Research Article

The Detection of Hepatocellular Carcinoma Using a Prospectively Developed and Validated Model Based on Serological Biomarkers

Philip J. Johnson[1,3], Sarah J. Pirrie[2], Trevor F. Cox[4], Sarah Berhane[1,3], Mabel Teng[1], Daniel Palmer[3], Janet Morse[1], Diana Hull[2], Gillian Patman[5], Chiaki Kagebayashi[7], Syed Hussain[3], Janine Graham[6], Helen Reeves[5], and Shinji Satomura[7]

Abstract

Background: Hepatocellular carcinoma is a common complication of chronic liver disease (CLD), and is conventionally diagnosed by radiological means. We aimed to build a statistical model that could determine the risk of hepatocellular carcinoma in individual patients with CLD using objective measures, particularly serological tumor markers.

Methods: A total of 670 patients with either CLD alone or hepatocellular carcinoma were recruited from a single UK center into a case–control study. Sera were collected prospectively and specifically for this study. A logistic regression analysis was used to determine independent factors associated with hepatocellular carcinoma and a model built and assessed in terms of sensitivity, specificity, and proportion of correct diagnoses.

Results: The final model involving gender, age, AFP-L3, α fetoprotein (AFP), and des-carboxy-prothrombin ("GALAD") was developed in a "discovery" data set and validated in independent data sets both from the same institution and from an external institution. When optimized for sensitivity and specificity, the model gave values of more than 0.88 irrespective of the disease stage.

Conclusions: The presence of hepatocellular carcinoma can be detected in patients with CLD on the basis of a model involving objective clinical and serological factors. It is now necessary to test the model's performance in a prospective manner and in a routine clinical practice setting, to determine if it may replace or, more likely, enhance current radiological approaches.

Impact: Our data provide evidence that an entirely objective serum biomarker–based model may facilitate the detection and diagnosis of hepatocellular carcinoma and form the basis for a prospective study comparing this approach with the standard radiological approaches. *Cancer Epidemiol Biomarkers Prev; 23(1); 144–53. ©2013 AACR.*

Introduction

Hepatocellular carcinoma was, until recently, diagnosed on the basis of histologic examination of tumor tissue but the diagnosis can now be established with a high degree of specificity on the basis of characteristic radiological features once tumors are greater than 1 cm in diameter (1–3). Furthermore, such an approach obviates both the risk of bleeding and tumor seeding along the biopsy tract (4, 5).

Surveillance, well accepted to be the key to effective delivery of potentially curative treatment (1, 2), involves ultrasound examination (USS; refs. 1, 2, 6) followed, where a suspicious lesion is detected, by confirmatory tests, including conventional computed tomography (CT) or MRI scanning with or without biopsy. Estimation of serum α-fetoprotein (AFP) has also been used for diagnosis, with grossly elevated levels being highly specific for hepatocellular carcinoma (7) but, as the importance of early diagnosis has become apparent, the limited sensitivity of AFP for hepatocellular carcinoma in smaller tumors has reduced its value (1, 2, 8, 9). Other serological diagnostic tests include des-carboxy-prothrombin (DCP; an abnormal prothrombin molecule derived from an acquired defect in the posttranslational carboxylation of

Authors' Affiliations: [1]School of Cancer Sciences; [2]Cancer Research UK Clinical Trials Unit, School of Cancer Sciences, University of Birmingham, Birmingham; [3]Department of Molecular and Clinical Cancer Medicine; [4]Liverpool Cancer Research UK Centre, Liverpool Cancer Trials Unit, University of Liverpool, Liverpool; [5]Northern Institute for Cancer Research, Newcastle University; [6]Northern Centre for Cancer Care, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom; and [7]Wako Life Sciences, Inc., Mountain View, California

Corresponding Author: Philip J. Johnson, Department of Molecular and Clinical Cancer Medicine, The Duncan Building, Daulby Street, Liverpool L69 3GA, United Kingdom. Phone: 151-706-4172; Fax: 151-706-5826; E-mail: philip.johnson@liverpool.ac.uk

American Association for Cancer Research

the prothrombin precursor; refs. 10, 11) and AFP-L3 (an isoform of AFP characterized by the presence of an α 1-6–linked residue on the AFP carbohydrate side chain; refs. 12, 13).

The limitations of radiological diagnosis of hepatocellular carcinoma are, however, being increasingly recognized in terms of diagnosis of new cases and in the screening setting (14). Both the number of non-hepatocellular carcinoma lesions and pseudolesions, which may collectively be more common than hepatocellular carcinoma lesions in the cirrhotic liver, and the necessity for considerable expertise in liver-imaging, have been noted (14). The limitations of ultrasound for screening are also becoming apparent. It has limited sensitivity, usually quoted at between 65% and 80%, but rather lower in early disease in which appearances are not specific and performance characteristics have not been well defined in nodular cirrhotic liver undergoing surveillance. Furthermore, it is subjective and dependent on operator experience and the available equipment (15–20). Again, it is increasingly recognized that although screening by USS may be effective in specialized centers, this does not necessarily translate into an effective screening system in the wider community (21). Increasing levels of obesity in the West also limit the sensitivity of USS (22).

For all these reasons, we consider here the possibility of establishing the diagnosis of hepatocellular carcinoma in the clinical and, potentially, in the screening setting, using entirely objective measurements, mainly the three serological tests AFP, DCP, and AFP-L3 by developing a statistical model. Being cognizant of the concern that any statistical model would need to perform well in the early-disease setting, we were careful to collect the clinical material such that patients could be rigorously classified as to their disease stage.

Materials and Methods

This case–control study involved 670 patients, 331 with hepatocellular carcinoma and a control group of 339 patients with chronic liver disease (CLD), alone. The patients with hepatocellular carcinoma were recruited at the Queen Elizabeth Hospital (Birmingham, UK) from among patients who were approached and consented to the study between 2007 and 2012 (Table 1). For all patients, the diagnosis was established by the histologic examination of tumor tissue (23%) or characteristic radiotherapy according to international guidelines (1, 2). Samples were taken at the time of first referral for treatment or further investigation. The median period between sample acquisition and formal diagnosis on the basis of CT scan or histology was 1.7 months. No treatment was administered between sample acquisition and formal diagnosis of hepatocellular carcinoma. CLD control samples ($n = 339$) were recruited from patients who were attending outpatient clinics for CLD in the same institution and classified as hepatitis B virus (HBV)-related, hepatitis C virus (HCV)-related, alcoholic-related, and "other" or "no" underlying

CLD. The "other" group comprised patients with hemochromatosis, primary biliary cirrhosis, nonalcoholic steatohepatitis, or cryptogenic cirrhosis. The diagnosis of CLD was made on the basis of liver biopsy and/or typical clinical and imaging features. None of the CLD-control group had evidence of hepatocellular carcinoma at the time the relevant serum sample was taken or within a minimum follow-up period of 6 months (Table 1), but three of them developed hepatocellular carcinoma between 6 and 12 months. For the purpose of analysis, these three patients remained assigned to the CLD group. An age- and sex-matched control group of 92 subjects without any evidence of liver disease were recruited from patients with upper gastrointestinal symptoms who had no clinically significant abnormal findings at endoscopy. All patients gave informed consent for donating blood and the study procedure was approved by the South Birmingham Research Ethics Committee or The Newcastle and North Tyneside Ethics Committee. A standard operating procedure was applied to all blood collection. The Birmingham serum samples were collected prospectively for the discovery and internal validation sets, specifically for this research project and according to the REMARK guidelines (23, 24). The "discovery" set comprised 218 patients with hepatocellular carcinoma seen between April 2007 and January 2011. This sample size was based upon the calculation that approximately 200 subjects per group would be sufficient, using a two-sided test, to reject the null hypothesis of AUROC (area under the receiver operator curve) = 0.75 in favor of AUROC = 0.85 with 90% power for a significance level 0.05. On the basis of Harrell's rule of thumb, this sample size is also sufficient to allow the fitting of up to 20 candidate variables within a logistic discrimination model. The validation set comprised 113 patients with hepatocellular carcinoma seen between February 2011 and March 2012. The external validation set was from a previously reported study (25) designed to assess surveillance biomarkers in fatty liver disease (alcoholic and nonalcoholic). All markers were measured again on stored sera (Table 1) and collected specifically and prospectively for biomarker assessment.

Patients were classified as having "early" or "late" disease on the basis of three staging systems: tumor–node–metastasis (TNM) 6, Barcelona Liver Cancer Clinic (BCLC; refs. 26–30), Milan criteria (31), or on an "operational" basis. Stages I and II of TNM 6 and BCLC stages 0 and A were classified as early disease and, as an additional measure to disease stage, tumor size equal to or below 5 cm was considered early, whereas a size of more than 5 cm was considered late irrespective of tumor number. Those within and outside Milan criteria were categorized as early and late, respectively. Early and late disease was "operationally" classified on the basis of whether or not an experienced multidisciplinary team recommended potentially curative treatment. Where patients were listed for transplantation but had TransArterial ChemoEmbolization as initial treatment

Table 1. Characteristics of patients with hepatocellular carcinoma and CLD

Variable	Birmingham		Newcastle	
	HCC ($n = 331$)	CLD ($n = 339$)	HCC ($n = 63$)	CLD ($n = 100$)
Demographics				
Median age, y	66.3 (59.2–72.9)	53 (45–63)	68.6 (62.2–75.0)	63.6 (57.3–69.3)
Mean age, y ($\pm$ SD)	65.3 ($\pm$ 9.9)	52.8 ($\pm$ 13.7)	68.3 ($\pm$ 8.5)	62.2 ($\pm$ 11.2)
Gender (M:F)	272:59	214:125	53:10	42:58
Etiology				
Alcohol	81	53	27	17
HCV	43	74	0	0
HBV	30	58	0	0
HBV+HCV	2	6	0	0
Other	78	75	27	83
Hemochromatosis	6	3	2	0
Autoimmune	2	28	0	0
PBC	11	18	2	1
NASH/NAFLD	17	12	12	81
Cryptogenic/other	42	14	11	1
Noncirrhotic	34	10	9	0
Multiple (more than one etiology)[a]	48	58	0	0
Unknown	15	5	0	0
CLD (No:Yes:NK)	37:283:11	13:320:6	9:54:0	0:100:0
HCC biomarkers				
AFP (ng/mL)	57 (8.3–1438) $n = 331$	2.8 (2–4.7) $n = 339$	44.5 (6.1–1501.9) $n = 63$	3.2 (2.3–4.7) $n = 99$
Log_{10} AFP (ng/mL)	1.76 (0.92–3.16) $n = 331$	0.4 (0.3–0.7) $n = 339$	2.1 ($\pm$ 1.6) $n = 63$	0.5 (0.4–0.7) $n = 99$
L3 (%)	16.6 (7–51.9) $n = 319$	1(1–7.1) $n = 339$	24.5 (8.1–49.4) $n = 63$	1 (1–7.7) $n = 99$
Log_{10} L3 (%)	1.2 (0.85–1.72) $n = 319$	0 (0–0.8) $n = 339$	1.2 ($\pm$ 0.6) $n = 63$	0 (0–0.9) $n = 99$
DCP (ng/mL)	20.8 (2.6–169.7) $n = 320$	0.35 (0.27–0.6) $n = 339$	16.3 (3.0–102.7) $n = 63$	0.5 (0.4–0.8) $n = 99$
Log_{10} DCP (ng/mL)	1.37 ($\pm$ 1.2) $n = 320$	−0.5 (−0.6 to −0.2) $n = 339$	1.4 ($\pm$ 1.2) $n = 63$	−0.3 (−0.4 − −0.1) $n = 99$
LFTs				
Albumin (g/L)	39 (34.3–43) $n = 330$	44 (40–46) $n = 339$	35.9 ($\pm$ 5.6) $n = 63$	44 (41–47) $n = 100$
ALP (U/L)	370 (258.3–568.8) $n = 330$	213 (163–313) $n = 339$	177 (127–260) $n = 63$	89 (71.8–123.8) $n = 100$
INR	1.1 (1.0–1.2) $n = 324$	1 (1–1.1) $n = 333$	1 (1–1.2) $n = 63$	1 (0.9–1.1) $n = 100$
Bilirubin (μmol/L)	17 (11–28) $n = 330$	11 (8–19.5) $n = 339$	17 (12–30) $n = 63$	9 (6–13) $n = 100$
Creatinine (μmol/L)	77 (64–96) $n = 330$	72 (63–85) $n = 339$	100.7 ($\pm$28.6) $n = 63$	88.7 ($\pm$ 18.7) $n = 100$
Child–Pugh score				
A:B:C:NK	245:73:10:3	291:43:4	40:12:11	NK
Tumor characteristics				
Vascular invasion (No:Yes:NK)	225:87:19	NA	45:18:0	NA
Milan criteria (No:Yes: NK)	223:82:26	NA	49:14:0	NA
Treatments				
Curative (intended:actual)	67:52	NA	NK	NA
Palliative (intended:actual)	255:255	NA	NK	NA
Not known (intended:actual)	9:24			
Biopsy confirmation of HCC/CLD	77 (23%)	88 (26%)	27 (42.9%)	NA

NOTE: For all continuous variables, values are presented either as median (interquartile range) or mean ($\pm$SD), the latter for normal distributions where appropriate.

Abbreviations: ALP, alkaline phosphatase; F, female; HCC, hepatocellular carcinoma; INR, international normalized ratio; M, male; NA, not applicable; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NK, not known; PBC, primary biliary cirrhosis.

[a]For example, alcoholic and HCV-positive.

as a "bridge" to transplantation, they were classified as having early disease.

Routine liver and renal function tests (LFT and RFT) were measured on an automated analytical platform (the Roche Cobas 8000 Modular system) and the severity of the liver disease was defined according to the Child–Pugh score (32, 33). The hepatitis B surface antigen (HBsAg) and anti-HCV antibodies were measured using the e602 module (employing electrochemiluminescence technology) on the Roche Cobas 8000 system.

Assays of AFP, AFP-L3%, and DCP

AFP, AFP-L3%, and DCP were all measured in the same serum sample. The measurements of all three biomarkers were undertaken using a microchip capillary electrophoresis and liquid-phase binding assay on a μTASWako i30 auto analyzer (Wako Pure Chemical Industries Ltd.; ref. 34). Analytical sensitivity of μTASWako i30 is 0.3 ng/mL AFP and 0.1 ng/mL DCP, and the percentage of AFP-L3 can be measured when AFP-L3 is more than 0.3 ng/mL (34). All aspects of the test system performance have been reported (34). The assays were undertaken in a commercial laboratory with extensive experience of the μTASWako i30 auto analyzer; the operators had no knowledge of the diagnosis associated with the patient sample. There were no adverse events attributable to the biomarker tests.

Statistical methods

Model development. Continuous measurements are presented as medians (ranges) and categorical measurements are presented as frequencies. Odds ratios (ORs) are calculated using logistic regression for univariate and multivariate analyses to assess the strength of the association with hepatocellular carcinoma. Age, sex, albumin, bilirubin, AFP, DCP, and AFP-L3 were considered for inclusion in the multivariable models. Complete data for the GALAD score were available for more than 95% of cases. Patients with missing data were dropped from the statistical analysis. Factors such as symptoms and performance score were excluded to limit subjectivity in the model. A log transformation was made to AFP and DCP due to extreme skewness. Logistic regression analyses were based on a complete case analysis using a parsimonious forward–backward stepwise approach, keeping variables significant at the 1% level but that also increase AUROC. Fractional polynomials (35) were also used to investigate whether a more sophisticated transformation than the log transformation could improve the prediction.

Model accuracy is presented as sensitivity, specificity, proportion of false positives and negatives, and overall percentage of correct predictions. Having developed the model as described above, three cutoff points for classifying patients to the hepatocellular carcinoma group were used. The first optimized for maximum sensitivity while maintaining a prespecified specificity, the second for maximum specificity while maintaining a prespecified sensitivity, and the third for the maximum of the sum of specificity and sensitivity. Patients can then be classified by the model as being predicted to have hepatocellular carcinoma or not and this can then be directly compared against true diagnosis.

Model validation. Model validation is carried out on independent data sets in which again patients were diagnosed as hepatocellular carcinoma or CLD. The predictive score for each patient, based on fitted models, is used to classify patients as having hepatocellular carcinoma or not, and this is then directly compared against true diagnosis.

Results

Of the 331 patients with hepatocellular carcinoma, 283 (85.5%) had clear evidence of associated CLD, 37 (11.2%) seemed to have no underlying benign liver disease, and in 11 (3.3%) the presence or absence of underlying CLD could not be ascertained with certainty. The corresponding figures for the CLD group were 96%, 2.6%, and 1.8%, respectively. Among all data considered in the derived statistical models, data completeness was greater than 98%.

The median values for log(AFP), log(DCP), and AFP-L3 were significantly higher in the patients with hepatocellular carcinoma than in those with CLD, and both groups had median values higher than those for healthy control subjects (Fig. 1). All three biomarkers showed considerable discriminatory ability for distinguishing between hepatocellular carcinoma and CLD (AUROCs: log(AFP) 0.88, AFP-L3 0.84, and log(DCP) 0.90; Fig. 2A).

The optimal model (model 1), built on the discovery data set, included log(DCP), log(AFP), and AFP-L3, as well as age and sex, and had an AUC of 0.97. The Child–Pugh score was included in the logistic regression analysis but proved not to be a significant factor in the models. The model utility was maintained irrespective of the Child–Pugh class. Application of functional polynomials leads to a model incorporating AFP-L3, $DCP^{(-0.5)}$, and $AFP^{(0.5)}$ (model 2).

Table 2 shows the estimated coefficients (SE) and OR (95% confidence intervals, CI) from the univariate analyses as well as from the multivariate analysis using the discovery set data based on model 1. Table 3 shows the true positives/negatives, false positives/negatives, sensitivity, specificity, and proportion correctly classified when the multivariate model is used on the discovery data set (discovery) and on the validation data set (internal validation). For the validation set, AUROC = 0.98.
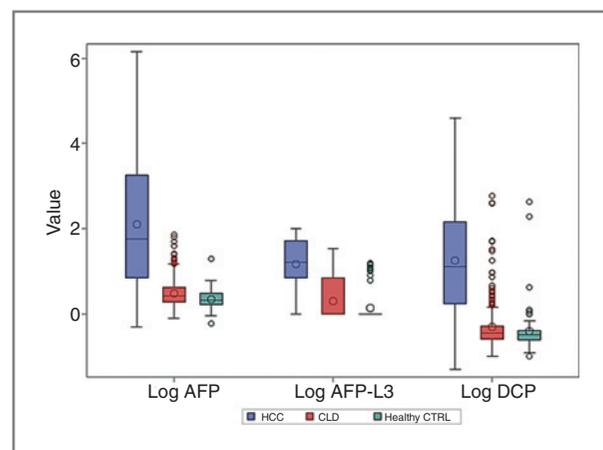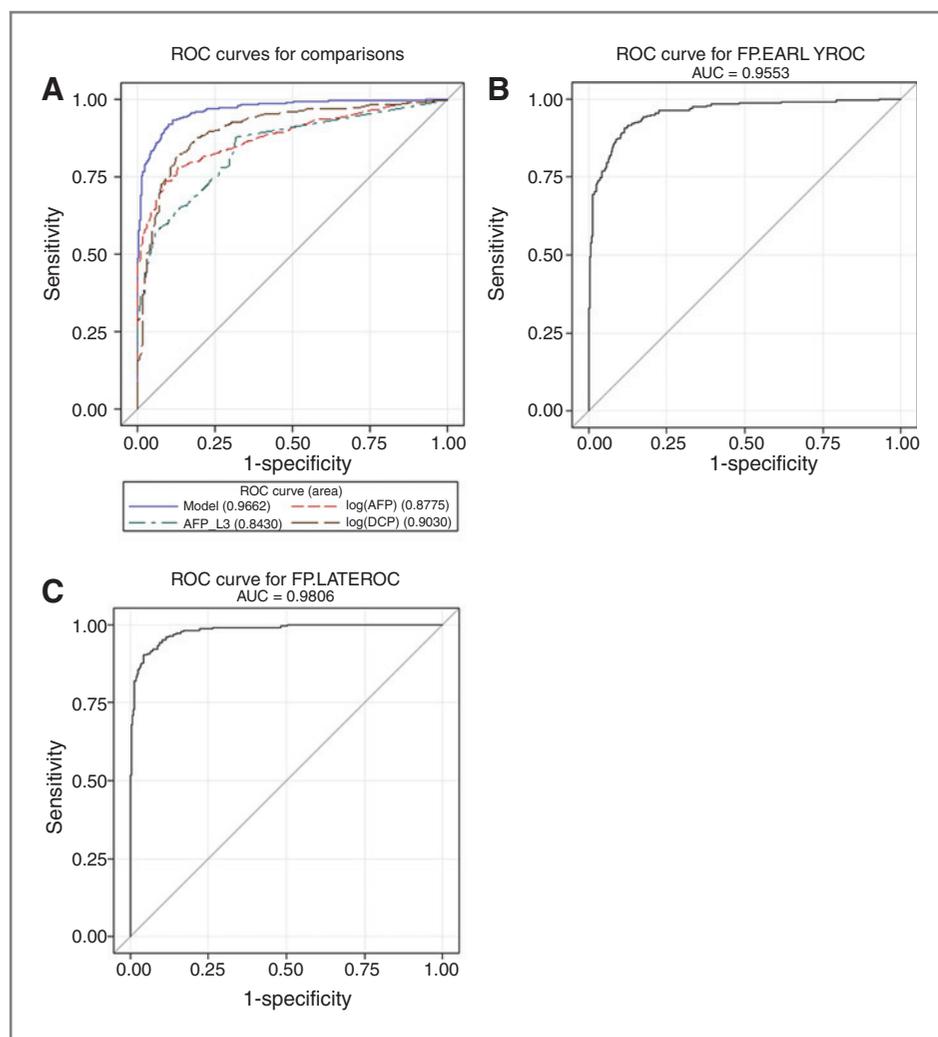

Figure 1. Log AFP, log AFP-L3, and log DCP values in the hepatocellular carcinoma, CLD, and healthy control patient cohorts, showing the median value and 25th and 75th percentiles. Outliers are shown as empty circles.

Figure 2. A–C, ROC curves showing the performance of model 3 (based on all Birmingham data) in comparison with individual biomarkers (A), patients with early disease as classified by TNM 6 (B), and in patients with late disease also classified by TNM 6 (C).

Also shown are subsets of the results for the fractional polynomial (model 2).

Because the validation results were extremely supportive of the model, the discovery and validation data sets (subsequently referred to as "Birmingham data") were combined and a new model was fitted (model 3).

For this model, AUROC = 0.97; parameter estimates (SE) and OR (95% CI) for the model variables are shown in Table 2. When the model was used on the Newcastle data set (external validation), AUROC = 0.95. A fractional polynomial model was also fitted (model 4).

Given the homogeneity of the Newcastle and Birmingham data, a final model based on all these data was found:

$$Z = -10.08 + 0.09 \times \text{age} + 1.67 \times \text{sex} + 2.34 \times \log(\text{AFP})$$
$$+ 0.04 \times \text{AFP-L3} + 1.33 \times \log(\text{DCP})$$

(model 5), where $\Pr(\text{HCC}) = \exp(Z)/(1 + \exp(Z))$ is the probability of hepatocellular carcinoma in a patient.

Table 3 shows the performance of models 1 and 2 on the Birmingham discovery and internal validation set, models 3 and 4 on all the Birmingham data and the external validation set, and model 5 on the total Birmingham and Newcastle data. Comparative figures are presented for models when maximized for either sensitivity and specificity or both in the discovery set, and then used in the validation sets. For example, using model 1, maximum sensitivity for specificity of 80% was achieved at a cutoff point of –1.58 and resulted in figures of 97% sensitivity in the discovery set and 98% sensitivity/74% specificity in the internal validation set. The corresponding cutoff point was –1.55 for the whole of the Birmingham data, leading to 97% sensitivity in the Birmingham data and 98% sensitivity/62% specificity in the external validation set. Note that the application of fractional polynomials did not improve the results.

The overall percentage of patients classified as having early disease varied widely depending on the definition applied (Table 4). Using BCLC, the percentage with early stage was 10%; with TNM 6 the percentage was 52%, 51% based on tumor size and 20% based on potentially curative therapies after multidisciplinary team review. Table 4

Table 2. Parameter estimates (SE) and OR (95% CI) of variables based on model 1

Variable	β (SE)	OR (95% CI)	χ^2	P
BRM discovery data set				
Constant	−10.25 (1.46)			
Age	0.10 (0.02)	1.11 (1.07–1.15)	29.63	<0.001
Sex	1.34 (0.46)	3.83 (1.56–9.42)	8.57	0.003
Log(AFP)	2.36 (0.48)	10.57 (4.16–26.86)	24.57	<0.001
AFP-L3	0.05 (0.02)	1.05 (1.00–1.09)	4.36	0.037
Log(DCP)	1.55 (0.24)	4.73 (2.93–7.63)	40.31	<0.001
Full BRM data set				
Constant	−10.32 (1.24)			
Age	0.10 (0.02)	1.11 (1.07–1.14)	39.97	<0.001
Sex	1.39 (0.38)	4.01 (1.89–8.49)	13.14	<0.001
Log(AFP)	2.43 (0.38)	11.37 (5.35–24.14)	40.04	<0.001
AFP-L3	0.04 (0.02)	1.04 (1.00–1.08)	4.66	0.031
Log(DCP)	1.45 (0.20)	4.28 (2.92–6.27)	55.33	<0.001
All data: BRM and NCL				
Constant	−10.08 (1.08)			
Age	0.09 (0.01)	1.10 (1.07–1.13)	44.87	<0.001
Sex	1.67 (0.33)	5.30 (2.79–10.07)	25.89	<0.001
Log(AFP)	2.34 (0.33)	10.34 (5.40–19.79)	49.73	<0.001
AFP-L3	0.04 (0.01)	1.04 (1.01–1.07)	8.66	0.003
Log(DCP)	1.33 (0.17)	3.77 (2.73–5.21)	64.56	<0.001

Abbreviations: BRM, Birmingham; NCL, Newcastle.

shows how model 5 performs on the early and late groups according to different classifications. Figure 2B and C show AUROC curves showing the performance of model 3 for the early and late groups (defined in terms of TNM 6) compared with controls.

Discussion

We have developed a model, now referred to as "GALAD," that generates a figure for the probability of an individual patient with CLD having hepatocellular carcinoma. The physician can then determine at what level of likelihood further investigation, in the form of CT or MRI scanning, should be instituted. The model performance was only slightly poorer for early compared with late hepatocellular carcinoma however these terms were defined.

AFP used on its own has been the most widely used biomarker for hepatocellular carcinoma but fluctuating, albeit low levels in patients with CLD (without hepatocellular carcinoma) and low sensitivity in patients with early hepatocellular carcinoma have resulted in the recommendation that it should not be used in the screening setting, although this view remains contentious (1, 2, 8, 20, 36–38). The development of the AFP-L3 assay has increased the sensitivity and specificity of AFP (39) because it retains significant discriminatory ability even at low levels of total AFP (13, 40–43). The recent development of highly sensitive AFP-L3 using an automated microfluidic based assay has made the isoform even more

sensitive and specific (13, 34). Because AFP, high-sensitivity AFP-L3, and DCP can now be measured on a single platform (44), the risk figure could be routinely reported. AFP-L3 and DCP have been approved by the Food and Drug Administration (FDA) in the United States and European Medicines Agency (EMEA) in Europe for the diagnosis of hepatocellular carcinoma. In Japan, all three markers are approved by the Japanese FDA (Korou-sho).

These markers have been investigated in Western populations using a study design similar to that applied here (45). Using receiver operating characteristic (ROC) analysis, Durazo and colleagues (45) determined optimal cutoff points for sensitivity, specificity, and positive predictive values, and concluded that DCP had optimal performance and that combining the markers did not achieve an additional predictive value to differentiate patients with hepatocellular carcinoma from those without hepatocellular carcinoma. In contrast, Carr and colleagues (46) considered that the combination of all three markers was superior to the individual although, unlike in the present study, formal statistical models were not proposed (47). The study most closely aligned to ours is that reported by Marrero and colleagues (47), in that a large number of Western patients were included, the disease etiologies were broadly similar, and account was taken of disease stage. Because overall the test performances will be affected by the methods of disease staging (discussed below), and Marrero and colleagues used a cut off point for individual markers rather than, as in our

Table 3. Model performance on the discovery and internal and external validation sets

Set		True HCC	True non-HCC	False HCC	False non-HCC	Sensitivity	Specificity	Correctly classified	Cutoff
Discovery	Max. sens. (spec. = 0.80)	200	198	49	7	97	80	88	−1.58
Model 1	Max. spec. (sens. = 0.80)	166	238	9	41	80	96	89	0.86
	Max. sens.+ spec.	192	226	21	15	93	91	92	−0.44
Internal validation	Max. sens. (spec. = 0.80)	110	68	24	2	98	74	87	−1.58
Model 1	Max. spec. (sens. = 0.80)	98	86	6	14	88	93	90	0.86
	Max. sens.+ spec.	106	79	13	6	95	86	91	−0.44
All Birmingham	Max. sens. (spec. = 0.80)	309	271	68	10	97	80	88	−1.55
Model 3	Max. spec. (sens. = 0.80)	256	329	10	63	80	97	89	1.10
	Max. sens.+ spec.	297	305	34	22	93	90	91	−0.48
External validation	Max. sens. (spec. = 0.80)	62	59	36	1	98	62	77	−1.55
Model 3	Max. spec. (sens. = 0.80)	51	90	5	12	81	95	89	1.10
	Max. sens.+ spec.	59	79	16	4	94	83	87	−0.48
All data: BRM and NCL	Max. sens. (spec. = 0.80)	367	347	87	15	96	80	88	−1.36
Model 5	Max. spec. (sens. = 0.80)	306	420	14	76	80	97	89	0.88
	Max. sens.+ spec.	356	385	49	26	93	89	91	−0.63
Fractional polynomial	Max. sens.+ spec. discovery	195	220	27	12	94	89	91	−0.64
Model 2	Max. sens.+ spec. internal validation	107	76	16	5	96	83	90	−0.64
Model 4	Max. sens.+ spec. All BRM	296	306	33	23	93	90	91	−0.40
Model 4	Max. sens.+ spec. external validation	58	80	15	5	92	84	87	−0.40

Abbreviations: BRM, Birmingham; HCC, hepatocellular carcinoma; NCL, Newcastle; Max. sens., maximum sensitivity; Max. spec., maximum specificity.

study, an overall model, direct comparison of results is not possible.

The three markers have also been studied in patients with cirrhosis or CLD who were followed-up and in whom hepatocellular carcinoma developed, although in neither study was biomarker validation the primary objective (45, 48). Two of the markers were investigated in patients in the Halt C trial using a nested case–control design. This involved 37 patients with hepatocellular carcinoma and 79 control subjects, all with HCV infection in the setting of a prospective study. The authors concluded that at the time of diagnosis, the combination provided a sensitivity of 91% and a specificity of 71% (at defined cutoff points; refs. 45, 48). In a second large prospective study, Kumada and colleagues (49) reported a series in which 623 patients with HCV-related CLD were prospectively followed-up and showed clearly that increased levels of AFP and AFP-L3 were closely associated with an increased incidence of hepatocellular carcinoma. On this basis, they suggested that patients with ≥10 ng/mL AFP levels or AFP-L3 ≥5% should receive intensive imaging at 3 to 6 month intervals. Our study included a third biomarker AFP-L3, a discovery, and two validation

sets. The external validation set was not ideal because the cohort was derived from patients with fatty liver disease (alcoholic and nonalcoholic) and, as such, had a different spectrum of etiologies to that on which the model was developed and validated. Ultimately, the model will require validation on larger external data sets with differing etiologies.

All our patients had a minimum follow-up of 6 months to exclude occult hepatocellular carcinoma in the CLD cohort. We did not focus entirely on patients with cirrhosis because the risk of hepatocellular carcinoma is associated with CLD rather than just at the stage of cirrhosis, and we aimed to set our study as close to a real clinical situation as possible (36).

Precise estimation of sensitivity and specificity demands a rigid "gold standard," and it is apparent that no definitive diagnosis of hepatocellular carcinoma or CLD *without* hepatocellular carcinoma is available. Thus, although current guidelines give figures of 85% and >95% for sensitivity and specificity, respectively, radiological diagnosis by CT or MRI imaging is still recognized to be "not infallible" (1, 2), especially in small and hypovascular tumors (50), Similarly, it is conceivable that a serological

Table 4. Performance of model 5 on early- and late-stage patients

Staging system/ treatment type	Criteria for early or late disease	Number of patients	Sensitivity		
			Maximum sensitivity Cutoff = −1.36	Maximum specificity Cutoff = 0.88	Maximum both Cutoff = −0.63
BCLC					
Early	0 and A	42	93	55	86
Late	B, C, and D	327	96	83	94
TNM 6					
Early	1 and 2	154	93	70	89
Late	3 and 4	143	99	90	97
Tumor size					
Early	≤5 cm	169	92	67	88
Late	>5 cm	166	99	92	98
Treatment intent					
Early	Curative	61	85	56	75
Late	Palliative	252	98	86	98

test might detect hepatocellular carcinoma in a cirrhotic liver before it is detectable on MRI scanning, thus resulting in the positive serological test being regarded as "false," and decreasing its apparent sensitivity. It would be surprising, therefore, if any new (in this case serologically based) test could achieve 100% sensitivity and specificity; if the three patients in the control group who developed hepatocellular carcinoma between 6 and 12 months into the study had been reclassified as being in the hepatocellular carcinoma group, the results of our study would have been even better.

The role of surveillance in the early detection of hepatocellular carcinoma is widely accepted, and it has been estimated that approximately 70% of patients who are detected when the lesion is <5 cm or has three tumors each less than 5 cm can receive potentially curative therapy (1, 2, 51). The optimal approach to surveillance remains contentious, some arguing that ultrasound alone should be the primary procedure, others arguing that ultrasound should be combined with AFP estimation (6, 52, 53).

Our study does not determine how such a model will perform in a prospective screening setting or in a routine practice outside specialist units, but we believe that these results are sufficiently encouraging to warrant a prospective study of this model to be run in parallel with conventional staging with USS with the aim of supplanting or, more likely, enhancing USS as an effective screening procedure for hepatocellular carcinoma.

References

1. Bruix J, Sherman M. Management of hepatocellular carcinoma. Hepatology 2005;42:1208–36.
2. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020–2.
3. Lee J, Yoon JH, Joo I, Woo H. Recent advances in CT and MR imaging for evaluation of hepatocellular carcinoma. Liver Cancer 2012;1: 22–40.
4. Silva MA, Hegab B, Hyde C, Guo B, Buckels JA, Mirza DF. Needle track seeding following biopsy of liver lesions in the diagnosis of hepatocellular cancer: a systematic review and meta-analysis. Gut 2008;57: 1592–6.
5. Stigliano R, Marelli L, Yu D, Davies N, Patch D, Burroughs AK. Seeding following percutaneous diagnostic and therapeutic approaches for hepatocellular carcinoma. What is the risk and the outcome? Seeding

risk for percutaneous approach of HCC. Cancer Treat Rev 2007;33: 437–47.

6. Bolondi L, Sofia S, Siringo S, Gaiani S, Casali A, Zironi G, et al. Surveillance programme of cirrhotic patients for early diagnosis and treatment of hepatocellular carcinoma: a cost effectiveness analysis. Gut 2001;48:251–9.

7. Johnson PJ. The role of serum alpha-fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. Clin Liver Dis 2001;5:145–59.

8. Marrero JA, El-Serag HB. Alpha-fetoprotein should be included in the hepatocellular carcinoma surveillance guidelines of the American Association for the Study of Liver Diseases. Hepatology 2011;53: 1060–1; author reply 1-2.

9. Trevisani F, D'Intino PE, Morselli-Labate AM, Mazzella G, Accogli E, Caraceni P, et al. Serum alpha-fetoprotein for diagnosis of hepato-cellular carcinoma in patients with chronic liver disease: Influence of HBsAg and anti-HCV status. J Hepatol 2001;34:570–5.

10. Liebman HA. Isolation and characterization of a hepatoma-associated abnormal (des-gamma-carboxy)prothrombin. Cancer Res 1989;49: 6493–7.

11. Ikoma J, Kaito M, Ishihara T, Nakagawa N, Kamei A, Fujita N, et al. Early diagnosis of hepatocellular carcinoma using a sensitive assay for serum des-gamma-carboxy prothrombin: a prospective study. Hepa-togastroenterology 2002;49:235–8.

12. Oda K, Ido A, Tamai T, Matsushita M, Kumagai K, Mawatari S, et al. Highly sensitive lens culinaris agglutinin-reactive alpha-fetoprotein is useful for early detection of hepatocellular carcinoma in patients with chronic liver disease. Oncol Rep 2011;26:1227–33.

13. Toyoda H, Kumada T, Tada T. Highly sensitive lens culinaris agglutinin-reactive alpha-fetoprotein: a new tool for the management of hepa-tocellular carcinoma. Oncology 2011;81Suppl 1:61–5.

14. Tang A, Cruite I, Sirlin CB. Toward a standardized system for hepa-tocellular carcinoma diagnosis using computed tomography and MRI. Expert Rev Gastroenterol Hepatol 2013;7:269–79.

15. Gambarin-Gelwan M, Wolf DC, Shapiro R, Schwartz ME, Min AD. Sensitivity of commonly available screening tests in detecting hepa-tocellular carcinoma in cirrhotic patients undergoing liver transplan-tation. Am J Gastroenterol 2000;95:1535–8.

16. Libbrecht L, Bielen D, Verslype C, Vanbeckevoort D, Pirenne J, Nevens F, et al. Focal lesions in cirrhotic explant livers: pathological evaluation and accuracy of pretransplantation imaging examinations. Liver Transpl 2002;8:749–61.

17. Teefey SA, Hildeboldt CC, Dehdashti F, Siegel BA, Peters MG, Heiken JP, et al. Detection of primary hepatic malignancy in liver transplant candidates: prospective comparison of CT, MR imaging, US, and PET. Radiology 2003;226:533–42.

18. Tong MJ, Blatt LM, Kao VW. Surveillance for hepatocellular carcinoma in patients with chronic viral hepatitis in the United States of America. J Gastroenterol Hepatol 2001;16:553–9.

19. Yao FY, Ferrell L, Bass NM, Watson JJ, Bacchetti P, Venook A, et al. Liver transplantation for hepatocellular carcinoma: expansion of the tumor size limits does not adversely impact survival. Hepatology 2001;33:1394–403.

20. Yu NC, Chaudhari V, Raman SS, Lassman C, Tong MJ, Busuttil RW, et al. CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis. Clin Gas-troenterol Hepatol 2011;9:161–7.

21. El-Serag HB, Kramer JR, Chen GJ, Duan Z, Richardson PA, Davila JA. Effectiveness of AFP and ultrasound tests on hepatocellular carcinoma mortality in HCV-infected patients in the USA. Gut 2011;60:992–7.

22. Singal A, Volk M, Waljee A, Salgia R, Higgins P, Rogers M, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. Aliment Pharmacol Ther 2009;30: 37–47.

23. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.

24. Alonzo TA. Standards for reporting prognostic tumor marker studies. J Clin Oncol 2005;23:9053–4.

25. Beale G, Chattopadhyay D, Gray J, Stewart S, Hudson M, Day C, et al. AFP, PIVKAII, GP3, SCCA-1 and follistatin as surveillance biomarkers for hepatocellular cancer in non-alcoholic and alcoholic fatty liver disease. BMC Cancer 2008;8:200.

26. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG, et al. AJCC cancer staging manual. New York:Springer; 2002.

27. Sobin LH, Wittekind C. TNM classification of malignant tumours. International Union Against Cancer 2002.

28. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. Lancet 2003;362:1907–17.

29. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19:329–38.

30. Bruix J, Llovet JM. Prognostic prediction and treatment strategy in hepatocellular carcinoma. Hepatology 2002;35:519–24.

31. 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: 693–700.

32. Child CG, Turcotte JG. Surgery and portal hypertension. Major Probl Clin Surg 1964;1:1–85.

33. Pugh RN, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 1973;60:646–9.

34. Kagebayashi C, Yamaguchi I, Akinaga A, Kitano H, Yokoyama K, Satomura M, et al. Automated immunoassay system for AFP-L3% using on-chip electrokinetic reaction and separation by affinity elec-trophoresis. Anal Biochem 2009;388:306–11.

35. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Applied Statistics 1994;43:429–67.

36. El-Serag HB, Davila JA. Surveillance for hepatocellular carcinoma: in whom and how? Therap Adv Gastroenterol 2011;4:5–10.

37. Lai Q, Melandro F, Pinheiro RS, Donfrancesco A, Fadel BA, Levi Sandri GB, et al. Alpha-fetoprotein and novel tumor biomarkers as predictors of hepatocellular carcinoma recurrence after surgery: a brilliant star raises again. Int J Hepatol 2012;2012:893103.

38. Sherman M. Serological surveillance for hepatocellular carcinoma: time to quit. J Hepatol 2010;52:614–5.

39. Leerapun A, Suravarapu SV, Bida JP, Clark RJ, Sanders EL, Mettler TA, et al. The utility of Lens culinaris agglutinin-reactive alpha-fetoprotein in the diagnosis of hepatocellular carcinoma: evaluation in a United States referral population. Clin Gastroenterol Hepatol 2007;5:394–402; quiz 267.

40. Toyoda H, Kumada T, Tada T, Kaneoka Y, Maeda A, Kanke F, et al. Clinical utility of highly sensitive Lens culinaris agglutinin-reactive alpha-fetoprotein in hepatocellular carcinoma patients with alpha-fetoprotein < 20 ng/mL. Cancer Sci 2011;102:1025–31.

41. Sterling RK, Jeffers L, Gordon F, Sherman M, Venook AP, Reddy KR, et al. Clinical utility of AFP-L3% measurement in North American patients with HCV-related cirrhosis. Am J Gastroenterol 2007;102: 2196–205.

42. Sterling RK, Jeffers L, Gordon F, Venook AP, Reddy KR, Satomura S, et al. Utility of Lens culinaris agglutinin-reactive fraction of alpha-fetoprotein and des-gamma-carboxy prothrombin, alone or in com-bination, as biomarkers for hepatocellular carcinoma. Clin Gastroen-terol Hepatol 2009;7:104–13.

43. Toyoda H, Kumada T, Osaki Y, Oka H, Kudo M. Role of tumor markers in assessment of tumor progression and prediction of outcomes in patients with hepatocellular carcinoma. Hepatol Res 2007;37 Suppl 2: S166–71.

44. Yamaguchi I, Nakamura K, Kitano H, Masuda Y, Kanke F, Kobatake S, et al. Development of des-gamma-carboxy prothrombin (DCP) mea-suring reagent using the LiBASys clinical analyzer. Clin Chem Lab Med 2008;46:411–6.

45. Durazo FA, Blatt LM, Corey WG, Lin JH, Han S, Saab S, et al. Des-gamma-carboxyprothrombin, alpha-fetoprotein and AFP-L3 in patients with chronic hepatitis, cirrhosis and hepatocellular carcinoma. J Gastroenterol Hepatol 2008;23:1541–8.

46. Carr BI, Kanke F, Wise M, Satomura S. Clinical evaluation of lens culinaris agglutinin-reactive alpha-fetoprotein and des-gamma-carboxy

prothrombin in histologically proven hepatocellular carcinoma in the United States. Dig Dis Sci 2007;52:776–82.

47. Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. Gastroenterology 2009;137:110–8.

48. Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. Gastroenterology 2010;138:493–502.

49. Kumada T, Toyoda H, Kiriyama S, Tanikawa M, Hisanaga Y, Kanamori A, et al. Predictive value of tumor markers for hepatocarcinogenesis in patients with hepatitis C virus. J Gastroenterol 2011;46:536–44.

50. Okuda H, Saito A, Shiratori K, Yamamoto M, Takasaki K, Nakano M. Clinicopathologic features of patients with primary malignant hepatic tumors seropositive for alpha-fetoprotein-L3 alone in comparison with other patients seropositive for alpha-fetoprotein-L3. J Gastroenterol Hepatol 2005;20:759–64.

51. Llovet JM. Updated treatment approach to hepatocellular carcinoma. J Gastroenterol 2005;40:225–35.

52. Lin OS, Keeffe EB, Sanders GD, Owens DK. Cost-effectiveness of screening for hepatocellular carcinoma in patients with cirrhosis due to chronic hepatitis C. Aliment Pharmacol Ther 2004;19:1159–72.

53. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. J Cancer Res Clin Oncol 2004;130: 417–22.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

The Detection of Hepatocellular Carcinoma Using a Prospectively Developed and Validated Model Based on Serological Biomarkers

Philip J. Johnson, Sarah J. Pirrie, Trevor F. Cox, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-13-0870

Cited articles	This article cites 51 articles, 5 of which you can access for free at: http://cebp.aacrjournals.org/content/23/1/144.full#ref-list-1
Citing articles	This article has been cited by 6 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/23/1/144.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/23/1/144. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.