (38.4%) were also classified as IGF-CTP-B and had median OS of 6.5 months (95% CI = 5.1 to 13.6 months), whereas 28 (38.4%) were reclassified as lower risk (IGF-CTP-A) and had longer median OS of 23.5 months (95% CI = 7.6 to 40.6 months) and 17 (23.2%) were reclassified as higher risk (IGF-CTP-C) and had shorter median OS of 3.7 months (95% CI = 2.2 to 6.9 months). This subset of CTP-B patients reclassified as IGF-CTP-A had statistically significantly reduced hazard ratios than the CTP-B patients classified as IGF-CTP-B (HR = 0.48; 95% CI = 0.26 to 0.87; P = .02), whereas the CTP-B patients reclassified as higher risk IGF-CTP-C had statistically significantly increased hazard ratios than CTP-B patients classified as IGF-CTP-B (HR = 3.23; 95% CI = 1.76 to 6.09; P < .001) (Table 7). Thus, IGF-CTP found a subset of CTP-B patients in the training data with better prognosis than other CTP-B patients and a subset with worse prognosis. Very

few patients in the training cohort were reclassified from high risk (CTP-C) to low (IGF-CTP-A) or intermediate (IGF-CTP-B) risk or from low risk (CTP-A) to high (IGF-CTP-C) risk.

In the validation cohort, 67 of 126 (53.2%) of CTP-A patients were classified as IGF-CTP-A and had median OS of 25.9 months (95% CI = 18.4 months to not applicable), whereas 58 of 126 (46.0%) were reclassified as intermediate (IGF-CTP-B) risk and had lower median OS of 11.0 months (95% CI = 7.7 to 16.9 months) but had similar prognosis as other intermediate-risk (IGF-CTP-B) patients who were CTP-B (median OS = 8.8 months; 95% CI = 4.6 months to not applicable). This subset of CPT-A patients reclassified as IGF-CTP-B had statistically significantly worse prognosis than the CPT-A patients classified as IGF-CTP-A (HR = 2.83; 95% CI = 1.65 to 4.85; P < .001). This represents a subset of the CTP-A patients with worse prognosis.

Discussion

Despite the mounting evidence for plasma IGF-1 as a liver function assessment test, a strategy to incorporate it into clinical practice is

lacking. The results of this study confirm our biologically driven hypothesis that plasma IGF-1 level is a surrogate marker for functional liver reserve in HCC, which we propose to effectively replace the subjective assessment of encephalopathy and ascites severity in the CTP score. IGF-CTP offered statistically significantly more accurate survival prediction and prognostic stratification than the CTP score in two independent HCC cohorts. The new score, IGF-CTP, uses exclusively objective laboratory variables that are routinely assessed in clinical practice, including plasma IGF-1, which is a routine, reproducible, cost-effective, and standardized Clinical Laboratory Improvement Amendments–certified laboratory test and is therefore ready for implementation in routine practice.

Notably, because decreased hepatic reserve, caused by the underlying liver disease in addition to the HCC space-occupying tumors, has the ability to cause physiologic derangements in the whole body, including those related to decreased plasma IGF-1 (23–25), assessment of the liver reserve using the IGF score can guide initial and subsequent therapy decisions and hence decrease morbidity and mortality of HCC patients. Furthermore, CTP score cannot discriminate among the majority of HCC patients undergoing therapy because most treatment candidates are of CTP class A. Our study indicated that the IGF-CTP score resulted in more accurate prediction of OS in both cohorts, in addition to reclassification of a substantial proportion of HCC patients from class A using the CTP score (presumably indicating good hepatic reserve) to class B using the new score (indicating poorer hepatic reserve) (Table 7; Figure 1, A and B). This refinement of survival prediction and risk assessment of CTP class A patients is extremely important clinically because these patients are generally expected to do well because of their presumably good hepatic reserve and thus are considered the standard patient population for active therapy and clinical trial entry. Therefore, the new IGF score may help clinicians in designing clinical trials with comparable stratification criteria, which is critical to meaningful interpretation of trial results and comparison of results from different trials.

Conversely, some patients among CTP class B patients moved to IGF-CTP-A and were found to have longer OS. These patients did considerably better than generally expected, a reclassification that would afford these patients access to a broader array of therapeutic options. Unfortunately, the very small number of patients reclassified from CTP-B to IGF-CTP-A in the validation cohort resulted in low power to confirm this result from the training data, so future validation studies are warranted to confirm these findings.

We compared the prognostic scales using C-index, which yields values ranging from 0 (no discrimination) to 1 (perfect separation) to compare both scores. Notably, C-index has been used to compare the prognostic accuracy of liver scores in several other studies, including the landmark prospective study of 3437 adult liver transplant candidates with chronic liver disease to estimate 3-month mortality, which established the universal use of MELD score to determine organ allocation priorities (35). The study reported a C-index of 0.83 for MELD score as compared with 0.76 for the CTP score. Although the difference between C-indices was not large (0.07), it was statistically significant ($P < .001$) (35). Similarly, the differences between C-indices in our study, although not large, were statistically significant because the C-index computes the ability to predict survival for all patients

in the cohort, including those whose CTP and IGF-CTP scores were the same and those whose scores were different.

Furthermore, HCC staging systems depend on CTP score to assess hepatic reserve (5). Thus, future studies to evaluate replacing CTP score with the IGF-CTP score in HCC staging systems may lead to a more accurate approach to patients' stratification and treatment assignment and may decrease futility rates in trials of investigational drugs.

One of the major strengths of our study is that our hypothesis was tested in two independent cohorts, including a retrospective training cohort and a prospective validation cohort. Additionally, there is a remarkable consistency of the independently validated IGF-1 ranges in our study with prior reports that correlated IGF-1 level with degree of hepatic dysfunction. The highest category for IGF-1 (<26 ng/mL = score 3) in our study is consistent with recent reports of the association between advanced cirrhosis (CTP class C) and IGF-1 levels of less than 25 ng/mL (26) and less than 27 ng/mL (27). Our middle IGF-1 score range (26–50 ng/mL = score 2) is also supported by studies that found that mean circulating IGF-1 levels of 45 ng/mL (27) and 57 ng/mL (28) correlated with CTP class B and moderate histologic liver fibrosis in patients with cirrhosis, respectively. Furthermore, we did not observe statistically significant demographic or clinicopathologic differences between enrolled patients with or without available samples; therefore, there is no evidence of selection bias. Finally, prognostic scores provide information about patients regardless of type of therapy received, if any. Therefore, because the IGF-CTP score testing and validation studies included patients with different stages and treatment and at different time points of their disease course, the superior prognostic value of the IGF-CTP score as compared with CTP score is clinically significant.

Our study has some limitations. Because of the retrospective nature of the IGF-1 measurement, the study could not assess the new score's ability to predict specific treatment outcome and rate of adverse events in relation to pretreatment score. Future trials to assess this correlation are therefore warranted. Furthermore, selecting optimal cutoff points for circulating biomarkers remains challenging because of the potential for daily variations, in addition to variations based on patient genetics, sex, age, and other demographic characteristics. That said, our studies have clearly shown no statistically significant differences in mean IGF-1 levels among patients of different ages, sexes, or ethnicities and also confirmed the independent prognostic information obtained from plasma IGF-1 level.

In summary, the CTP score, despite its limitations, has remained the standard tool for assessing hepatic reserve in patients with HCC to guide therapy decisions, predict treatment outcome, and stratify patients in clinical trials. In this study, we incorporated plasma IGF-1 level into a new score by eliminating the CTP subjective clinical variables of ascites and encephalopathy. The new IGF score considerably improved the accuracy of survival prediction and resulted in reclassification of a large number of patients. However, international validation is required given the noted heterogeneity in HCC.

References

1. Child CG. Remote results of portal surgery in liver cirrhosis. *Rev Int Hepatol*. 1964;14:287–288.

2. Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973;60(8):646–649.

3. Llovet JM, Di Bisceglie AM, Bruix J, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *J Natl Cancer Inst*. 2008;100(10):698–711.

4. Wilson SR, Greig P, Kaseb AO. Pretreatment assessment of hepatocellular cancer: expert consensus conference. *HPB (Oxford)*. 2010;12(5):300–301.

5. Vauthey JN, Dixon E, Abdalla EK, et al. Pretreatment assessment of hepatocellular carcinoma: expert consensus statement. *HPB (Oxford)*. 2010;12(5):289–299.

6. Durand F, Valla D. Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. *J Hepatol*. 2005;42(Suppl1):S100–S107.

7. Oellerich M, Burdelski M, Lautz HU, et al. Assessment of pre-transplant prognosis in patients with cirrhosis. *Transplantation*. 1991;51(4):801–806.

8. Testa R, Valente U, Risso D, et al. Can the MEGX test and serum bile acids improve the prognostic ability of Child-Pugh's score in liver cirrhosis? *Eur J Gastroenterol Hepatol*. 1999;11(5):559–563.

9. Malinchoc M, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology (Baltimore)*. 2000;31(4):864–871.

10. Botta F, Giannini E, Romagnoli P, et al. MELD scoring system is useful for predicting prognosis in patients with liver cirrhosis and is correlated with residual liver function: a European study. *Gut*. 2003;52(1):134–139.

11. LeRoith D, Roberts CT Jr. The insulin-like growth factor system and cancer. *Cancer Lett*. 2003;195(2):127–137.

12. Jones JI, Clemmons DR. Insulin-like growth factors and their binding proteins: biological actions. *Endocrine Rev*. 1995;16(1):3–34.

13. Chan JM, Stampfer MJ, Giovannucci E, et al. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. *Science (New York)*. 1998;279(5350):563–566.

14. Hankinson SE, Willett WC, Colditz GA, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet*. 1998;351(9113):1393–1396.

15. Ma J, Pollak MN, Giovannucci E, et al. Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J Natl Cancer Inst*. 1999;91(7):620–625.

16. Yu H, Spitz MR, Mistry J, et al. Plasma levels of insulin-like growth factor-I and lung cancer risk: a case-control analysis. *J Natl Cancer Inst*. 1999;91(2):151–156.

17. Buzzelli G, Dattolo P, Pinzani M, et al. Circulating growth hormone and insulin-like growth factor-I in nonalcoholic liver cirrhosis with or without superimposed hepatocarcinoma: evidence of an altered circadian rhythm. *Am J Gastroenterol*. 1993;88(10):1744–1748.

18. Luo SM, Tan WM, Deng WX, et al. Expression of albumin, IGF-1, IGFBP-3 in tumor tissues and adjacent non-tumor tissues of hepatocellular carcinoma patients with cirrhosis. *World J Gastroenterol*. 2005;11(27):4272–4276.

19. Stuver SO, Kuper H, Tzonou A, et al. Insulin-like growth factor 1 in hepatocellular carcinoma and metastatic liver cancer in men. *Int J Cancer*. 2000;87(1):118–121.

20. Su WW, Lee KT, Yeh YT, et al. Association of circulating insulin-like growth factor 1 with hepatocellular carcinoma: one cross-sectional correlation study. *J Clin Lab Anal*. 2010;24(3):195–200.

21. Garcia-Galiano D, Sanchez-Garrido MA, Espejo I, et al. IL-6 and IGF-1 are independent prognostic factors of liver steatosis and non-alcoholic steatohepatitis in morbidly obese patients. *Obes Surg*. 2007;17(4):493–503.

22. Volzke H, Nauck M, Rettig R, et al. Association between hepatic steatosis and serum IGF1 and IGFBP-3 levels in a population-based sample. *Eur J Endocrinol*. 2009;161(5):705–713.

23. Goral V, Simsek M, Mete N. Hepatic osteodystrophy and liver cirrhosis. *World J Gastroenterol*. 2010;16(13):1639–1643.

24. Bonefeld K, Moller S. Insulin-like growth factor-I and the liver. *Liver Int*. 2011;31(7):911–919.

25. Mitchell R, McDermid J, Ma MM, et al. MELD score, insulin-like growth factor 1 and cytokines on bone density in end-stage liver disease. *World J Hepatol*. 2011;3(6):157–163.

26. Rehem RN, El-Shikh WM. Serum IGF-1, IGF-2 and IGFBP-3 as parameters in the assessment of liver dysfunction in patients with hepatic cirrhosis and in the diagnosis of hepatocellular carcinoma. *Hepatogastroenterology*. 2011;58(107–108):949–954.

27. Wu YL, Ye J, Zhang S, Zhong J, et al. Clinical significance of serum IGF-I, IGF-II and IGFBP-3 in liver cirrhosis. *World J Gastroenterol*. 2004;10(18):2740–2743.

28. Colak Y, Senates E, Ozturk O, et al. Serum concentrations of human insulin-like growth factor-1 and levels of insulin-like growth factor-binding protein-5 in patients with nonalcoholic fatty liver disease: association with liver histology. *Eur J Gastroenterol Hepatol*. 2012;24(3):255–261.

29. Castro GR, Coelho JC, Parolin MB, et al. Insulin-like growth factor I correlates with MELD and returns to normal level after liver transplantation. *Ann Transpl*. 2013;18:57–62.

30. Assy N, Pruzansky Y, Gaitini D, et al. Growth hormone-stimulated IGF-1 generation in cirrhosis reflects hepatocellular dysfunction. *J Hepatol*. 2008;49(1):34–42.

31. Kaseb AO, Morris JS, Hassan MM, et al. Clinical and prognostic implications of plasma insulin-like growth factor-1 and vascular endothelial growth factor in patients with hepatocellular carcinoma. *J Clin Oncol*. 2011;29(29):3892–3899.

32. Kaseb AO, Abbruzzese JL, Vauthey JN, et al. I-CLIP: improved stratification of advanced hepatocellular carcinoma patients by integrating plasma IGF-1 into CLIP score. *Oncology*. 2011;80(5–6):373–381.

33. Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology (Baltimore)*. 2005;42(5):1208–1236.

34. Grambsch PMaT, TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515–526

35. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91–96.

Affiliations of authors: Department of Gastrointestinal Medical Oncology (AOK, MMH, YKC, SC, HM H, KR, JLA, RW), Department of Biostatistics (LX, JSM), Department of Systems Biology (JSL), Department of Surgical Oncology (JNV, TA, CC, SAC), Department of Radiation Oncology (SK), Department of Interventional Radiology (MW), and Department of Pathology (AR), University of Texas MD Anderson Cancer Center, Houston, TX; Department of Hepatology, Baylor College of Medicine, Houston, TX (JMV, PJ).