Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Development and validation of a new prognostic score of death for patients with hepatocellular carcinoma in palliative setting

Caroline Tournoux-Facon[1], Xavier Paoletti[2,3,*], Jean-Claude Barbare[4], Olivier Bouché[5], Philippe Rougier[6], Laetitia Dahan[7], Catherine Lombard-Bohas[8], Roger Faroux[9], Jean Luc Raoul[10], Laurent Bedenne[11], Franck Bonnetain[12]

[1]*Unité d'Epidémiologie et Biostatistique, Inserm CIC P802, CHU Poitiers, Université de Poitiers, France;* [2]*Service de Biostatistique, Institut Curie, Paris, France;* [3]*INSERM U900, Paris, France;* [4]*Délégation à la Recherche Clinique et à l'Innovation, CHU Amiens, France;* [5]*Service d'Hépatogastroentérologie, CHU Reims, France;* [6]*Hépato-gastro-entérologie, oncologie digestive, Hôpital Ambroise Paré, Boulogne, France;* [7]*Service d'Hépatogastroentérologie et Oncologie Digestive, CHRU AP-HM La Timone, Université de la Méditerranée, Marseille, France;* [8]*Service d'Hépatogastroentérologie, CHRU Lyon, France;* [9]*Service d'Hépatogastroentérologie, CH La Roche sur Yon, France;* [10]*Centre Régional de Lutte Contre le Cancer Eugène Marquis, Rennes, European University in Brittany, France;* [11]*FFCD, INSERM U-866, Dijon, France;* [12]*Unité de biostatistiques et de méthodologie, FFCD, INSERM U-866, Dijon, France*

Background & Aims: Patients with hepatocellular carcinoma (HCC) in a palliative setting have a poor prognosis despite recent therapeutic progress. Several prognostic scores, such as the BCLC and the CLIP, have been shown to be useful in helping select treatment options ranging from transplantation to palliative care. However, the discriminatory ability of these scores is inadequate in palliative settings, which concern about 70% of HCC patients. In this paper, we propose and validate a new prognostic score for patients in the palliative setting.

Methods: The prognostic score was developed on a set of 416 patients from a negative randomized clinical trial conducted by the Fédération Francophone de Cancers Digestifs. It was then subsequently validated on a second set of 271 patients from another negative trial. Backward selection was used to identify independent baseline characteristics. Measures of discrimination and predictive values were computed to assess the quality of the developed score. Comparisons with the BCLC and the CLIP – with and without the WHO performance status (PS) score – were performed.

Results: Tumour morphology, portal vein obstruction, metastasis, ascites, jaundice, alpha-foetoprotein, and serum alkaline phosphatase were included in the final score. From the training dataset, three groups of increasing risk were defined, and these were associated with hazard ratios (HR) of 2.13 and HR = 5.72. Similar results were obtained on the validation dataset. This score provides a better discriminatory ability than BCLC and CLIP in this setting. Unfortunately, absolute performances for these scores remain poor.

Conclusions: The new prognostic score and CLIP + PS are recommended in palliative settings. However, new prognostic variables are necessary.

Introduction

Staging systems used to define the prognosis of a particular disease are useful tools to guide treatment options. They are also essential for the selection of patients in clinical trials and for the adequate stratification of the population of interest for randomization. Overall survival (OS) in patients with hepatocellular carcinoma (HCC) is generally predicted using one of the four classical staging systems developed over the last two decades. The Okuda classification system was the first to consider both tumour factors and liver function [1] and the other scores, namely: the Cancer of the Liver Italian Program (CLIP) score [2–4], the Barcelona Clinic Liver Cancer (BCLC) system [5], and the Groupe d'Etudes et de Traitement du Carcinome Hépatocellulaire (GRETCH) classification [6] were developed thereafter. While BCLC is used successfully in clinical practice to help select the most appropriate treatment option – ranging from transplantation to palliative care – no consensus has been reached concerning the best tool to be used when considering patients in the palliative setting. This setting nevertheless represents about 70% of all HCC patients and is the main target for clinical trials. Previous studies comparing these classifications have reported controversial results [7–9]. Several explanations can be proposed. It is widely accepted that the prognosis of HCC patients depends

Keywords: Prognostic factors; Discrimination; Palliative setting; Overall survival.
Received 21 September 2009; received in revised form 2 June 2010; accepted 18 June 2010; available online 20 August 2010

*Corresponding author at: Service de Biostatistique, Institut Curie, Paris, France. Tel.: +33 1 56 24 56 47; fax: +33 1 53 10 40 20.
E-mail addresses: caroline.tournoux-facon@chu-poitiers.fr (C. Tournoux-Facon), xavier.paoletti@curie.net (X. Paoletti), Barbare.Jeanclaude@chu-amiens.fr (J.-C. Barbare), obouche@chu-reims.fr (O. Bouché), philippe.rougier@apr.aphp.fr (P. Rougier), laetitia.dahan@mail.ap-hm.fr (L. Dahan), catherine.lombard@chu-lyon.fr (C. Lombard-Bohas), roger.faroux@chd-vendee.fr (R. Faroux), raoul@rennes.fnclcc.fr (J.L. Raoul), lbedenne@u-bourgogne.fr (L. Bedenne), FBonnetain@dijon.fnclcc.fr (F. Bonnetain).

Abbreviations: HCC, hepatocellular carcinoma; BCLC, Barcelona Clinic Liver Cancer; CLIP, Cancer of the Liver Italian Program; WHO, World health organization; PS, performance status; HR, hazard ratio; OS, overall survival; AFP, alpha-foetoprotein; AIC, Akaike information criterion; 95% CI, 95% confidence interval.

on tumour factors and liver function. The presence and the causes of cirrhosis lead to a wide heterogeneity in the disease and in the population on which the prognostic classifications are applied. Moreover, the statistical analyses used to evaluate score performance were not always appropriate. In particular, only an association between variables and OS was analysed statistically by 'information criterion' or 'measure of gradient', which is necessary but not sufficient to make a good prognostic variable [10]. Specific measures of discrimination ability are required. As suggested by Wildi et al., it may also be useful to establish a different staging system for patients undergoing palliative care [9] and to evaluate this system by appropriate measures.

Collette et al. recently investigated the value of prognostic classifications applied to patients in the palliative setting, mainly with HCC on alcoholic cirrhosis [11]. Using statistical measures specifically dedicated to the evaluation of prognostic models, they studied the performance of the Okuda, CLIP, and BCLC scores. Based on a pooled analysis of 2 randomized clinical trials [12,13], they demonstrated that the CLIP score had the best predictive value when applied to patients in the palliative setting. They noted that the performance of the CLIP could be improved by the addition of the WHO performance status criteria (WHO PS). They also found that the overall performance of the investigated prognostic classifications remained low in this setting. Based on these results, this study aimed to: (i) develop a new prognostic score of HCC for patients in the palliative setting (ii) explore the performance of the CLIP plus the WHO PS score, and (iii) validate these findings on independent data.

Material and methods

Patients

The population of interest consisted of patients non-amenable to curative treatments, that is, surgical resection, liver transplantation, percutaneous ablation, or transarterial chemoembolization. The prognostic score was developed on a first set of patients with HCC in palliative settings from a multicentre randomized clinical trial conducted in France by the Fédération Francophone de Cancérologie Digestive (FFCD). The trial comparing tamoxifen to supportive care failed to demonstrate any benefit on OS [13]. The main eligibility criterion was HCC not suitable for curative treatments. Exclusion criteria were: serum creatinine greater than 130 μmol/L, Child-Pugh class C, WHO performance status greater than 2, and prior treatment with tamoxifen.

Score validation was performed on a second set of palliative care patients from another negative randomized clinical trial (the CHOC trial) [12]. The CHOC trial compared octreotide-retard versus placebo. Patients had to have a diagnosis of HCC that was either histologically or cytogenetically confirmed. They were also selected based on the following three criteria: (i) presence of cirrhosis and a tumour with a measurable mass at least 3 cm in diameter; (ii) presenting clinical features consistent with the diagnosis of HCC as determined by two contrast-enhanced imaging techniques (ultrasonography and/or computed-tomography scan and/or MRI); and (iii) serum alpha-foetoprotein (AFP) $\geqslant$500 μg/L. In addition, patients were required to be ineligible for curative treatments (transplantation, surgery, percutaneous ablation, or chemoembolization), or had relapsed following potentially curative therapy; with a CLIP score of 0–3 and measurable disease. Main exclusion criteria were: presence of hyperglycaemia ($\geqslant$2.5 g/L) or hypoglycaemia; decreased prothrombin time (<50%); and low platelet count (<50,000/μl). All patients in the two trials gave their written consent.

The model was developed with all collected candidate baseline predictors of OS. It included (i) demographic characteristics such as gender and age at diagnosis; (ii) tumour characteristics such as presence of cirrhosis and its origin (alcoholic or viral), Child-Pugh stage, number and size of tumours according to the Milan criteria [14], search for vascular extension (portal vein obstruction investigated by ultrasonography), and metastasis (investigated by chest CT scan and bone scan); (iii) clinical variables including WHO-PS, oedema, encephalopathy, and jaundice and ascites (defined clinically); and (iv) laboratory variables including alpha-foetoprotein (AFP), prothrombin activity, serum bilirubin, albumin, and alkaline phosphatase. Hepatomegaly was recorded but not tested due to the lack of standardisation of this measurement. Missing data were imputed to the median for continuous variables and to the empirical estimate for categorical variables. As first analysis, continuous variables were dichotomized. The cut-off levels were based on the published literature for AFP [15] and at the lower/upper limits of normal values for prothrombin, bilirubin, albumin, and alkaline phosphatase.

Statistical analysis

OS was defined as the time from the date of randomization to either the date of death (all causes) or the date of last follow-up. Survival distributions were estimated by the Kaplan–Meier method.

Variables associated with univariate Log-rank *p*-value less than 10% were selected for multivariate analysis. When variables were highly statistically correlated (Pearson's correlation coefficient >0.60), only the most relevant variable – as judged by the principal investigator – was retained for the multivariate model. Backward selection was performed with a Cox proportional hazards model to identify independent baseline predictors at the 5% level. Proportional hazards assumption was graphically assessed. The prognostic value for a given patient was the combination of the different variables weighted with the regression coefficients included in the multivariate analysis. In order to establish a practical and an easy-to-compute score, the said score was calibrated so that each variable contributed 1 or 2 units. The score was then calculated for each patient by adding the points corresponding to each prognostic factor. Three risk groups were determined by dividing the score range into 3 equal intervals: low risk, intermediate risk, and high risk of death. A sensitivity analysis was also performed to evaluate how much discriminatory power was lost due to dichotomization and simplification. Continuous variables were then modelled using fractional polynomials and the equation score was not simplified. This score equation is provided as supplementary material.

The CLIP score, modified to account for PS, was investigated. Based on multivariate regression adjusted for the CLIP, we estimated the added contribution of the PS. The CLIP score then varied as follows: unchanged for patients with PS 0, +1 point for patients with PS 1, and +2 points for patients with PS 2 or 3.

The validity of this new classification and of the CLIP + PS was then assessed on the second set of patients referred to as the independent validation cohort. No re-estimation was performed. The predictive accuracy of the new scoring system was examined and compared to the classical scoring system by calculating (ii) the explained variation measure proposed by Schemper that quantifies how much variability is captured by the score [16] bearing in mind that perfect scores tend to explain 100% of the variability, (iii) Royston's D statistic [17] and, (iv) Harrell's C discrimination index extended for survival data [18] that measure the ability of the score to distinguish the various groups. D-statistics may be interpreted as the separation (log hazard ratio) between the survival distributions for two independent prognostic groups and C-statistics can be considered to be a proportion of correct predictions, i.e. the proportion of patients with a better prognostic stage who have a better survival. C-index typically ranges between 0.5 (random classification) and 1 (perfect classification).

SAS v9.1 software was used for all analysis except for calculation of C-statistics which was performed with R software.

Results

Baseline patient characteristics

Baseline patient characteristics in the 2 sets are summarized in Table 1. More information has been presented in previous reports [12,13]. A total of 416 patients were included in the training set. Three hundred and eighty-nine patients died and the median OS was 4.4 months (95% confidence interval 95% CI 3.8–5.0) (Supplementary Fig. 1) with a median follow-up of 48 months. Three hundred and seventy-two patients (89%) were male and the median age was 67 ± 8. An alcoholic aetiology was observed in 318 of the 376 patients with cirrhosis (85%). However, most patients were classified as Child-Pugh class A (52%). Only 39 (9%) patients presented a small HCC as defined by 2 or 3 nodes smaller than 3 cm or one node $\leqslant$5 cm (Milan criteria), 164 (39%) presented

Table 1. Baseline characteristics of the 687 patients included in the training and validation sets.

	Training cohort (n = 416)	Validation cohort (n = 271)
Demographic		
Male sex, n (%)	372 (89)	202 (75)
Age, years (mean ± sd)	67 ± 8	68 ± 9
Tumour characteristics, n (%)		
Cirrhosis	376 (90)	216 (80)
Alcoholic origin	318 (76)	160 (77)
Child-Pugh stage		
A	218 (52)	195 (72)
B	181 (44)	73 (27)
C	17 (4)	3 (1)
Non-small HCC[a]	377 (91)	182 (67)
Portal vein obstruction[b]	164 (39)	39 (14)
Metastasis (chest or bone)	68 (16)	63 (23)
Clinical impact, n (%)		
WHO PS score 2-3	130 (31)	56 (21)
Oedema	100 (24)	41(15)
Jaundice	83 (20)	19 (7)
Hepatomegaly	322 (77)	164 (61)
Encephalopathy	7 (2)	2 (1)
Ascites	144 (35)	44 (16)
Laboratory impact, n (%)		
Alpha-foetoprotein >200µg/L	219 (53)	125 (46)
Serum bilirubin >17µM	275 (66)	156 (58)
Prothrombin <80%	249 (60)	150 (55)
Serum albumin <35g/L	238 (57)	125 (46)
Serum alkaline-phosphatase (ULN[c])	151 (36)	62 (23)
Classical prognostic scores, n(%)		
Okuda		
I	139 (33)	72 (27)
II	240 (58)	187 (69)
III	37 (9)	12 (4)
CLIP		
0	18 (4)	9 (3)
1	85 (20)	26 (10)
2	115 (28)	106 (39)
3	112 (27)	96 (35)
4	64 (15)	27 (10)
5	21 (5)	7 (3)
6	1 (1)	0 (0)
BCLC		
A	3 (1)	27 (10)
B	42 (10)	39 (14)
C	327 (79)	187 (69)
D	44 (11)	18 (6)
GRETCH		
A	63 (15)	57 (21)
B	258 (62)	195 (72)
C	95 (23)	19 (7)

[a] With small HCC defined according to Milan criteria: Single <5 cm or 3 tumours <3 cm.
[b] Ultrasonography.
[c] Twice the upper limit of normal.

portal vein obstruction, and 68 (16%) presented chest and/or bone metastasis. Performance status was good, as 68% of patients were class 0 or 1.

According to classical prognostic scores, patients were mostly classified in the intermediate classes: 58% were Okuda II, 55% were CLIP 2 or 3, 79% were BCLC class C, and 62% were GRETCH class B. Similar baseline characteristics were observed in the validation set [13]. A total of 271 patients were included and 251 patients died. As presented in Fig. 1, patients had a slightly longer OS (median 6.8 months 95% CI 5.9–7.9). The median follow-up was 34 months.

Score development

Results from univariate analysis are presented in Table 2. All tested baseline variables were related to OS with a p value less than 10%. Due to the correlation between ascites and serum albumin and between jaundice and serum bilirubin, only clinical variables (ascites and jaundice) were tested in multivariate analysis.

The final model contained tumour morphology, portal vein obstruction, metastasis, WHO PS score, ascites, jaundice, AFP, and serum alkaline phosphatase (more than twice the upper limit of normal) (Table 3).

The calculated scores are presented in Table 4. The new score ranges from 0 to 10. A low-risk group (0–3 points), an intermediate-risk group (4–6 points), and a high-risk group (7–10 points) were defined (Table 5, Fig. 1). Thirty-five percent (35%) of patients were classified in the low-risk group, 49% in the intermediate-risk group, and 16% in the high-risk group. Each group had a statistically different prognosis (log rank for trend $p < 10^{-4}$).

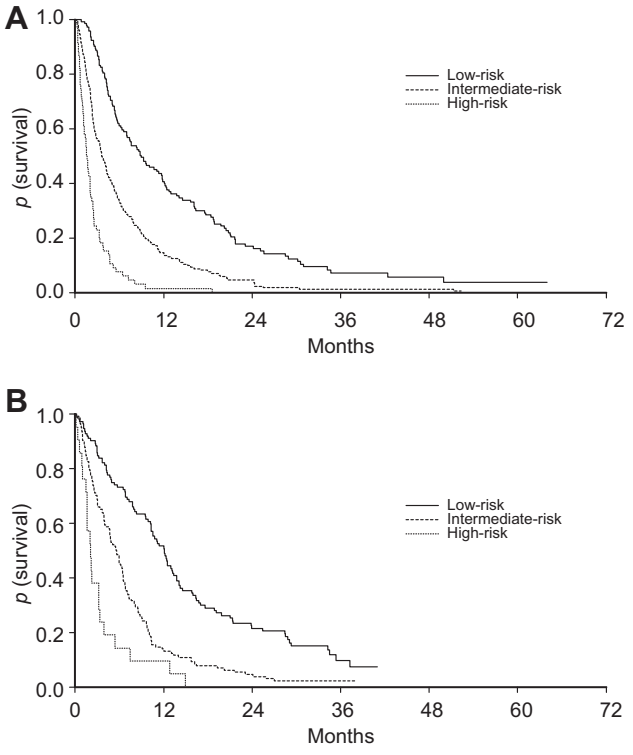

Fig. 1. Kaplan–Meier survival curves. (A) Training cohort. (B) Validation cohort according to the new prognostic score.

Table 2. Univariate prognostic analyses on the training set.

Characteristics, n(%)	Censored n = 27	Death n = 389	*p* value
Tumour characteristics			
Alcoholic origin	19 (70)	299 (77)	0.056
Non-small HCC[a]	25 (93)	352 (90)	0.089
Portal vein obstruction[b]	7 (24)	157 (40)	$<10^{-4}$
Metastasis (chest or bone)	2 (7)	66 (17)	$<10^{-2}$
Clinical impact			
WHO PS score 1	13 (48)	197 (51)	$<10^{-2}$
WHO PS score 2-3	3 (11)	127 (33)	$<10^{-4}$
Oedema	3 (11)	97 (25)	0.083
Jaundice	1 (4)	82 (21)	$<10^{-4}$
Ascites	1 (4)	143 (37)	$<10^{-4}$
Biological impact			
Alpha-foetoprotein >200µg/L	8 (30)	211 (54)	$<10^{-4}$
Serum bilirubin >17µM	13 (48)	262 (67)	$<10^{-4}$
Prothrombin <80%	12 (44)	237 (61)	0.011
Serum albumin <35g/L	9 (33)	229 (59)	$<10^{-4}$
Serum alkaline-phosphatase (>2N[c])	7 (26)	144 (37)	$<10^{-4}$

[a] With small HCC defined according to Milan criteria: Single <5 cm or 3 tumours <3 cm.
[b] Ultrasonography (US).
[c] Twice the upper limit of normal.

Compared to the low-risk category, the intermediate and the high-risk categories were associated with HR = 2.13 (95% CI 1.7–2.68) and HR = 5.72 (95% CI 4.16–7.86), respectively.

Validation

The distribution of the population between the three defined groups with the new score was similar for the validation cohort: 41% were classified in the low-risk group, 51% in the intermediate-risk group, and 8% in the high-risk group. The estimated HRs were also similar between the two sets: intermediate-risk group compared to low-risk group was associated with HR = 2.29 (95% CI 1.75–3.02) in the validation set and the HR of the high-risk group compared to low-risk group was equal to 4.90 (95% CI 3.01–7.98) in the validation set (Table 5).

The new prognostic score and the CLIP + PS provided a better predictive accuracy, as they explained between 10.8% and 11.9% of the variability corresponding to a 17% increase compared to the CLIP score (Table 6). They also presented the highest separability index (*D*-statistic = 1.01 and 0.86, respectively) and *C*-statistics. This discriminatory measure was similar for the new score and CLIP (0.63 ± 0.034 and 0.63 ± 0.036, respectively) and not significantly superior to CLIP + PS (0.66 ± 0.036). The similar good relative performances of the CLIP + PS and the new score were confirmed in the validation set. However, the overall absolute performances of all these scores were low. As a sensitivity analysis, the concordance index of a more complete (and less easy-to-use) equation score without dichotomization was 0.65 (95% CI 0.61–0.69), which was not significantly different from 0.63.

Discussion

In this study, we have investigated the prognostic value of the CLIP + WHO PS and we have developed and validated on two independent samples a score specifically designed for the 70% of HCC patients non-eligible for curative treatments. Eight variables were selected to define the new prognostic score: number and size of tumours Milan criteria), presence of US portal vein obstruction and metastasis, three clinical variables (PS score, presence of ascites and jaundice), and two laboratory variables (AFP and alkaline phosphatase).

Three risk groups were derived with increasing estimated risk of death. The quality of this score was investigated in terms of its association with OS, its discriminatory performance, and its predictive accuracy. This score showed a better performance than

Table 3. Multivariate Cox proportional-hazards analysis of the training set and assignment of points.

Prognostic factor	Hazard ratio [95% CI] n = 416	*p* value	β-regression coefficient	Points[†]
Non-small HCC	1.53 [1.07-2.19]	0.02	0.428	1
Portal vein obstruction (US)	1.40 [1.13-1.73]	$<10^{-2}$	0.337	1
Metastasis	1.60 [1.22-2.10]	$<10^{-4}$	0.471	1
WHO PS score 1	1.41 [1.07-1.88]	0.02	0.346	1
WHO PS score 2-3	1.88 [1.38-2.58]	$<10^{-4}$	0.635	2
Jaundice	1.38 [1.07-1.79]	0.01	0.322	1
Ascites	1.59 [1.28-1.99]	$<10^{-4}$	0.466	1
Alpha-foetoprotein >200µg/L	1.49 [1.21-1.83]	$<10^{-3}$	0.397	1
Serum alkaline-phosphatase (>2N[1])	1.65 [1.32-2.07]	$<10^{-4}$	0.502	2

[a] Twice the upper limit of normal.
[†] Assignment of points to risk factors was based on a linear transformation of the corresponding regression coefficient. The coefficient of each variable was divided by the lowest value and rounded to the nearest integer.

Table 4. New scoring system.

Prognostic factor	0 point	1 points	2 points
Tumour morphology	Single <5 or 3 tumours <3cm	Other	
Portal vein obstruction (US)	no	yes	
Metastasis	no	yes	
WHO PS score	class 0	class 1	class 2-3
Jaundice	no	yes	
Ascites	no	yes	
Alpha-foetoprotein >200μg/L	no	yes	
Serum alkaline-phosphatase (>2N[a])	no		yes

[a] Twice the upper limit of normal.

previous tools on both the training and the validation sets. Previous results [11] that had demonstrated that the CLIP performed well in this setting were also validated; moreover, the addition of the WHO performance status to the CLIP was shown to significantly improve the precision for the prognosis of OS. Both combinations of variables led to similar discriminatory performances.

Previous scores were developed on broader therapeutic situations ranging from transplantation to palliative care. The CLIP and BCLC, the two scores most commonly used, have been validated as practical tools to support decision-making [7,19]. Our findings suggest that patients in the palliative setting with BCLC stage C or CLIP score 0–3 have a fairly heterogeneous prognosis. Clinical trials in this latter group could benefit from refinements in the stratification of included patients. This would improve control of heterogeneity in the trial and would allow for better-orienting of patients to trials in specific clinical phases. Three components appear to be essential in the evaluation of the prognosis of these patients: tumour characteristics, underlying liver disease, and general performance status. Adding PS to the CLIP or measurement of extrahepatic tumour extension for BCLC [11] improves the performance of these scores in the palliative setting.

This setting was defined in a rather pragmatic way: all patients who were not eligible for a curative treatment (for instance, surgery or chemo-embolization) were considered in the palliative setting. This definition is not uncommon in clinical trials [20], and is largely used in the current practice in several countries as stated by different recommendations [21,22]. This population essentially corresponds to the BCLC stages C and D that represented 90% of our population. In the trial used for external validation, some additional exclusion criteria (hyperglycaemia or low platelet count) related to the treatment under study (octreotide) applied. As these two criteria are not a part of any definition of the palliative setting in HCC and are not known to be prognostic factors, it was assumed that they would not bias our conclusions and limit their applicability.

An important issue is that the scores have been constructed, compared, and validated on patients recruited at the pre-molecular targeted agents (MTA) era. In most developed countries, these patients would receive Sorafenib. However, to our knowledge,

Cancer

Table 5. Hazard ratio of the risk of death in the training and validation sets according to risk category.

Risk category	Training set		Validation set	
	n (%)	Hazard ratio [95% CI]	n (%)	Hazard ratio [95% CI]
Low [0-3 points]	143 (35)	1	112 (41)	1
Intermediate [4-6 points]	206 (49)	2.13 [1.70-2.68]	138 (51)	2.29 [1.75-3.02]
High [7-11 points]	67 (16)	5.72 [4.16-7.86]	21 (8)	4.90 [3.01-7.98]

Table 6. Comparison of prognostic scores on the training and validation set.

	Training set				Validation set			
	AIC	Schemper	Royston	C-Harrell (se)	AIC	Schemper	Royston	C-Harrell (se)
Okuda	3913	9.63	0.99	0.62 (0.028)	2389	2.0	0.44	0.54 (0.033)
BCLC	3928	8.63	0.94	0.61 (0.024)	2385	3.8	0.53	0.58 (0.033)
GRETCH	3910	10.00	0.97	0.63 (0.027)	2369	6.3	0.79	0.59 (0.032)
CLIP	3906	11.55	0.79	0.64 (0.031)	2361	9.3	0.78	0.63 (0.036)
CLIP+PS	3894	13.28	0.89	**0.67 (0.031)**	2351	**11.9**	0.86	**0.66 (0.036)**
New score	**3884**	**13.51**	**1.16**	0.66 (0.025)	**2346**	10.8	**1.01**	0.63 (0.034)

Bold figures correspond to the best values.
AIC, Akaike Information Criterion.

the Sorafenib has not been shown to have a differential effect in subpopulations defined by the variables included in our scores. No interaction between the variables investigated in our prognostic scores and the Sorafenib treatment have been reported. Moreover, Sorafenib will modify the prognosis of the patients but cannot be considered as a curative treatment. Therefore, our proposed score and the CLIP + PS should be applicable in a Sorafenib-treated population.

The use of patients derived from randomized clinical trials to assess the prognostic classification ensured good quality data with few missing data and adequate follow-up. As the two trials failed to demonstrate any benefit of experimental treatment [12,13], the influence of treatment should be limited when considering the prognostic value of the variables investigated. Patient samples in clinical trials are often more homogeneous due to inclusion criteria that exclude patients with poor general prognosis and patients with less aggressive disease (to be eligible to chemoembolization). The prognostic variability and representativity of the results can therefore be considered to be more limited. Nevertheless, the inclusion criteria in these trials were fairly broad and probably relatively representative of the typical palliative HCC population eligible for clinical trials. In addition, several protocol violation led to include patients with more advanced disease such as 20 patients categorized Child-Pugh C after review of the baseline characteristics. Likewise, the percentage of patients with an alcoholic aetiology was high (70%) corresponding to the usual percentage reported in France [23].

The present prognostic score was based on the most recent progress in HCC clinical practice, using refined cut-off values to define pathological values for serum bilirubin, AFP [15], prothrombin, and serum albumin, from those used in the CLIP or BCLC. Our results and conclusions were also supported by stringent statistical analysis of the required prognostic properties, based on performance measures that are designed to evaluate the discriminatory and predictive ability of the score. The *p*-values and information criteria (AIC) are the most commonly reported measures of prognostic accuracy, and although they are useful to construct scores, they are influenced by the sample size, the number of variables included in the model, and the model construction methods, making them inappropriate as the unique measure of prognostic performance [10]. Moreover, validation of results and construction of a new score must be based on at least two independent samples: the first sample is used to construct and calibrate the new proposed score and the second sample is used for validation. As a result of this methodology, the results for the proposed score should not be too excessively subject to a risk of overfitting and the associated overestimation of performances.

A potential limitation of this study concerns the high prevalence of alcoholic cirrhosis in the study population (70%). This prevalence varies considerably from one country to another and even across large countries where major differences are observed between urban and rural areas or according to ethnicity [24,25]. Previous publications in a population of patients wherein 30% have hepatitis cirrhosis (HBV or HCV only) tend to indicate that the aetiology of cirrhosis should not have a major impact on the prognosis of patients [26]. We did not find that aetiology was an important prognostic factor in the two trials considered for this study, although the impact of aetiology has not been clearly elucidated.

Our results also emphasize the poor "absolute" performance of all prognostic criteria. A Harrell C less than 0.66 indicates a result that is slightly better than a simple random classification ($c = 0.5$). None of the proposed scores can successfully explain the variability of time to death. Even though our statistical modelling choices (variables selection, continuous variables dichotomization, type of model) may have contributed to the prognostic qualities of the new proposed score, this degradation is probably weak. As shown by the supplementary analysis investigating a more refined equation score, extra complexity entailed only a very modest improvement. This suggests that major variables influencing the prognosis of HCC have not yet been identified. Other candidate variables should be investigated to more accurately predict OS. Quality of life has been identified as an important prognostic factor in this setting and could be proposed as an informative variable for the prognosis of these patients [26,27].

Although the individual OS cannot be accurately predicted, we suggest using the proposed score or the CLIP + WHO PS in clinical trials investigating HCC in palliative settings. They include simple clinical and laboratory variables that are used in practice. New prognostic variables, more specific of the stage of disease and treatment modalities, are necessary to allow better selection of HCC patients for clinical trials in the non curative setting.

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this paper.

Acknowledgements

We are grateful to all the patients who participated in the studies and to all investigators who treated them and who collected the data. We express our gratitude to Professors M. Schemper (Medical University of Vienna, Austria) and P. Royston (MRC, Clinical Trials Unit, England) for providing SAS macros. We thank Dr. Antony Saul for editing the manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhep.2010.06.015.

References

[1] Okuda K, Ohtsuki T, Obata H, Tomimatsu M, Okazaki N, Hasegawa H, et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. Cancer 1985;56:918–928.
[2] A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. Hepatology 1998;28:751–55.
[3] Ueno S, Tanabe G, Sako K, Hiwaki T, Hokotate H, Fukukura Y, et al. Discrimination value of the new western prognostic system (CLIP score) for hepatocellular carcinoma in 662 Japanese patients. Cancer of the Liver Italian Program. Hepatology 2001;34:529–534.
[4] Prospective validation of the CLIP score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. The Cancer of the Liver Italian Program (CLIP) Investigators. Hepatology 2000;31:840–45.
[5] Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19:329–338.
[6] Chevret S, Trinchet JC, Mathieu D, Rached AA, Beaugrand M, Chastang C. 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.
[7] Huang YH, Chen CH, Chang TT, Chen SC, Wang SY, Lee HS, et al. Evaluation of predictive value of CLIP, Okuda, TNM and JIS staging systems for hepatocellular carcinoma patients undergoing surgery. J Gastroenterol Hepatol 2005;20:765–771.
[8] Levy I, Sherman M. 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.
[9] 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.
[10] Pepe M. the statistical evaluation of medical tests for classification and prediction. Oxford Statistical University 2003.
[11] Collette S, Bonnetain F, Paoletti X, Doffoel M, Bouche O, Raoul JL, et al. Prognosis of advanced hepatocellular carcinoma: comparison of three staging systems in two French clinical trials. Ann Oncol 2008;19:1117–1126.
[12] Barbare JC, Bouche O, Bonnetain F, Dahan L, Lombard-Bohas C, Faroux R, et al. Treatment of advanced hepatocellular carcinoma with long-acting octreotide: a phase III multicentre, randomised, double blind placebo-controlled study. Eur J Cancer 2009;45:1788–1797.
[13] Barbare JC, Bouche O, Bonnetain F, Raoul JL, Rougier P, Abergel A, et al. Randomized controlled trial of tamoxifen in advanced hepatocellular carcinoma. J Clin Oncol 2005;23:4338–4346.
[14] Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 1996;334:693–699.
[15] Soresi M, Magliarisi C, Campagna P, Leto G, Bonfissuto G, Riili A, et al. Usefulness of alpha-fetoprotein in the diagnosis of hepatocellular carcinoma. Anticancer Res 2003;23:1747–1753.
[16] Schemper M. Predictive accuracy and explained variation. Stat Med 2003;22:2299–2308.
[17] Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;23:723–748.
[18] Harrell Jr 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.
[19] Cillo U, Vitale A, Grigoletto F, Farinati F, Brolese A, Zanus G, et al. Prospective validation of the Barcelona Clinic Liver Cancer staging system. J Hepatol 2006;44:723–731.
[20] Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Gane E, et al. Sorafenib in advanced hepatocellular carcinoma. NEJM 2008;359 (4):378–390.
[21] HAS/Service des recommandations professionnelles – service évaluation médico-économique et santé publique Septembre 2007.
[22] Verslype C, Van Cutsem E, Dicato M, Arber N, Berlin JD, Cunningham D et al. The management of hepatocellular carcinoma. Current expert opinion and recommendations derived from the 10th World Congress on Gastrointestinal Cancer, Barcelona, 2008. Ann Oncol. 2009;20:vii1–vii6.
[23] Borie F, Tretarre B, Bouvier AM, Faivre J, Binder F, Launoy G, et al. Primitive liver cancers: epidemiology and geographical study in France. Eur J Gastroenterol Hepatol 2009;21:984–989.
[24] Marrero JA. Hepatocellular carcinoma. Curr Opin Gastroenterol 2003;19:243–249.
[25] Wong R, Corley DA. Racial and ethnic variations in hepatocellular carcinoma incidence within the United States. Am J Med 2008;121: 525–531.
[26] Trevisani F, Magini G, Santi V, Morselli-Labate AM, Cantarini MC, Di Nolfo MA, 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.
[27] Bonnetain F, Paoletti X, Collette S, Doffoel M, Bouche O, Raoul JL, et al. Quality of life as a prognostic factor of overall survival in patients with advanced hepatocellular carcinoma: results from two French clinical trials. Qual Life Res 2008;17:831–843.