original article

Annals of Oncology 19: 1117–1126, 2008
doi:10.1093/annonc/mdn030
Published online 25 February 2008

Prognosis of advanced hepatocellular carcinoma: comparison of three staging systems in two French clinical trials

S. Collette[1,2], F. Bonnetain[2]*, X. Paoletti[1], M. Doffoel[3], O. Bouché[4], J. L. Raoul[5], P. Rougier[2], F. Masskouri[2], L. Bedenne[2] & J. C. Barbare[1,6]

[1]*Institut National du Cancer, Paris;* [2]*Methodological and Biostatistical Unit, Fédération Francophone de Cancérologie Digestive INSERM U866, Dijon;* [3]*CHU Strasbourg, Strasbourg;* [4]*Service d'Hépatogastroentérologie, Centre Hospitalo-Universitaire R. Debré, Reims;* [5]*Centre Eugène Marquis, Rennes;* [6]*CHRU, Amiens, France*

Received 16 November 2007; revised 14 January 2008; accepted 15 January 2008

Objective: The objective of this study was to assess the performance of three staging systems [Okuda, Cancer of the Liver Italian Program (CLIP) and Barcelona Clinic Liver Cancer group (BCLC)], for predicting survival in patients with hepatocellular carcinoma (HCC) and to explore how to improve prognostic classification among French patients with HCC whose main etiology is alcoholic cirrhosis.

Methods: We have pooled two randomized clinical trials in palliative condition from the Fédération Francophone de Cancerologie Digestive. They had included 416 and 122 patients. Performances of Okuda, CLIP and BCLC scores have been compared using Akaike information criterion, discriminatory ability (Harrell's *C* and the Royston's *D* statistics), monotonicity of gradients and predictive accuracy (Schemper statistics Vs). To explore how to improve classifications, univariate and multivariate Cox model analyses were carried out.

Results: The pooled database included 538 patients. The median survival was 5.3 months (95% confidence interval 4.6–6.2). For all statistics CLIP staging system had a better prognostic ability. Performances of all staging systems were rather disappointing. World Health Organization performance status (WHO PS) for CLIP or α-fetoprotein for BCLC allowed a significant improvement of prognostic information.

Conclusion: Our results indicate that CLIP staging seems to be most adapted to palliative setting and that it could be better by associating WHO PS.

Key words: hepatocellular carcinoma, overall survival, prognostic factor, validation

introduction

Primary liver cancer is the fifth most frequent cancer and the third most common cause of cancer-related death in the world [1]. Hepatocellular carcinoma (HCC) is a main form of liver cancer; this cancer generally develops on cirrhosis or hepatitis B or C infections. Incidence of HCC has substantially increased in developed countries during the last three decades [2, 3]. In France 6000 deaths per year are due to this cancer, whose main etiology is related to alcohol.

Classification of patients according to their prognosis is a central issue since inclusion criteria in clinical trials suppose that homogenous groups of patients can be identified. Various prognostic factors of overall survival (OS) have then been explored and several classifications have been proposed [4–7]. Among the most commonly used scores, one can quote the Okuda stage, Cancer of the Liver Italian Program (CLIP), Barcelona Clinic Liver Cancer Group (BCLC) and Groupe d'Etude et de Traitement du Carcinome Hépatocellulaire. Different studies have compared and ranked these classifications [8–16] according to their prognostic value. Results were discordant between studies and remained controversial. This can probably be partly explained by the difference in the investigated populations and by the different statistical methodologies applied. Furthermore, most of the studies focused on patients with mainly hepatitis B virus/hepatitis C virus etiology. Their conclusions may then not be consistent with studies based on alcoholic HCC.

This study focuses on patients with HCC in an advanced setting mainly associated with alcoholic cirrhosis etiology. On the basis of a pooled analysis of two randomized clinical trials (RCTs) carried out by the Fédération Francophone de Cancerologie Digestive (FFCD), we have assessed and compared the performance of three prognostic classifications (Okuda, CLIP and BCLC) for predicting OS. We also explore whether the staging systems could be improved by adding other clinical or biological variables.

Correspondence to: Dr F. Bonnetain, Methodological and Biostatistical Unit, Fédération Francophone de Cancérologie Digestive, INSERM U866, Dijon, France.
Tel: +33-3-80-73-77-84; Fax: +33-3-80-73-77-34; E-mail: fbonnetain@dijon.fnclcc.fr

patients and methods

patients

We carried out a pooled analysis of two RCTs of patients with HCC in a palliative setting.

The FFCD 9403 trial evaluated survival benefit of adding tamoxifen over best supportive care. In this trial, 420 eligible patients were entered in a randomized study from 78 French institutions [17]. Eligibility criteria were HCC not suitable for surgical resection, liver transplantation, percutaneous ablation or transarterial chemoembolization. Diagnosis of HCC was either cytologically or histologically confirmed or made by the association of an established diagnosis of cirrhosis: demonstration in ultrasonography, and/or computed tomography scan (CT scan), and/or magnetic resonance imaging (MRI) of a space-occupying lesion having an image consistent with the diagnosis of HCC and persistently elevated α-fetoprotein (AFP) values >500 μg/l. Exclusion criteria were renal failure (serum creatinine level >130 μmol/l), advanced liver disease (Child–Pugh class C), World Health Organization performance status (WHO PS) two or more and prior treatment with tamoxifen.

The FFCD 9402 trial evaluated survival benefit of adding transarterial lipiodol chemoembolization over tamoxifen alone. In this trial, 122 eligible patients from 15 French institutions were randomly assigned. [18]. Eligibility criteria were HCC not suitable for surgical resection, liver transplantation or percutaneous ablation; all patients were cirrhotic (cirrhosis diagnosis was histologically proven or based on clinical and biological parameter). Diagnosis of HCC was based on biopsy or persistently elevated AFP levels (>400 μg/l) with one typical imaging finding (ultrasonography, CT scan or MRI) or normal AFP levels with two concordant imaging findings. Exclusion criteria were advanced liver disease (Child–Pugh class C), advanced HCC (Okuda stage III), portal vein thrombosis (trunk and primary branches) or arteriovenous shunting, extrahepatic metastases, renal failure (serum creatinine level >120 μmol/l or creatinine clearance <80 ml/min), platelet count $<50 \times 10^9$/l, prothrombin time <50% and cardiac ejection fraction <35%.

We further selected patients with <60% of missing data studied.

prognostic scores. Table 1 presents definitions of Okuda, CLIP and BCLC prognostic scores.

Furthermore, Child–Pugh score, that is necessary to calculate the CLIP, was generated on the basis of ascite, encephalopathy, total bilirubin, prothrombin time and albumin.

collected variables and reconciliation. The following baseline variables were retained to calculate the prognostic classification and to explore whether the staging systems can be improved: age, sex, date and modality of HCC diagnosis, date of death or of last information on vital status, presence of cirrhosis and its etiology, clinical parameters (weight, edemas of the lower limbs, jaundice, hepatomegaly, hepatalgy, ascite and encephalopathy), serological parameters (total bilirubin; prothrombin time; and creatinine, albumin and AFP serum levels), tumoral characteristics (site of the principal tumor, maximum tumor diameter, number of localization, tumoral extension, portal vein thrombosis and extrahepatic metastases) and the WHO PS.

Biological parameters have been dichotomized according to the literature and age according to the median.

Portal vein thrombosis had been reported with different modalities in the two trials. Reconciliation has been carried out by the physician in charge of the study.

Table 1. Definitions of Okuda, CLIP and BCLC classifications

	Scores						
Okuda	0	1					
Ascites	Absent	Present					
Tumor size	≤50%	>50%					
Bilirubin (μmol/l)	≤50	>50					
Albumin (g/l)	≤30	>30					
CLIP	0	1	2				
Child–Pugh	A	B	C				
Tumor morphology	Uninodular and extension ≤50%	Multinodular and extension ≤50%	Massive or extension >50%				
AFP (ng/dl)	≤400	>400					
Portal vein thrombosis	No	Yes					
BCLC	A1	A2	A3	A4	B	C	D
PS	0	0	0	0	0	1–2	3–4
Tumor stage	Single	Single	Single	3 tumors <3 cm	Multinodular	Vascular invasion or extrahepatic spread	Any
Okuda	I	I	I	I–II	I–II	I–II	II
Liver functional status	No portal hypertension and normal bilirubin	Portal hypertension and normal bilirubin	Portal hypertension and abnormal bilirubin	Child–Pugh class A–B	Child–Pugh class A–B	Child–Pugh class A–B	Child–Pugh class C

Okuda stages: I = 0 points; II = 1–2 points; III = 3–4 points. BCLC (Barcelona Clinic Liver Cancer group) staging classification: stages A and B all criteria should be fulfilled; stage C at least one criteria performance status (PS) 1–2 or vascular invasion or extrahepatic spread; stage D at least one criteria PS 3–4 or Okuda stage III/Child–Pugh class C.

AFP, α-fetoprotein; BCLC, Barcelona Clinic Liver Cancer group; CLIP, Cancer of the Liver Italian Program.

'Small HCC' has been defined according to the Milan criteria [19] that is one nodule <50 mm or two to three nodules <30 mm.

statistical analysis

All statistical analyses were carried out on the pooled database stratified on trial to take into account trial heterogeneity. Per-trial analyses were then carried out and enabled to check robustness of our results.

Baseline variables were described as mean [standard deviation (SD)] or frequencies and percents. OS was defined as the time interval between date of inclusion and the date of death or the date of the last follow-up. Survival was estimated using the Kaplan–Meier approach and was compared using stratified log-rank test. Median of survival was calculated with its 95% confidence interval (CI). Univariate and multivariate Cox analyses stratified on trial were carried out to estimate hazard ratio (95% CI). Following the guidelines of Altman and Royston to validate prognostic model [20], we have investigated the following.

Two information criteria: the likelihood ratio (LR, χ^2) and the Akaike information criterion (AIC). LR χ^2 estimates loss of adjustment by calculating the difference of the deviance between models with and without the score. A smaller AIC value or a higher LR χ^2 indicates a better model.

Monocity of gradients has been checked by comparing the median of survival. A group of patients with better prognostic stage should have a higher median as compared with patients with poorest prognostic stage. Significant log-rank was considered as reflecting this monotonicity.

The discriminatory capacity was tested using two statistical methods: the Harrell's *C* statistics [21] and the Royston's *D* statistics [22]. Harrell's *C* statistics estimates the proportion of correct predictions, i.e. the proportion of patients with a better prognostic stage who have a better survival. Result of the Harrell's *C* index varied from 0.5 (no discrimination) to 1 (perfect discrimination).

Royston's *D* statistics [22] estimates separation between independent survival distributions under the proportional hazards assumption. Higher is the *D* statistic, better is the discriminatory capacity.

The added precision of the prediction and the explained variation were measured by the Schemper statistics Vs [23]. This statistic represents the part of the survival variability explained by the score. The higher the explained variability, the better the prognostic score is.

In order to explore staging systems improvement by other clinical or biological variables, we carried out univariate Cox analyses of all potential baseline predictors including the variables constituting each score. We have tested a multivariate model including all variables with univariate $P < 0.10$. The final model only included multivariate significant predictors. These variables were eligible to test an improvement of each scoring system.

Finally, multivariate Cox model analyses were carried out for each score. The best models were built with forward and backward procedures among baseline variables not redundant with the score. In the aim to retain the best prognostic variable to add from the final model, we have compared AIC, LR χ^2 and the log likelihood.

All data analyses were carried out using SAS 9.1.3 and R 2.3.0. A *P* value <0.05 was considered significant.

results

patient characteristics

In the 9403 trial, four patients, who had >60% of missing data, have been excluded and in the 9402 trial, one patient, who had a WHO PS of four, has been excluded. Finally, 122 patients in the 9402 trial and 416 patients in the 9403 trial have been retained and pooled (N = 538).

Baseline patient characteristics are described in Table 2. Mean age was 66 years (SD = 8.3 years, minimum = 35 years and maximum = 87 years) and males were in majority (four men for one woman). All patients of the 9402 trial were cirrhotic (inclusion criteria) and 90.4% of patients of the 9403 trial were cirrhotic. Among them ($n = 498$), 83.1% of patients had alcoholic cirrhosis and 19.9% of patients had hepatitis B or C etiology. In the 9402 trial, WHO PS of zero was more frequent than that in the 9403 trial (39% versus 18%, $P < 0.0001$). Finally, patients of the 9402 trial had a better clinical, biological and tumoral status (Table 2). Due to inclusion criteria, majority of patients were Child–Pugh class A or B, Okuda I and II, CLIP 1–3 and BCLC B or C.

overall survival

At the time the databases were frozen, 502 (93%) patients had died and only 36 patients (7%) were alive. The median survival was 5.3 months (95% CI 4.6–6.2) and 1-, 2- and 3-year OS rates were, respectively, 27.8%, 11.8% and 5.0%.

OS differed significantly according to trials (log-rank test: $P < 0.0001$) requiring to stratify analyses on the trial. Median survival was longer in the 9402 trial: 11.6 months (8.1–15.8) versus 4.4 months (3.8–5.0).

performance of prognostic scores

monotonicity of gradients. Whatever scoring system, monotonicity was respected: the higher the score, the longer the OS (Table 3, Figure 1).

According to Okuda stages I, II and III, medians were, respectively, 11.1 months (8.4–12.4), 4.1 months (3.5–5.0) and 1.5 months (0.9–1.8).

According to CLIP, median CIs overlapped between CLIP 0 and 1/CLIP 2 and 3/CLIP 4 and 5–6 (Table 3, Figure 1). Due to overlap, CLIP scores were regrouped into three classes. Medians were then, respectively, 14.7 months (11.8–17.7) for CLIP 0–1, 4.6 months (4.1–5.4) for CLIP 1–2 and 1.9 months (1.5–2.4) for CLIP 4 to 5–6.

Regarding BCLC scores, median CIs of BCLC A and BCLC B overlapped slightly. Median OS was 20.0 months (13.7–40.4), 12.4 months (9.3–17.5), 5.0 months (4.3–5.7) and 1.9 months (1.5–2.4) for BCLC A, B, C and D, respectively.

added information. According to information criteria, CLIP had the lowest AIC and the highest LR χ^2 (Table 3). CLIP in three or six classes seemed to be more informative to explain survival than Okuda and BCLC.

discriminatory capacity. Harrell's *C* statistics varied from 0.66 for CLIP score to 0.61 for BCLC score (Table 3). So the proportion of correct predictions according to prognostic stage was better for CLIP score. However, whatever score, these *C* statistics were close to 0.5, highlighting limited discriminatory properties.

Royston's *D* statistics varied from 1.01 for CLIP (three classes) and Okuda to 0.79 for BCLC score (Table 3).

Table 2. Baseline characteristics of patients with HCC according to the clinical trial, N = 538

	Trial				Total		*P* value
	9402		9403				
	n	%	*n*	%	*n*	%	
Gender							
Male	106	86.89	372	89.42	478	88.85	0.434
Female	16	13.11	44	10.58	60	11.15	
Age (years)							
<65	57	46.72	144	34.62	201	37.36	0.015
≥65	65	53.28	272	65.38	337	62.64	
Cirrhosis							
Absent	0	0.00	40	9.62	40	7.43	0.000
Present	122	100.00	376	90.38	498	92.57	
Alcoholic cirrhosis	29	23.77	95	22.84	124	23.05	
	93	76.23	321	77.16	414	76.95	
Jaundice							
No	112	91.80	333	80.05	445	82.71	0.003
Yes	10	8.20	83	19.95	93	17.29	
Hepatomegalia							
No	43	35.25	94	22.60	137	25.46	0.005
Yes	79	64.75	322	77.40	401	74.54	
Hepatalgia							
No	100	81.97	307	73.80	407	75.65	0.065
Yes	22	18.03	109	26.20	131	24.35	
Involved liver volume							
≤50%	107	87.70	287	68.99	394	73.23	<0.0001
>50%	15	12.30	129	31.01	144	26.77	
Extrahepatic metastases							
No	122	100.00	346	83.17	468	86.99	<0.0001
Yes	0	0.00	70	16.83	70	13.01	
Portal vein thrombosis							
No	90	73.77	251	60.34	341	63.38	0.007
Yes	32	26.23	165	39.66	197	36.62	
α-Fetoprotein serum level (μg/l)							
<200	76	62.30	197	47.36	273	50.74	0.004
≥200	46	37.70	219	52.64	265	49.26	
Total bilirubin (μmol/l)							
<20	61	50.00	181	43.51	242	44.98	0.205
≥20	61	50.00	235	56.49	296	55.02	
Prothrombin time (%)							
<80	61	50.00	236	56.73	297	55.20	0.189
≥80	61	50.00	180	43.27	241	44.80	
Albumin (g/l)							
<35	41	33.61	237	56.97	278	51.67	<0.0001
≥35	81	66.39	179	43.03	260	48.33	
Creatinine (μmol/l)							
<80	63	51.64	205	49.28	268	49.81	0.647
≥80	59	48.36	211	50.72	270	50.19	
Small HCC							
No	98	80.33	375	90.14	473	87.92	0.003
Yes	24	19.67	41	9.86	65	12.08	
WHO PS							
0	47	38.52	76	18.27	123	22.86	<0.0001
1	67	54.92	209	50.24	276	51.30	
2	8	6.56	131	31.49	139	25.84	
Child–Pugh							
Class A	87	71.31	217	52.16	304	56.51	0.000
Class B	35	28.69	182	43.75	217	40.33	
Class C	0	0.00	17	4.09	17	3.16	
Okuda stage							
I	83	68.03	138	33.17	221	41.08	<0.0001
II	38	31.15	241	57.93	279	51.86	
III	1	0.82	37	8.89	38	7.06	
CLIP score							
0	14	11.48	18	4.33	32	5.95	<0.0001
1	41	33.61	84	20.19	125	23.23	
2	40	32.79	115	27.64	155	28.81	
3	19	15.57	113	27.16	132	24.54	
4	8	6.56	64	15.38	72	13.38	
5–6	0	0.00	22	5.29	22	4.09	
BCLC classification							
A	9	7.38	5	1.20	14	2.60	<0.0001
B	28	22.95	40	9.61	68	12.64	
C	84	68.85	327	78.61	411	76.39	
D	1	0.82	44	10.58	45	8.36	

BCLC, Barcelona Clinic Liver Cancer group; CLIP, Cancer of the Liver Italian Program; HCC, hepatocellular carcinoma; WHO PS, World Health Organization performance status.

precision of the prediction and the explained variation. According to the survival variability explained by the score, Schemper statistics Vs varied from 13.91 for CLIP (six classes) to 8.73 for BCLC (Table 3). These results highlight higher explained variability by CLIP.

improvement of prognostic scores

Univariate Cox analyses stratified on trial showed that the following variables were significantly associated with lower OS (Table 4): age ≥65 years, alcoholic cirrhosis, jaundice, hepatomegaly, hepatalgia, ascites, involved liver volume >50%, portal vein thrombosis, AFP serum level ≥200 μg/l, total bilirubin and WHO PS greater than zero. Likewise, albumin, prothrombin time and small HCC improved survival.

Multivariate analysis among all these variables retained the following independent and significant baseline predictors: alcoholic cirrhosis, jaundice, hepatomegaly, hepatalgia, ascites, involved liver volume >50%, portal vein thrombosis, AFP level, albumin level, small HCC and WHO PS greater than zero.

Then the three investigated scores could be improved with the following variables (Table 5).

For CLIP: alcoholic cirrhosis, jaundice, hepatalgy and WHO PS.

Table 3. Overall survival and statistical performance according to Okuda, CLIP and BCLC (trial stratification)

	Monocity of gradient					Added information			Discriminatory capacity		% Explained variation
	Overall survival										
	n	Median	95% CI	Log-rank	*P* value	AIC	LR χ^2	*P* value	Royston's *D* (SE)	Harrell's *C*	Schemper
Okuda stage						4780	94.89	<0.0001	1.01 (0.11)	0.64	11.67
I	221	11.10	8.37–12.37								
II	279	4.10	3.47–4.97								
III	38	1.47	0.93–1.80								
Okuda[a]				170.98	<0.0001	4786	86.44	<0.0001			
CLIP score						4759	121.61	<0.0001	0.81 (0.08)	0.66	13.91
0	32	21.00	15.73–26.17								
1	125	12.37	9.60–16.70								
2	155	4.35	3.87–5.73								
3	132	4.73	3.67–5.93								
4	72	2.13	1.37–2.57								
5–6	22	1.73	1.10–2.50								
CLIP score											
0–1	157	14.67	11.83–17.70			4756	119.00	<0.0001	1.01 (0.09)	0.65	13.71
2–3	287	4.57	4.10–5.37								
4–6	94	1.93	1.50–2.40								
CLIP[a]				158.07	<0.0001	4775	97.38	<0.0001			
BCLC classification						4810	66.74	<0.0001	0.79 (0.11)	0.61	8.73
A	14	20.05	13.77–40.43								
B	68	12.37	9.30–17.53								
C	411	4.97	4.33–5.73								
D	45	1.57	1.07–1.90								
BCLC[a]				113.84	<0.0001	4818	55.24	<0.0001			

[a]Score was treated as ordinary instead of dummy.

AIC, Akaike information criterion; BCLC, Barcelona Clinic Liver Cancer group; CI, confidence interval; CLIP, Cancer of the Liver Italian Program; LR, likelihood ratio; SE, standard error.

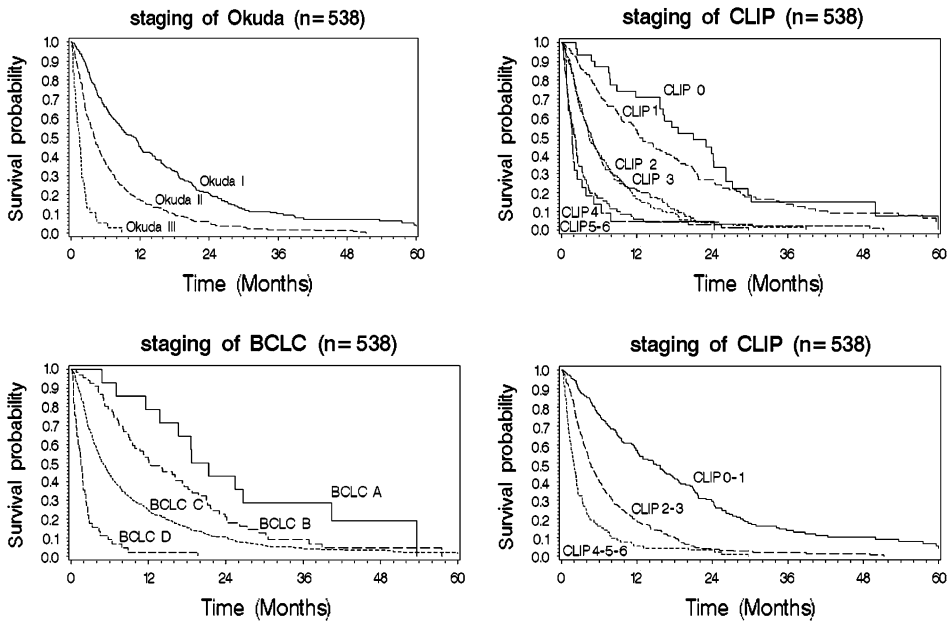

Figure 1. Overall survival (Kaplan–Meier estimate) according to Okuda, BCLC and Cancer of the Liver Italian Program (six and three stages), N = 538.

Table 4. Univariate and multivariate baseline prognostic factor analysis (Cox model)

	Univariate Cox model				Multivariate Cox model (N = 538)			Final multivariate Cox model (N = 538)		
	n	HR	95% CI	*P* value	HR	95% CI	*P* value	HR	95% CI	*P* value
Gender										
Male	478	Ref								
Female	60	0.90	0.68–1.19	0.4596						
Weight (kg)										
<65 (F) and <75 (M)	270	Ref								
≥65 (F) and ≥75 (M)	268	0.94	0.79–1.12	0.5026						
Age (years)										
<65	201	Ref								
≥65	337	0.85	0.70–1.02	0.0718	1.00	0.82–1.21	0.9659			
Cirrhosis										
No	40	Ref								
Yes	498	1.29	0.93–1.81	0.1331						
Alcoholic cirrhosis										
No	124	Ref								
Yes	414	1.30	1.05–1.61	0.0160	1.25	1.00–1.57	0.0481	1.26	1.01–1.57	0.0381
HBV or HCV cirrhosis										
No	439	Ref								
Yes	99	0.89	0.70–1.11	0.2979						
Jaundice										
No	445	Ref								
Yes	93	1.85	1.47–2.33	<0.0001	1.33	1.02–1.73	0.0333	1.42	1.11–1.82	0.0055
Hepatomegalia										
No	137	Ref								
Yes	401	1.56	1.27–1.92	<0.0001	1.34	1.07–1.68	0.0114	1.35	1.08–1.70	0.0085
Edemas of the lower limbs										
No	421	Ref								
Yes	117	1.17	0.95–1.45	0.1436						
Hepatalgia										
No	407	Ref								
Yes	131	1.60	1.31–1.97	<0.0001	1.50	1.20–1.87	0.0004	1.47	1.18–1.83	0.0007
Ascite										
No	378	Ref								
Minimal	121	1.95	1.57–2.41	<0.0001	1.48	1.18–1.86	0.0015	1.52	1.21–1.91	0.0005
Abundant	39	1.99	1.42–2.77	<0.0001	1.48	1.02–2.14		1.53	1.06–2.22	
Tumor localization										
Right	353	Ref								
Left	141	1.02	0.83–1.25	0.8383						
Bilateral	44	1.43	1.02–2.00	0.0393						
Tumor morphology										
Uninodular	166	Ref								
Unilateral multinodular	107	1.19	0.92–1.53	0.1971						
Bilateral multinodular	265	1.23	1.01–1.51	0.0420						
Involved liver volume >50%										
No	394	Ref								
Yes	144	1.61	1.31–1.98	<0.0001	1.28	1.02–1.59	0.0321	1.31	1.05–1.62	0.0172
Portal thrombosis										
No	341	Ref								
Yes	197	1.67	1.39–2.01	<0.0001	1.25	1.03–1.52	0.0219	1.25	1.03–1.52	0.0214
α-Fetoprotein serum level (μg/l)										
<200	273	Ref								
≥200	265	1.87	1.56–2.25	<0.0001	1.71	1.41–2.07	<0.0001	1.75	1.45–2.12	<0.0001
Total bilirubin (μmol/l)										
<20	242	Ref								
≥20	296	1.60	1.33–1.91	<0.0001	1.14	0.93–1.42	0.2156			

Table 4. (*Continued*)

	Univariate Cox model				Multivariate Cox model (N = 538)			Final multivariate Cox model (N = 538)		
	n	HR	95% CI	*P* value	HR	95% CI	*P* value	HR	95% CI	*P* value
Prothrombin time (%)										
<80	297	Ref								
≥80	241	0.77	0.65–0.92	0.0045	0.91	0.75–1.11	0.3369			
Albumin (g/l)										
<35	278	Ref								
≥35	260	0.62	0.52–0.74	<0.0001	0.82	0.67–1.00	0.0446	0.79	0.65–0.96	0.0160
Creatinine (μmol/l)										
<80	268	Ref								
≥80	270	0.94	0.79–1.12	0.4888						
Small HCC										
No	473	Ref								
Yes	65	0.62	0.47–0.82	0.0007	0.57	0.43–0.77	0.0002	0.60	0.45–0.81	0.0006
WHO PS										
0	123	Ref								
1	276	1.48	1.18–1.85	0.0007	1.27	1.01–1.60	<0.0001	1.27	1.01–1.60	<0.0001
2	139	2.31	1.77–3.01	<0.0001	1.85	1.40–2.44		1.84	1.40–2.43	

CI, confidence interval; F, female; HBV, hepatitis virus A; HCV, hepatitis virus B; HCC, hepatocellular carcinoma; HR, hazard ratio; M, male; WHO PS, World Health Organization performance status.

For Okuda: alcoholic cirrhosis, hepatomegaly, jaundice, hepatalgia, portal vein thrombosis, AFP serum level, small HCC and WHO PS.
For BCLC: alcoholic cirrhosis, hepatomegaly, jaundice, hepatalgia and AFP.

AIC and LR $\chi 2$ statistics highlighted that WHO PS and AFP serum level were, respectively, the more informative variables to be added to the CLIP and BCLC scores. Prognostic information of Okuda could be improved by adding AFP and WHO PS (Table 5).

While alcoholic cirrhosis was an independent prognosis factor, whatever score, it was the least informative variable to add.

discussion

Evaluation of prognostic scores on independent population is essential to assess their relative performances and to identify on which population they can be applied. We found that the CLIP staging system produced the best performances on a French population with HCC in palliative setting; discriminatory ability and predictive accuracy were superior to what was measured with the BCLC and the Okuda scores. Nevertheless, differences between scores remain low and none clearly emerges as an unquestionable reference. Overall, predictive accuracy is low, indicating that the investigated variables partly explain the patient prognosis. After adjusting on the prognostic score, other variables remain associated to OS, indicating that patient prognosis prediction should be improved.

This study is the first one to compare prognostic scores on a population with mainly alcoholic HCC etiology which is associated with older age at diagnosis, poor living conditions and other complications due to alcoholism. A recent paper, however, indicates that prognosis of HCCs detected during surveillance is independent of etiology [24]. The major strengths of our study are the quality of the data and the methodology applied to evaluate prognostic scores. We used statistical methods to investigate the calibration, the discrimination, the added information and the predictive accuracy. This study indicates that future works would benefit from following the proposal of Altman and Royston to validate prognostic scores. Furthermore, this analysis pooled data from two RCTs, which limits most of potential biases observed with cohort or case–control studies. In particular, a high standard of follow-up was applied, resulting in a minimal rate of lost to follow-up, a large number of events and an adequate overall statistical power as compared with previous publications.

The population of our study is limited to patients with advanced HCC, representing between 60% and 75% of all patients treated in France [25]. They formed a rather homogenous sample not representative of the whole HCC population, limiting conclusions to the palliative setting. Child–Pugh class C in both trials and portal vein thrombosis in one trial were exclusion criteria, while they are both constituents of the prognostic systems. These may partly explain the low predictive accuracy of all three staging systems studied. Extension to less advanced patients is not straightforward, requiring a separate study.

Identification of the 'best' score is a controversial issue. Several comparative studies [10–12] concluded that the CLIP was superior, others [14, 26] that the BCLC was. Three

Table 5. Evaluation of the independent contribution of baseline variables and prognostic scores

		Multivariate Cox model				−2 log *L*	AIC	LR χ^2
		Trial stratification						
		HR	95% CI	*P* value				
CLIP	CLIP 1	1.18	0.77–0.81		Without covariates	4870.83	4870.83	–
	CLIP 2	2.79	1.83–4.24		CLIP	4749.22	4759.22	121.61
	CLIP 3	2.29	1.49–3.52	<0.0001	+WHO PS	4721.24	4735.24	149.59
	CLIP 4	4.57	2.88–7.24		+Jaundice	4736.86	4748.86	133.97
	CLIP 5–6	3.80	2.11–6.85		+Hepatalgia	4739.28	4751.28	131.55
Alcoholic cirrhosis	Yes	1.37	1.10–1.71	0.0050	+Alcoholic cirrhosis	4744.53	4756.53	126.30
Jaundice	Yes	1.54	1.22–1.94	0.0003	Full model	4692.90	4712.90	177.93
Hepatalgia	Yes	1.49	1.20–1.85	0.0003	−WHO PS	4718.06	4734.06	152.77
WHO PS	PS 1	1.25	1.00–1.58	<0.0001	−Hepatalgia	4705.37	4723.37	165.46
	PS 2	1.96	1.50–2.57		−Jaundice	4704.84	4722.84	165.99
					−Alcoholic cirrhosis	4701.19	4719.19	169.64
Okuda	Okuda stage II	1.53	1.24–1.88	<0.0001	Without covariates	4870.83	4870.83	
	Okuda stage III	4.14	2.78–6.15		Okuda	4775.94	4779.94	94.89
Alcoholic cirrhosis	Yes	1.25	1.00–1.56	0.0461	+α-Fetoprotein	4745.50	4751.50	125.33
Hepatomegalia	Yes	1.40	1.12–1.74	0.0029	+WHO PS	4753.30	4761.30	117.53
Hepatalgia	Yes	1.40	1.13–1.74	0.0022	+Hepatalgia	4760.06	4766.06	110.78
Portal vein thrombosis	Yes	1.26	1.04–1.53	0.0183	+Portal thrombosis	4762.19	4768.19	108.64
α-Fetoprotein (μg/l)	≥200	1.73	1.43–2.08	<0.0001	+Small HCC	4762.87	4768.87	107.96
Small HCC	Yes	0.62	0.47–0.83	0.0010	+Hepatomegalia	4768.56	4774.56	102.27
WHO PS	PS 1	1.30	1.04–1.64	<0.0001	+Alcoholic cirrhosis	4774.06	4780.06	96.77
	PS 2	1.94	1.48–2.54		Full model	4675.39	4695.39	195.44
					−α-Fetoprotein	4707.72	4725.72	163.11
					−WHO PS	4698.83	4714.83	172.00
					−Small HCC	4687.33	4705.33	183.50
					−Hepatomegalia	4684.58	4702.58	186.25
					−Hepatalgia	4684.36	4702.36	186.47
					−Portal thrombosis	4680.86	4698.86	189.97
					−Alcoholic cirrhosis	4679.51	4697.51	191.32
BCLC	BCLC B	1.20	0.64–2.25	0.5667	Without covariates	4870.83	4870.83	
	BCLC C	1.79	0.99–3.22	0.0537	BCLC	4804.09	4810.09	66.74
	BCLC D	4.53	2.29–8.95	<0.0001	+α-Fetoprotein	4769.41	4777.41	101.42
Alcoholic cirrhosis	Yes	1.29	1.04–1.60	0.0212	+Hepatomegalia	4786.03	4794.03	84.81
Jaundice	Yes	1.33	1.03–1.72	0.0293	+Hepatalgia	4787.56	4795.56	83.27
Hepatomegalia	Yes	1.47	1.18–1.83	0.0005	+Jaundice	4795.25	4803.25	75.58
Hepatalgia	Yes	1.55	1.25–1.92	<0.0001	+Alcoholic cirrhosis	4801.66	4809.66	69.17
α-Fetoprotein (μg/l)	≥200	1.86	1.55–2.24	<0.0001	Full model	4722.66	4738.66	148.17
					−α-Fetoprotein	4765.47	4779.47	105.36
					−Hepatalgia	4737.68	4751.68	133.15
					−Hepatomegalia	4735.16	4749.16	135.67
					−Alcoholic cirrhosis	4728.19	4742.19	142.64
					−Jaundice	4727.21	4741.21	143.62

+ seems that variable have been add to the model; − seems that variable was removed from the full model.
AIC, Akaike information criterion; BCLC, Barcelona Clinic Liver Cancer group; CI, confidence interval; CLIP, Cancer of the Liver Italian Program; HCC, hepatocellular carcinoma; HR, hazard ratio; LR, likelihood ratio; WHO PS, World Health Organization performance status.

factors may explain these conflicting results; first of all the investigated population is crucial. Most of the comparative studies were carried out on patients with mainly viral etiology and longer expected survival [10, 11, 13, 17]. BCLC staging system was developed on this kind of population and does not seem adapted to this situation as shown by its lack of discriminatory ability. Previous studies [12, 17] indicate that it is a valuable tool in the choice of treatment on a broader HCC population. A second factor is that overall performances are not strongly different. Values of the investigated measures of performance belong to a tight interval. Random fluctuations as well as inclusion criteria can easily explain the modification in score ranking. It is essential to underline that all scores have limited performances either to

discriminate between high- and low-risk patients or to predict the outcome. Last, it comes up from the review of the literature that statistical analyses carried out to evaluate score performance are not the most appropriate. In particular, only association between variables and survival are evaluated through 'information criterion' or 'measure of gradient'. These statistical criteria, however, have been shown to produce biased results and to depend on the sample size, the number of model variables and model construction among others. Even though such tools are useful to construct scores, they are insufficient to evaluate their competing performances. Wald et al. [27] highlighted that a strong association was necessary but not sufficient to make a good diagnostic variable. Similar arguments have been developed by Pepe et al. [28] for prognostic factors.

Improvement of the scores is a delicate challenge. Due to the specificity of HCC that generally develops on a liver disease, it is appealing to have a score that takes into account the gravity of the hepatic disease, the extension of the tumor as well as the general status of the patient. In our study, WHO PS was associated with survival after adjustment on the CLIP score, making it a good candidate for construction of a new score. Other variables of interest, which were not reported in previous studies [10, 11, 16, 17], include presence of jaundice, hepatomegalia or hepatalgy. They, however, raise concerns due to their dependence upon the clinical exam. Likewise, it would be interesting to investigate whether quality-of-life measures could be more prognostic information than measures of the general PS [29].

Validation of our results and construction of a new score require having at least two independent samples: the first one to construct and calibrate the new proposal and a second for validation. Failing this methodological process would lead to overestimate the performances of any new prognostic score. To continue the statistical analyses on the patients included in the randomized FFCD clinical trial investigating long-acting octreodid treatment versus placebo is promising.

CLIP is already used in advanced HCC as stratification or eligibility criteria for clinical trials [30–32]. Considering the relatively disappointing performances of the three staging systems in terms of discriminatory power, however, it is unquestionable that new prognostic markers of the HCC progression are needed. There is a huge need that fundamental and transfer researches are carried out to better understand the interaction between the liver disease preeminent HCC.

acknowledgements

This work has been presented at the Journées Francophones de Pathologie Digestive in March 2007 (Lyon), at the 2007 American Society for Clinical Oncology Annual Meeting, and in June 2007 at the 9th World Congress on Gastrointestinal Cancer in Barcelona.

references

1. Parkin DM, Bray F, Ferlay J, Pisani P. Estimating the world cancer burden. Globocan 2000. Int J Cancer 2001; 94: 153–156.
2. El-Serag HB, Davila JA, Petersen NJ, McGlynn KA. The continuing increase in the incidence of hepatocellular carcinoma in the United States—an update. Ann Intern Med 2003; 139: 817–823.
3. Remontet L, Esteve J, Bouvier AM et al. Cancer incidence and mortality in France over the period 1978–2000. Rev Epidemiol Santé Publique 2003; 51: 3–30.
4. Okuda K, Ohtsuki T, Obata H et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. Cancer 1985; 56: 918–928.
5. CLIP Group. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. Hepatology 1998; 28: 751–755.
6. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999; 19: 329–338.
7. Chevret S, Trinchet JC, Mathieu D et al. A new prognostic classification for predicting survival in patients with hepatocellular carcinoma. Groupe d'Etude et de Traitement du Carcinome Hepatocellulaire. J Hepatol 1999; 31: 133–141.
8. Ueno S, Tanabe G, Sako K et al. Discrimination value of the new western prognostic system (CLIP score) for hepatocellular carcinoma in 662 Japanese patients. Cancer of Liver Italian Program. Hepatology 2001; 34: 529–534.
9. Dilou N, Patouillard B, Audigier JC. Les classifications de prédiction de survie du carcinome hépatocellulaire. Gastroenterol Clin Biol 2004; 28: 359–366.
10. The Cancer of Liver Italian Program Investigators. Prospective validation of the CLIP score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. Hepatology 2000; 31: 840–845.
11. Levy I, Sherman M, the Liver Cancer Study Group of the University of Toronto. Staging of hepatocellular carcinoma: assessment of the CLIP, Okuda, and Child-Pugh staging systems in a cohort of 257 patients in Toronto. Gut 2002; 50: 881–885.
12. Wildi S, Pestalozzi BC, McCormack L, Clavien PA. Critical evaluation of the different staging systems for hepatocellular carcinoma. Br J Surg 2004; 91: 400–408.
13. Farinati F, Rinaldi M, Gianni S, Naccarato R. How should patients with hepatocellular carcinoma be staged? Validation of a new prognostic system. Cancer 2000; 89: 2266–2273.
14. Marrero JA, Fontana RJ, Barrat A et al. Prognosis of hepatocellular carcinoma: comparison of 7 staging systems in an American cohort. Hepatology 2005; 41: 707–716.
15. Leung TW, Tang AM, Zee B et al. Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system: a study based on 926 patients. Cancer 2002; 94: 1760–1769.
16. Cillo U, Vitale A, Grigoletto F et al. Prospective validation of the Barcelona Clinic Liver Cancer staging system. J Hepatol 2006; 44: 723–731.
17. Barbare JC, Bouché O, Bonnetain F et al. Randomized controlled trial of tamoxifen in advanced hepatocellular carcinoma. J Clin Oncol 2005; 23: 4338–4346.
18. Doffoël M, Bonnetain F, Bouché O et al. Multicentre randomised phase III trial comparing tamoxifen alone or with transarterial lipiodol chemoembolisation (TLC) for unresectable hepatocellular carcinoma (HCC) in cirrhotic patients (Fédération Francophone de Cancérologie Digestive 9402. Eur J Cancer 2008; doi: 10.1016/j.ejca.2008.01.004.
19. Mazzaferro V, Regalia E, Doci R et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 1996; 334: 693–699.
20. Altman D, Royston P. What do we mean by validating a prognostic model? Stat Med 2000; 19: 453–473.
21. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15: 361–387.
22. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004; 23: 723–748.
23. Schemper M. Predictive accuracy and explained variation. Stat Med 2003; 22: 2299–2308.
24. Trevisani F, Magini G, Santi V et al. Impact of etiology of cirrhosis on the survival of patients diagnosed with hepatocellular carcinoma during surveillance. Am J Gastroenterol 2007; 102: 1022–1031.

25. Caumes JL, Nousbaum JB, Bessaguet C et al. Epidemiology of hepatocellular carcinoma in Finistère. Prospective study from June 2002 to May 2003. Gastroenterol Clin Biol 2007; 1: 259–264.
26. Cillo U, Bassanello M, Vitale A et al. The critical issue of hepatocellular carcinoma prognostic classification: which is the best tool available? J Hepatol 2004; 40: 124–131.
27. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? BMJ 1999; 11: 1562–1565.
28. Pepe MS, Janes H, Longton G et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004; 159: 882–890.
29. Yeo W, Mo FK, Koh J, Allen J et al. Quality of life is predictive of survival in patients with unresectable hepatocellular carcinoma. Ann Oncol 2006; 17: 1083–1089.
30. O'Neil BH, Morse MA et al. A phase II study of octreotide LAR in patients with advanced hepatocellular carcinoma and CLIP score ≥3. Gastrointest Cancers Symp 2007 (Abstr 161).
31. Boige V, Raoul JL, Pignon JP et al. Multicentre phase II trial of capecitabine plus oxaliplatin (XELOX) in patients with advanced hepatocellular carcinoma: FFCD 03-03 trial. Br J Cancer 2007; 97: 862–867.
32. A Randomised, Double-Blind, Parallel Group, Multi-Centre, Phase II Study to Assess the Efficacy and Safety of Best Support Care (BSC) Plus ZD 6474 300 mg, BSC Plus ZD 6474 100 mg, and BSC Plus Placebo in Patients With Inoperable Hepatocellular Carcinoma (HCC). http://clinicaltrials.gov/ct2/show/NCT00508001?term=A+Randomised%2C+Double-Blind%2C+Parallel+Group%2C+Multi-Centre%2C+Phase+II+Study+to&rank=1 (6 February 2008, date last acessed).