INTRODUCTION

We have recently developed a serum-based tool ("GALAD") for detection of hepatocellular carcinoma (HCC) based on the objective measures of Gender, Age and three serological biomarkers, AFP, AFP-L3 and DCP[2], all of which are commercially available on a single standard platform[3]. The model has the potential to be used in the surveillance setting and may mitigate some of the limitations of ultrasound scanning (USS) including limited sensitivity in obese patients and those with advanced cirrhosis. The former is of particular importance as obesity-related HCC is accounting for an increasing percentage of HCC[4-8]. However, the model has not been validated in other countries where the underlying aetiology of HCC is different. Although it appeared to perform as well in patients with early disease (defined as tumour size < 5cm) as in advanced disease we did not undertake detailed analysis of the impact of tumour size on the utility of the model. This is of importance in the screening setting since the earlier the disease is detected the better the chance of curative treatment.

The same three biomarkers were combined with liver function tests (serum bilirubin and albumin) by Toyoda et al [9], to form the BALAD model for prognostication in HCC. A more rigorous statistical approach generated a second model (BALAD-2) which applied the same variables in a continuous rather than a categorical manner[10] but, again, the model has not been validated in the international setting or at different disease stages.

We now describe application of these two models to cohorts from Germany, Japan and Hong Kong.

We used cohorts from Germany, Japan and Hong Kong (Table 1). Both HCC and CLD cohorts were used for GALAD validation and HCC patients only for BALAD-2 validation. CLD refers to disease of the liver which has lasted over a period of six months. Table 1 also reports the percentage of cases with cirrhosis.

The German cohort came from four large centers based in The University Hospital Essen (collected between 2005 and 2008), Hannover Medical High School (collected between 2008 and 2014), Leipzig (Evangelisches Krankenhaus Duisburg-Nord, collected between 2010 and 2013) and Mainz University Medical Centre (collected between 2003 and 2012). Overall they comprised 1278 patients (275 HCC and 1003 patients with CLD alone).

The Japanese patients comprised 4476 patients (1514 with HCC and 2962 with CLD alone) and these were recruited from Ogaki Municipal Hospital where they were initially diagnosed as having HCC between 1988 and 2013.

The Hong Kong cohort (247 HCC patients) was recruited from the Prince of Wales Hospital, Department of Clinical Oncology, Chinese University of Hong Kong between 2009 and 2013[11].

For reference, the original UK cohort (on which the GALAD model was initially built and BALAD-2 validated) was included in the analysis. These were recruited at the Queen Elizabeth Hospital, Birmingham, UK and Newcastle Upon Tyne NHS Foundation Trust between 2007 and 2012. The Birmingham cohort comprised 670 patients (331 with HCC and 339 with CLD alone), and the Newcastle cohort, 163 patients (63 HCC and 100 CLD alone).

We also included 229 patients with other hepatobiliary tract cancers (cholangiocarcinoma and pancreatic adenocarcinoma) patients (Table 2) and 92 healthy controls (Table 1) recruited also from the Queen

Elizabeth Hospital Birmingham UK, between 2006 – 2012 and 2009 – 2011 respectively, to test the ability of the GALAD model to discriminate HCC from other hepatobiliary cancers and healthy controls. The hepatobiliary cohort was further divided into three subgroups: intrahepatic cholangiocarcinoma, extrahepatic cholangiocarcinoma and pancreatic adenocarcinoma.

None of the CLD-control group had evidence of HCC at the time the relevant serum sample was taken or within a minimum follow-up period of 9 months (Table 1). They were considered typical of those that would be included in a surveillance programme. In all cohorts, the HCC patients had the three biomarkers measured within ±1.7 months of HCC diagnosis and before any treatment was administered.

The diagnosis of HCC was made according to international guidelines.[4, 5] Patients in the control groups had established chronic liver disease (on the basis of liver biopsy and/or typical clinical and imaging features). All had the three constituent biomarkers measured using the μTASWako i30 auto analyzer (see below). Patients with HCC were classified as having early (those receiving potentially curative therapy), intermediate (intra-arterial therapies) or advanced disease (systemic chemotherapy or supportive care).

In Japan, the biomarkers were, in both the HCC and control groups, undertaken as part of a surveillance programme. In the other centres, they were collected specifically for this research study. The inclusion criterion was chronic liver disease (as defined above) with presence or absence of cirrhosis specified.

In total, 7155 patients (2430 HCC, 4725 CLD) were involved in this study.

Assays of AFP, AFP-L3, and DCP

AFP, AFP-L3 and DCP were all measured in the same serum sample. The measurements of hs-AFP-L3 and DCP were achieved by using a microchip capillary electrophoresis and liquid-phase binding assay on a μTASWako i30 auto analyzer (Wako Pure Chemical Industries, Ltd. Osaka, Japan)[3]. The Analytical

sensitivity of auto analyzer is 0.3ng/mL AFP, and the percentage of AFP-L3 can be measured when AFP-L3 is over 0.3ng/mL[3].

Statistical Methods

Stata IC 12 software was used to perform the analyses.

GALAD

The GALAD model, as described previously[2], uses the equation:

$Z=-10.08 + 0.09 \times age + 1.67 \times sex + 2.34 \log_{10}(AFP) + 0.04 \times AFP\text{-}L3 + 1.33 \times \log_{10}(DCP)$

where sex=1 for male and 0 for female.

The linear predictor (Z) is used to estimate the probability of HCC in an individual patient (ranging from 0 to 1) using the following equation:

$Pr(HCC) = \exp(Z)/(1+\exp(Z))$

To estimate sensitivity and specificity, three different sets of cut-off points were used in this study:

1. The three original cut-off points as specified in the original GALAD study[2] (that were based on the ***overall*** UK cohort).

2. Using the original UK cohort, three new cut-off points were also generated using a ***subset*** of patients that had early stage HCC (within Milan Criteria). As in the original study, one cut-off point maximizes sensitivity while keeping specificity at 80%, the second maximizes specificity while keeping sensitivity at 80% and the third maximizes the sum of sensitivity and specificity.

3. The same three optimized cut-off points, specific for Japan and Germany are also reported. For Japan, this was generated from a subgroup within Milan Criteria whereas in Germany, this was based on the overall cohort as the number of early stage disease patients was too small for meaningful analysis.

The effect of tumour size (maximum tumour diameter) recorded on the basis of an imaging procedure (CT or MRI scan) and aetiology on the performance of the GALAD model was tested. Unifocal tumour size ranges applied were <2, <3, <4, <5 and <10cm, as well as those within Milan Criteria.

For the purpose of analysis, aetiology was classified as hepatitis C virus (HCV) or hepatitis B virus (HBV)-related or 'other' (mainly alcoholic liver disease). Although detailed data on alcohol consumption was available in the Japanese cohort, a distinct diagnosis of 'alcoholic cirrhosis' was only collected in the UK and German cohorts. These were combined under "Europe" to test the performance of the model in an alcoholic cirrhosis cohort for comparison with the other aetiologies. The model performance was also tested in HCV patients depending on whether or not they had achieved sustained viral response (SVR) status or, in case of HBV, if they were on active anti-viral treatment. Detailed information was available only in the Japanese cohort. The number of cases receiving antiviral therapy in the German cohort was too small for a meaningful analysis.

Using the GALAD model, the area under the receiver operating characteristic (ROC) curves (AUROC), sensitivities and specificities were generated for each cohort as a whole and then as subgroups according to the tumour size and aetiology classifications as described above. The 95% confidence intervals (C.I.) for the AUROCs were constructed assuming a normal distribution for the area under the curve (asymptotic normal confidence intervals). The performance of the GALAD model was also compared to that of the individual biomarkers. The equality of ROC curves were tested using the method as described by Delong et al[12].

BALAD-2

Survival was measured from date of HCC diagnosis until date of death or date of last follow up. Patients undergoing transplantation (4.8% and 3.6% of the UK and German cohorts respectively, Table 1) were not excluded from the analysis.

The BALAD-2 model[10] (built on a Japanese and validated on a UK, cohort), uses the equation:

Linear predictor (xb) = 0.02 * (AFP-2.57) + 0.012* (AFP-L3-14.19) + 0.19 * (ln(DCP)-1.93) +

$$0.17 * ((\text{bili (umoll)}^{1/2})-4.50) - 0.09 * (\text{alb(gl)}-35.11)$$

where AFP was capped at 50000 units. Both AFP and DCP are modelled as per 1000 units.

To generate the four prognostic groups, cut-points applied to the linear predictor were xb > 0.24 (risk 4, high), 0.24 to >-0.91 (risk 3), -0.91 to > -1.74 (risk 2) and <= -1.74 (risk 1, low).

This equation was applied to all the HCC cohorts and separation of the prognostic groups assessed by Kaplan-Meier survival curves. The utility of the model was also tested by applying to subgroups according to treatment delivered.

The prognosis of HCC is widely assumed to be determined by underlying liver dysfunction and tumor-related factors. We hypothesized that BALAD-2 accurately reflected prognosis because it combined both of these, the former through (**B**)ilirubin and (**A**)lbumin and the latter through the three biomarkers (**LAD**) . To test this hypothesis we first assessed the prognostic impact of 'B' and 'A', by applying the 'ALBI' grade (a recently proposed and validated instrument for assessing liver function in HCC[13]). We then used the European cohort (merged UK and Germany) the test the extent to which adding the tumor markers i.e. the BALAD model would increase discriminatory utility of the ALBI grade using Harrell-C statistic[14, 15] and Akaike information criterion (AIC)[16]. Harrell-C assesses the discriminative ability of the model by measuring the proportion of patient pairs for which the model correctly assigns lower risk to the patient that truly survives longest (i.e. is at least risk). A good discriminative performance corresponds to a higher C-statistic. AIC measures relative fit between models for a given set of data. A 4-point reduction (per additional covariate) is indicative of an improved model.

Cases with missing data (within bilirubin, albumin, AFP, AFP-L3, DCP, age and gender) were excluded from the analysis; however these make up just 1.7% and 1% of the GALAD and BALAD-2 data respectively.

RESULTS

GALAD

Demographic and clinical details of the cohorts are shown in Tables 1.

The model gave an overall AUROC figure of 0.93 (95% C.I. 0.92 – 0.94), and 0.94 (95% C.I. 0.93 – 0.96) in the Japanese and German validation cohorts respectively (Supplementary figure 1a), only marginally lower than the figure for the original UK cohort (0.97, 95% C.I. 0.96 – 0.98). GALAD also correctly classified HCC from other hepatobiliary cancers and healthy controls within the UK cohort showing AUROC figures of 0.95 (95% C.I. 0.93 – 0.96) and 0.97 (95% C.I. 0.96 – 0.99) respectively (Supplementary figure 1b and 1c). The ROC curve results of hepatobiliary cancer subgroups (pancreatic adenocarcinoma, extrahepatic cholangiocarcinoma and intrahepatic cholangiocarcinoma) are shown in Supplementary figure 2a-c. The AUROC derived from the model was superior ($p<0.0001$) to that obtained if the biomarkers were used individually (Figure 1a-c). This was true both in the cohort overall and within the subset of early stage HCC patients (within Milan Criteria) (Figure 1d-e). Table 2 shows the figures for sensitivity, specificity and AUROC for all patients as well as those within Milan Criteria. Cut-offs for the GALAD model in Table 2 maximise the sum of sensitivities and specificities for each cohort.

The utility of the model was slightly lower in the smaller unifocal tumours, but remained in range of 0.85 to 0.95 (Figure 2a-c). The AUROCs of patients with less than 2cm unifocal tumours were 0.92 (95% C.I. 0.85 – 0.997), 0.89 (95% C.I. 0.88 – 0.91) and 0.93 (95% C.I. 0.89 – 0.97) for UK, Japan and Germany, however in the latter group the numbers were very small. Unifocal <3cm in the German series generated AUROC of 0.87 (95% C.I. 0.81 – 0.94). In those patients that were within Milan Criteria, the corresponding AUROC figures were 0.93 (95% C.I. 0.90 – 0.96) and 0.91 (95% C.I. 0.90 – 0.92) in the UK and Japan cohorts respectively (Figure 2d-e).

There was no statistically significant difference ($p>0.05$) in model performance between HBV, HCV and other subgroups, in the European cohorts. Although there was a statistically significance difference between the aetiologies in the Japanese cohort ($p=0.0012$), this was unlikely to be of clinical significance as the

figures ranged only between 0.92-0.95 (Supplementary figure 3a – c). Due to lower numbers of HCV and HBV subgroups within the UK and German cohorts, these were combined under "Europe" and the alcoholic aetiology subgroup was added prior to generating GALAD AUROC curves (Supplementary figure 3d). This showed that the GALAD model performed equally well (p=0.7490) in all four 'aetiological' subgroups.

In the Japanese series, performance of the GALAD model was not affected (p>0.1) by the status of SVR or active viral treatment in HCV and HBV patients respectively (Supplementary figure 4a-b).

The AUROC values as well as the sensitivity, specificity and correctly classified percentages at the different cut-offs (whole cohort and subgroups) are summarized in Table 3 (UK-based cut-offs) and Supplementary Table 1 (country specific cut-offs). The same data for the different etiologies in the combined European cohort (UK and Germany) are shown in supplementary table 2.

BALAD-2

Applying the BALAD-2 model to the German and Hong Kong cohorts produced four clearly distinct and well separated HCC prognostic groups, hence confirming the utility of the model in the international setting (Figure 3a-d, median survival at each BALAD-2 score in Figure 3e). This separation remained when the analysis was confined to patients with a minimum of 18 months follow up (Supplementary figure 5a-e).

BALAD-2 was then compared to the ALBI grade (Supplementary figure 6a-c). The AIC and Harrell-C statistics clearly show that BALAD-2 model was a better fit to the data than the ALBI model, as demonstrated by the lower AIC and higher Harrell-C scores (5233.982 and 0.7012 for the BALAD-2, and 5469.059 and 0.6192 for the ALBI grade respectively).

The BALAD-2 model proved equally discriminatory in all treatment classes (Supplementary figure 7a-g). UK and German cohorts were merged (as "Europe") as sample size for each treatment subgroup was relatively small.

DISCUSSION

Our analysis validates the GALAD model by showing that it has utility outside the country in which it was developed (the UK). Both the German and Japanese datasets were particularly valuable in the setting of validation. Each of the study cohorts was multi-center, the relevant biomarkers were measured in an external laboratory by an operator who had no knowledge of the clinical diagnosis and the statistical analysis was undertaken at a third center where no further manipulation of the data was undertaken. In each of the four German centers, the AUROC was virtually identical and very close to that reported in the UK series. This was not surprising since the aetiology and clinical features of HCC were very similar between Germany and the UK.

In Japan these biomarkers are currently used individually or, more often, in concert to enhance routine ultrasound screening[17, 18], based on clinical experience rather than a formal statistical model. Combination of the markers is increasingly recognized to add utility to the individual biomarkers[19] and here we show that a formal, prospectively developed, statistical model that combines the markers is, superior to individual markers alone. Further, when the model performance was directly compared with results as obtained by the conventional combined use of the three markers in clinical practice in Japan[1], there was a clear improvement. Presumably this reflects the gain in information derived from the individual markers by treating them as continuous variables. Several recent studies have confirmed significant information loss when cut-off points are applied to continuous variables[20, 21].

The current analysis validates our original report where 'small tumors' were considered to be those with a maximum tumour size of 5cm but significantly extends the utility of the model by showing that, when unifocal tumours are considered, model performance remains remarkably good (AUROC >0.85 for Germany and >0.89 for other cohorts) and consistent down to <2 cm. This is perhaps not surprising since Marrero et al[22] reported that the overall performance of the individual biomarkers only decreased marginally in 'early tumors'. The utility of the GALAD model in early stage disease has also been reported in patients with early stage disease as assessed according to the BCLC staging system[23] with similar results - (BCLA 0, 0.97, BCLC A 0.98 and BCLC B 0.97)(Caviglia CP & Smedile 2015; personal communication).

The limitations of USS are acknowledged in the AASLD guidelines[5] in that performance characteristics have not been well-defined in nodular cirrhotic livers and that 'some patients, particularly the obese, are not good candidates [for surveillance] despite their risk'[5]. The serological approach has the specific advantage that it is not impacted upon by physical factors such as obesity, an increasingly recognized etiological factor for HCC[7, 8, 24]. Our data suggest that the GALAD model is likely to detect tumours within the range that potentially curative therapies will be applicable. A further limitation of USS is that other primary liver tumours may be detected. In a recent analysis of surveillance in Japan (1994 – 2005), 4.4% of detected hepatic tumours were ultimately classified as cholangiocarcinoma[25]. Our data suggests that the GALAD model can successfully discriminate between HCC and cholangiocarcinoma (Supplementary figure 1b and supplementary figure 2).

Having validated the GALAD model, it now requires testing in a prospective manner and we are aware that the three tumour markers involved in GALAD are currently being prospectively assessed in clinical trials in North America. From these it will be possible to assess the potential role of the GALAD model in the clinical surveillance setting.

Our analysis supports the clinically plausible view that the prognostic power of the BALAD-2 model is based on its ability to reflect both the degree of underlying liver dysfunction ('B' and 'A') and tumour related factors ('LAD'). The general applicability of the model is underlined by the observation that discrimination is equally good, irrespective to the treatment applied.

Figure 1a-e. Receiver operating characteristic (ROC) curves comparing overall performance of the GALAD model to the individual biomarkers in the (a) UK, (b)Japan, (c) German cohorts as well as in early stage HCC (within Milan Criteria) in the (d) UK and (e) Japan cohorts.

Figure 2a-e. Receiver operating characteristic (ROC) curves comparing the performance of the GALAD model in subgroups of patients with unifocal tumours of different sizes in (a) UK, (b) Japan and (c) Germany, in addition to those within Milan Criteria in (d) UK and (e) Japan cohorts.

Figure 3a-e. Survival according to BALAD-2 score in the (a) UK, (b) Japan (c) German and (d) Hong Kong datasets. Table in (e) shows median survival in each cohort according to BALAD-2 score.

1. Toyoda H, Kumada T, Kagebayashi C, et al. Diagnosis Of Hepatocellular Carcinoma Using A GALAD Model By Objective Clinical And Serological Factors. Hepatology 2013;58:1230A.
2. Johnson PJ, Pirrie SJ, Cox TF, et al. The detection of hepatocellular carcinoma using a prospectively developed and validated model based on serological biomarkers. Cancer Epidemiology Biomarkers & Prevention 2014;23:144-153.
3. Kagebayashi C, Yamaguchi I, Akinaga A, et al. Automated immunoassay system for AFP–L3% using on-chip electrokinetic reaction and separation by affinity electrophoresis. Analytical biochemistry 2009;388:306-311.
4. Bruix J, Sherman M. Management of hepatocellular carcinoma. Hepatology 2005;42:1208-36.
5. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020-1022.
6. Bolondi L, Sofia S, Siringo S, 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. Dyson J, Jaques B, Chattopadyhay D, et al. Hepatocellular cancer: the impact of obesity, type 2 diabetes and a multidisciplinary team. J Hepatol 2014;60:110-7.
8. Siegel AB, Zhu AX. Metabolic Syndrome and Hepatocellular Carcinoma: Two Growing Epidemics with a Potential Link. Cancer 2009;115:5651-5661.
9. Toyoda H, Kumada T, Osaki Y, et al. Staging hepatocellular carcinoma by a novel scoring system (BALAD score) based on serum markers. Clinical Gastroenterology and Hepatology 2006;4:1528-1536.
10. Fox R, Berhane S, Teng M, et al. Biomarker-based prognosis in hepatocellular carcinoma: validation and extension of the BALAD model. British journal of cancer 2014;110:2090-2098.
11. Chan SL, Mo F, Johnson P, et al. Applicability of BALAD score in prognostication of hepatitis B-related hepatocellular carcinoma. Journal of gastroenterology and hepatology 2015.
12. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988:837-845.

13. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol 2015;33:550-8.

14. Newson RB. Comparing the predictive powers of survival models using Harrell's C or Somers' D. Stata Journal 2010;10:339.

15. Taktak A, Eleuteri A, Lake S, et al. Evaluation of prognostic models: Discrimination and calibration performance. Computational Intelligence in Medicine 2007:2003-2005.

16. Akaike H. Information theory as an extension of the maximum likelihood principle. sin BN Petrov, and F, In Second International Symposium on Information Theory, 1973.

17. TheJapanSocietyofHepatology. Chapter 2: Diagnosis and surveillance. Hepatology Research 2010;40:16-47.

18. Kudo M, Matsui O, Izumi N, et al. JSH Consensus-Based Clinical Practice Guidelines for the Management of Hepatocellular Carcinoma: 2014 Update by the Liver Cancer Study Group of Japan. Liver Cancer 2014;3:458-468.

19. Ertle JM, Heider D, Wichert M, et al. A combination of α-fetoprotein and des-γ-carboxy prothrombin is superior in detection of hepatocellular carcinoma. Digestion 2013;87:121-131.

20. Del Priore G, Zandieh P, Lee M-J. Treatment of continuous data as categoric variables in obstetrics and gynecology. Obstetrics & Gynecology 1997;89:351-354.

21. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in medicine 2006;25:127-141.

22. Marrero JA, Feng Z, Wang Y, et al. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. Gastroenterology 2009;137:110-8.

23. Caviglia GP, Abate ML, Petrini E, et al. Highly sensitive alpha-fetoprotein, Lens culinaris agglutinin-reactive fraction of alpha-fetoprotein and des-gamma-carboxyprothrombin for hepatocellular carcinoma detection. Hepatol Res 2015;17:12544.

24. Singal A, Volk ML, Waljee A, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. Aliment Pharmacol Ther 2009;30:37-47.

25. Ikai I, Kudo M, Arii S, et al. Report of the 18th follow-up survey of primary liver cancer in Japan. Hepatology Research 2010;40:1043-1059.

Table 1. Patients characteristics

	HCC, CLD and healthy controls patients								cholangiocarcinoma/pancreatic adenocarcinoma patients			
	UK (Birmingham and Newcastle)		**Japan**		**Germany (Hanover, Leipzig, Essen and Mainz)**		**Hong Kong**	**Healthy controls**	**Pancreatic adenocarcinoma**	**Extrahepatic cholangiocarcinoma**	**Intrahepatic cholangiocarcinoma**	**Overall**
Variable	HCC (n=394)	CLD (n=439)	HCC (n=1514)	CLD (n=2962)	HCC (n=275)	CLD (n=1003)	HCC (n=247)	n=92	n=173	n=39	n=17	n=229
Demographics												
Median Age (IQR)	66.9 (59.6 – 73.5)	56.1 (46 – 64)	69 (62 – 75)	63 (53 – 71)	66 (60 – 72)	52 (41 – 61)	58 (52 – 68)	66 (57 – 74)	66 (59 – 72)	64 (56 – 69)	67 (63 – 71)	66 (59 – 71)
Mean Age (±SD)	65.8 (±9.7)	54.9 (±13.7)	67.8 (±9.4)	61.0 (±13.7)	65.4 (±9.2)	50.5 (±14.1)	60.2 (±11.4)	65.4 (±11.0)	64.4 (±12.5)	63.6 (±9.5)	65.0 (±10.7)	64.3 (±11.9)
Gender (% Male)	82.5	58.3	71.3	48.0	84.0	55.4 (n=903)	88.7	43.5	53.2	43.6	41.2	50.7
Aetiology												
HCV : HBV : Other (%)	18.3 : 9.0 : 72.7 n=377	24.1 : 14.3 : 61.7 n=428	69.2 : 15.4 : 15.4 n=1495	45.4 : 24.1 : 30.5 n=2920	20.9 : 12.2 : 66.9 n=263	30.3 : 36.1 : 33.7 n=835	8.1 : 80.1 : 11.8 n=246	NA	N/A	N/A	N/A	N/A
SVR achieved (HCV) (%)	NA	NA	1.1 n=1514	11.6 n=2942	8 n=25	47.1 n=34	NA	NA	N/A	N/A	N/A	N/A
On HBV treatment (%)	NA	NA	2.1 n=1514	5.3 n=2962	NA	NA	NA	NA	N/A	N/A	N/A	N/A
Alcohol (%)	40.2 (n=386)	25.6 (n=437)	NA	NA	43.6 (n=266)	19.6 (n=846)	NA	NA	N/A	N/A	N/A	N/A
Cirrhotic (%) in HCV group	93.7 (n=63)	100 (n=41)	NA	NA	93.8 (n=48)	50.4 (n=234)	75.0 (n=20)	0	N/A	N/A	N/A	N/A
Cirrhotic (%) in HBV group	59.3 (n=27)	100 (n=14)	NA	NA	92.6 (n=27)	30.3 (n=178)	48.2 (n=197)	0	N/A	N/A	N/A	N/A
HCC Biomarkers												
AFP, ng/ml	53.1 (7.6 – 1460.9) n=394	2.9 (2.1 – 4.7) n=438	22.3 (7.0 – 171.6) n=1514	2.5 (1.8 – 3.9) n=2962	42.0 (7.6 – 801.3) n=275	3 (1.9 – 5.5) n=1003	141.2 (13.2 – 9930.8) n=247	2.1 (1.7 – 3) n=92	1.9 (1.3 – 3.1) n=172	2.3 (1.6 – 3) n=39	3 (2.4 – 3.9) n=17	2.1 (1.3 – 3.1) n=228
L3, %	17 (7.2 – 51.8) n=382	1 (1 – 7.2) n=438	4.9 (0.5 – 16.8) n=1514	0.5 (0.5 – 0.5) n=2962	14.6 (6.1 – 47) n=275	0.1 (0.1 – 5.8) n=1003	20.6 (5.6 – 56) n=247	1 (1 – 1) n=92	1 (1 – 1) n=171	1 (1 – 1) n=39	1 (1 – 10.7) n=17	1 (1 – 1) n=227
DCP, ng/ml	20.1 (2.6 – 169.6) n=383	0.4 (0.3 – 0.7) n=438	0.7 (0.2 – 9.5) n=1514	0.20 (0.1 – 0.2) n=2962	10.6 (1.5 – 152.7) n=275	0.4 (0.2 – 0.6) n=1003	36.1 (1.7 – 357.1) n=247	0.3 (0.3 – 0.4) n=92	0.8 (0.4 – 4.2) n=172	0.5 (0.2 – 3.5) n=39	0.3 (0.2 – 1.1) n=17	0.7 (0.4 – 3.8) n=228
Liver Function Tests												
Albumin, g/L	38 (34 – 42) n=393	44 (40 – 46) n=439	36 (31 – 40) n=1514	43 (41 – 45) n=2962	37.7 (33 – 43.3) n=273	42 (36.2 – 46) n=697	38 (34 – 41) n=247	NA	N/A	N/A	N/A	N/A
INR	1.1 (1 – 1.2) n=387	1 (1 – 1.1) n=333	NA	NA	1.1 (1.0 – 1.2) n=87	1 (1 – 1.1) n=414	1.1 (1.0 – 1.2) n=247	NA	N/A	N/A	N/A	N/A
Bilirubin, µmol/L	17 (11 – 28) n=393	11 (7 – 18) n=439	13.7 (10.3 – 22.2) n=1514	12.0 (8.6 – 15.4) n=2962	15.7 (8.6 – 29.8) n=265	9.9 (5.6 – 15.4) n=696	19 (12 – 32) n=247	NA	N/A	N/A	N/A	N/A
Tumour Characteristics												
% Solitary	46.8 n=378	N/A	55.0 n=1512	NA	32.6 n=267	NA	42.5 n=247	NA	N/A	N/A	N/A	N/A
Maximum tumour size (cm), %												
< 5cm : ≥ 5cm	48.1 : 51.9 n=337	N/A	76.7 : 23.3 n=1489	NA	48.1 : 51.9 n=235	NA	29.6 : 70.4 n=243	NA	N/A	N/A	N/A	N/A
<3cm : 3 – 5cm : >5cm	21.1 : 33.2 : 45.7 n= 337	N/A	56.4 : 21.2 : 22.4 n=1489	NA	22.6 : 31.5 : 46 0 n=235	NA	14.4 : 17.3 : 68.3 n=243	NA	N/A	N/A	N/A	N/A
Treatments (HCC)												
N	n=313	NA	n=1513	NA	n=168	NA	n=247	NA	N/A	N/A	N/A	N/A
Transplantation (%)	4.8		0		3.6		0					
Resection (%)	2.6		32.9		4.2		14.6					
Ablative (%)	10.5		26.9		3.0		8.1					
TACE (%)	38.0		21.0		26.2		13.4					
Sorafenib/Chemotherapy (%)	16.0		1.1		14.3		26.3					
Supportive (%)	22.7		13.5		14.3		37.7					
Other palliative (%)	5.4		4.6		34.5		0					
Survival												
Median follow up time (months)	31.1 (28.3 – 39.8) n=387	24.6 (23.0 – 29.0) n=422	68.4 (60.8 – 74.0) n=1514	56.4 (n=2962)	47.9 (43.2 – 69.9) n=264	31.4 (29.7 – 33.4) n=295	37.7 (35.3 – 43.1) n=246	NA	N/A	N/A	N/A	N/A
Overall Survival (months)	12.3 (9.8 – 14.9) n=387	Not reached n=422	45.2 (40.5 – 50.1) n=1514	Not reached n=2962	12.7 (8.5 – 16.4) n=264	Not reached n=295	10.8 (5.9 – 15.0) n=246	NA	N/A	N/A	N/A	N/A

Abbreviations: AFP, alpha-fetoprotein; CLD, chronic liver disease; DCP, Des-gamma carboxyprothrombin; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; INR, international normalised ratio; N/A, not applicable; NA, not available; SD, standard deviation. All continuous variables are presented as median (with interquartile range).

Table 2: Comparison between GALAD model and the individual HCC biomarkers

		All					Early HCC (within Milan Criteria)				
Model/biomarker	**Cut-off**	**AUC**	**p-value (GALAD vs biomarker)**	**Sensitivity %**	**Specificity %**	**Correctly Classified %**	**AUC**	**p-value (GALAD vs biomarker)**	**Sensitivity %**	**Specificity %**	**Correctly Classified %**
UK											
GALAD model	-0.63	0.97 (0.96 – 0.98)	-	91.6	89.7	90.6	0.93 (0.90 – 0.96)	-	80.2	89.7	87.9
AFP	20 ng/mL*	0.88 (0.85 – 0.90)	<0.0001	60.7	96.4	79.5	0.84 (0.79 – 0.89)	<0.0001	49.1	96.4	86.9
AFP-L3	7%*	0.84 (0.82 – 0.87)	<0.0001	75.4	73.5	74.4	0.81 (0.76 – 0.85)	<0.0001	71.7	73.5	73.2
DCP	0.48 ng/mL*	0.90 (0.88 – 0.93)	<0.0001	62.4	93.8	79.2	0.81 (0.77 – 0.86)	<0.0001	86.8	63.7	68.2
AFP + AFP-L3 +DCP**	Same as above	0.75 (0.72 – 0.77)	<0.0001	99.2	50.0	72.9	0.75 (0.72 – 0.77)	<0.0001	99.1	50.0	59.6
Japan											
GALAD model	-1.95†	0.93 (0.92 – 0.94)	-	81.4	89.1	86.5	0.91 (0.90 – 0.92)	-	82.1	81.6	81.7
AFP	20 ng/mL*	0.89 (0.88 – 0.90)	<0.0001	51.3	97.3	81.8	0.87 (0.86 – 0.89)	<0.0001	42.2	97.3	84.6
AFP-L3	7%*	0.75 (0.74 – 0.77)	<0.0001	41.2	91.8	74.7	0.71 (0.70 – 0.73)	<0.0001	30.0	91.8	77.5
DCP	0.48 ng/mL*	0.84 (0.83 – 0.85)	<0.0001	57.3	97.4	83.8	0.78 (0.76 – 0.80)	<0.0001	41.4	97.4	84.5
AFP + AFP-L3 +DCP**	Same as above	0.84 (0.83 – 0.85)	<0.0001	79.3	88.3	85.3	0.80 (0.78 – 0.81)	<0.0001	71.2	88.3	84.3
Germany											
GALAD model	-0.68†	0.94 (0.93 – 0.96)	-	88.4	88.2	88.3	NA	NA	NA	NA	NA
AFP	20 ng/mL*	0.87 (0.85 – 0.89)	<0.0001	56.7	93.9	85.9	NA	NA	NA	NA	NA
AFP-L3	7%*	0.83 (0.80 – 0.86)	<0.0001	71.3	79.7	77.9	NA	NA	NA	NA	NA
DCP	0.48 ng/mL*	0.86 (0.83 – 0.89)	<0.0001	89.1	64.2	69.6	NA	NA	NA	NA	NA
AFP + AFP-L3 +DCP**	Same as above	0.73 (0.71 – 0.75)	<0.0001	95.3	54.3	63.2	NA	NA	NA	NA	NA

* Cut-off points for three biomarkers were based on the guideline of the Japan Society of Hepatology.
For the GALAD model, the optimum cut-off point was set from the ROC analysis.

**The combination (AFP+AFP-L3+DCP) represents the current method of using the markers in Japan. A positive result is recorded if any of the markers exceed their specified cut off point.

† Locally-based cut-off that maximises both sensitivity and specificity for GALAD in each cohort. Japan cut-offs generated from within Milan subgroup. UK and Germany based on whole cohort.

Table 3. GALAD model performance

			Original cut-offs									Cut-offs based on UK patients within Milan Criteria								
			Max sens: -1.36			Max spec: 0.88			Max sens and spec : -0.63			Max sens:-1.54			Max spec: -0.59			Max sens and spec : -0.77		
Country	Set	AUC	Sens. %	Spec. %	Corr. Class. %	Sens. %	Spec. %	Corr. Class. %	Sens. %	Spec. %	Corr. Class. %	Sens. %	Spec. %	Corr. Class. %	Sens. %	Spec. %	Corr. Class. %	Sens. %	Spec. %	Corr. Class. %
UK	**Whole cohort** (382 HCC, 437 CLD)	0.97 (C.I. 0.96 – 0.98)	94.8	83.1	88.5	77.2	97.5	88.0	91.6	89.7	90.6	96.1	80.1	87.6	91.6	90.2	90.8	92.7	88.8	90.6
	Unifocal <2cm (9 HCC, 437 CLD)	0.92 (0.85 – 0.997)	77.8	83.1	83.0	55.6	97.5	96.6	77.8	89.7	89.5	77.8	80.1	80.0	77.8	90.2	89.9	77.8	88.8	88.6
	Unifocal <3cm (29 HCC, 437 CLD)	0.92 (0.88 – 0.96)	72.4	83.1	82.4	58.6	97.5	95.1	72.4	89.7	88.6	82.8	80.1	80.3	72.4	90.2	89.1	72.4	88.8	87.8
	Unifocal <4cm (52 HCC, 437 CLD)	0.93 (0.90 – 0.96)	80.8	83.1	82.8	51.9	97.5	92.6	78.9	89.7	88.6	88.5	80.1	81.0	78.9	90.2	89.0	78.9	88.8	87.7
	Unifocal <5cm (75 HCC, 437 CLD)	0.93 (0.90 – 0.96)	84.0	83.1	83.2	57.3	97.5	91.6	82.7	89.7	88.7	89.3	80.1	81.5	82.7	90.2	89.1	82.7	88.8	87.9
	Unifocal <10cm (127 HCC, 437 CLD)	0.95 (0.93 – 0.97)	89.0	83.1	84.4	65.4	97.5	90.3	87.4	89.7	89.2	92.9	80.1	83.0	87.4	90.2	89.5	87.4	88.8	88.5
	Within Milan criteria (106 HCC, 437 CLD)	0.93 (0.90 – 0.96)	85.9	83.1	83.6	54.7	97.5	89.1	80.2	89.7	87.9	90.6	80.1	82.1	80.2	90.2	88.2	83.0	88.8	87.7
	HCV positive cohort (67 HCC, 103 CLD)	0.98 (C.I. 0.97 – 1.0)	95.5	87.4	90.6	77.6	99.0	90.6	91.0	94.2	92.9	97.0	85.4	90.0	91.0	95.2	93.5	94.0	94.2	94.1
	HBV positive cohort (33 HCC, 61 CLD)	0.99 (C.I. 0.96 – 1.00)	93.9	100.0	97.9	69.7	100.0	89.4	87.9	100.0	95.7	93.9	98.4	96.8	87.9	100.0	95.7	87.9	100.0	95.7
	Other aetiology (267 HCC, 262 CLD)	0.96 (C.I. 0.95 – 0.98)	95.5	77.5	86.6	78.7	96.2	87.3	92.5	85.5	89.0	96.3	73.7	85.1	92.5	86.3	89.4	93.3	84.4	88.9
Japan	**Whole cohort** (1514 HCC, 2962 CLD)	0.93 (C.I. 0.92 – 0.94)	79.9	90.0	86.6	50.1	99.1	82.5	70.5	95.8	87.2	81.9	87.7	85.8	69.6	96.1	87.1	72.5	94.9	87.3
	Unifocal <2cm (329 HCC, 2962 CLD)	0.89 (0.88 – 0.91)	68.7	90.0	87.9	27.7	99.1	91.9	54.4	95.8	91.7	71.1	87.7	86.1	53.2	96.0	91.7	56.5	94.9	91.0
	Unifocal <3cm (554 HCC, 2962 CLD)	0.90 (0.88 – 0.91)	70.8	90.0	87.0	31.8	99.1	88.5	58.1	95.8	89.9	72.7	87.7	85.4	56.7	96.0	89.8	60.3	94.9	89.4
	Unifocal <4cm (672 HCC, 2962 CLD)	0.90 (0.89 – 0.92)	72.5	90.0	86.8	33.6	99.1	87.0	59.8	95.8	89.2	74.6	87.7	85.3	58.2	96.0	89.0	62.2	94.9	88.8
	Unifocal <5cm (732 HCC, 2962 CLD)	0.91 (0.90 – 0.92)	73.1	90.0	86.7	35.1	99.1	86.4	60.4	95.8	88.8	75.1	87.7	85.3	58.9	96.0	88.6	62.6	94.9	88.5
	Unifocal <10cm (810 HCC, 2962 CLD)	0.91 (0.90 – 0.92)	74.2	90.0	86.6	38.0	99.1	86.0	62.6	95.8	88.7	76.2	87.7	85.3	61.2	96.0	88.5	64.6	94.9	88.4
	Within Milan criteria (888 HCC, 2962 CLD)	0.91 (0.90 – 0.92)	72.5	90.0	86.0	33.8	99.1	84.0	60.6	95.8	87.7	75.2	87.7	84.9	59.1	96.0	87.5	63.5	94.9	87.6
	Achieved SVR* (17 HCC, 341 CLD)	0.91 (0.84 – 0.97)	52.9	97.1	95.0	35.3	99.7	96.7	41.2	99.1	96.4	52.9	95.9	93.9	41.2	99.1	96.4	41.2	99.0	96.2
	Non-SVR* (1497 HCC, 2601 CLD)	0.93 (0.92 – 0.94)	80.2	89.0	85.8	50.2	98.9	81.1	70.8	95.4	86.4	82.2	86.6	85.0	70.0	95.5	86.2	72.9	94.3	86.5
	On active HBV Rx** (31 HCC, 156 CLD)	0.88 (0.82 – 0.94)	51.6	97.4	89.8	35.5	100.0	89.3	35.5	98.7	88.2	51.6	96.2	88.8	35.5	98.7	88.2	39.5	98.7	89.0
	No HBV Rx** (1483 HCC, 2806 CLD)	0.93 (0.92 – 0.94)	80.5	89.6	86.5	50.4	99.0	82.2	71.2	95.7	87.2	82.5	87.3	85.6	70.3	95.8	87.0	73.2	94.7	87.3
	HCV positive cohort (1035 HCC, 1325 CLD)	0.92 (C.I. 0.91 – 0.93)	81.5	85.4	83.6	47.5	98.4	76.1	71.1	93.6	83.7	83.2	82.6	82.8	70.1	93.7	83.4	73.6	92.2	84.1
	HBV positive cohort (230 HCC, 704 CLD)	0.93 (C.I. 0.92 – 0.95)	73.5	95.7	90.3	54.8	99.4	88.4	62.6	97.3	88.8	74.4	95.0	89.9	61.7	97.7	88.9	64.4	96.9	88.9
	Other aetiology (230 HCC, 891 CLD)	0.95 (C.I. 0.94 – 0.97)	80.9	92.4	90.0	57.0	99.7	90.9	76.5	98.1	93.7	83.9	89.6	88.4	76.1	98.2	93.7	76.5	97.2	93.0
Germany	**Whole cohort** (275 HCC, 900 CLD)	0.94 (C.I. 0.93 – 0.96)	91.6	81.3	83.7	75.3	95.8	91.0	87.6	88.6	88.3	91.6	79.2	82.1	87.3	88.8	88.4	89.1	86.8	87.3
	Unifocal <2cm (7 HCC, 900 CLD)	0.93 (0.89 – 0.97)	100.0	81.3	81.5	42.9	95.8	95.4	71.4	88.6	88.4	100.0	79.2	79.4	71.4	88.8	88.6	85.7	86.8	86.8
	Unifocal <3cm (23 HCC, 900 CLD)	0.87 (0.81 – 0.94)	73.9	81.3	81.2	47.8	95.8	94.6	65.2	88.6	88.0	73.9	79.2	79.1	65.2	88.8	88.2	69.6	86.8	86.4
	Unifocal <4cm (38 HCC, 900 CLD)	0.87 (0.81 – 0.93)	73.7	81.3	81.0	50.0	95.8	93.9	68.4	88.6	87.7	73.7	79.2	79.0	68.4	88.8	87.9	71.1	86.8	86.1
	Unifocal <5cm (46 HCC, 900 CLD)	0.85 (0.79 – 0.91)	71.7	81.3	80.9	52.2	95.8	93.7	67.4	88.6	87.5	71.7	79.2	78.9	67.4	88.8	87.7	69.6	86.8	85.9
	Unifocal <10cm (73 HCC, 900 CLD)	0.90 (0.86 – 0.94)	82.2	81.3	81.4	64.4	95.8	93.4	79.5	88.6	87.9	82.2	79.2	79.5	79.5	88.8	88.1	80.8	86.8	86.3
	Within Milan criteria	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	HCV positive cohort (55 HCC, 252 CLD)	0.93 (C.I. 0.90 – 0.97)	90.9	79.8	81.8	76.4	95.0	91.7	83.6	87.3	86.6	90.9	78.6	80.8	83.6	87.5	86.8	85.5	84.9	85.0
	HBV positive cohort (32 HCC, 200 CLD)	0.94 (C.I. 0.91 – 0.98)	78.1	92.0	90.1	59.4	96.3	91.2	75.0	94.0	91.4	78.1	90.0	88.4	75.0	94.0	91.4	78.1	93.8	91.7
	Other aetiology (176 HCC, 280 CLD)	0.94 (C.I. 0.91 – 0.96)	93.8	72.5	80.7	77.8	95.4	88.6	91.5	83.9	86.7	93.8	69.3	78.7	91.5	84.3	87.1	92.1	80.5	85.0

Abbreviations: AUC, area under curve; C.I., 95% confidence interval; CLD, chronic liver disease; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; INR, international normalised ratio; Max. Sens, maximum sensitivity; Max. Spec., maximum specificity; NA, not available; *p=0.5152, **p=0.1194

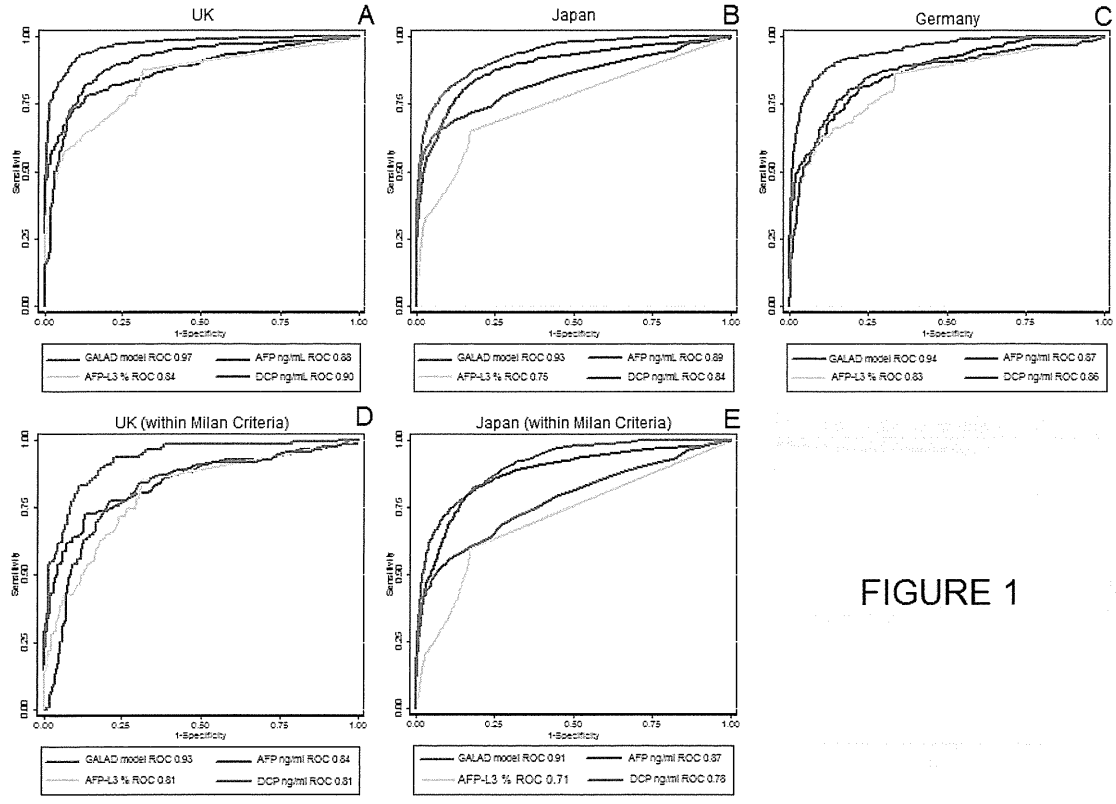

FIGURE 1

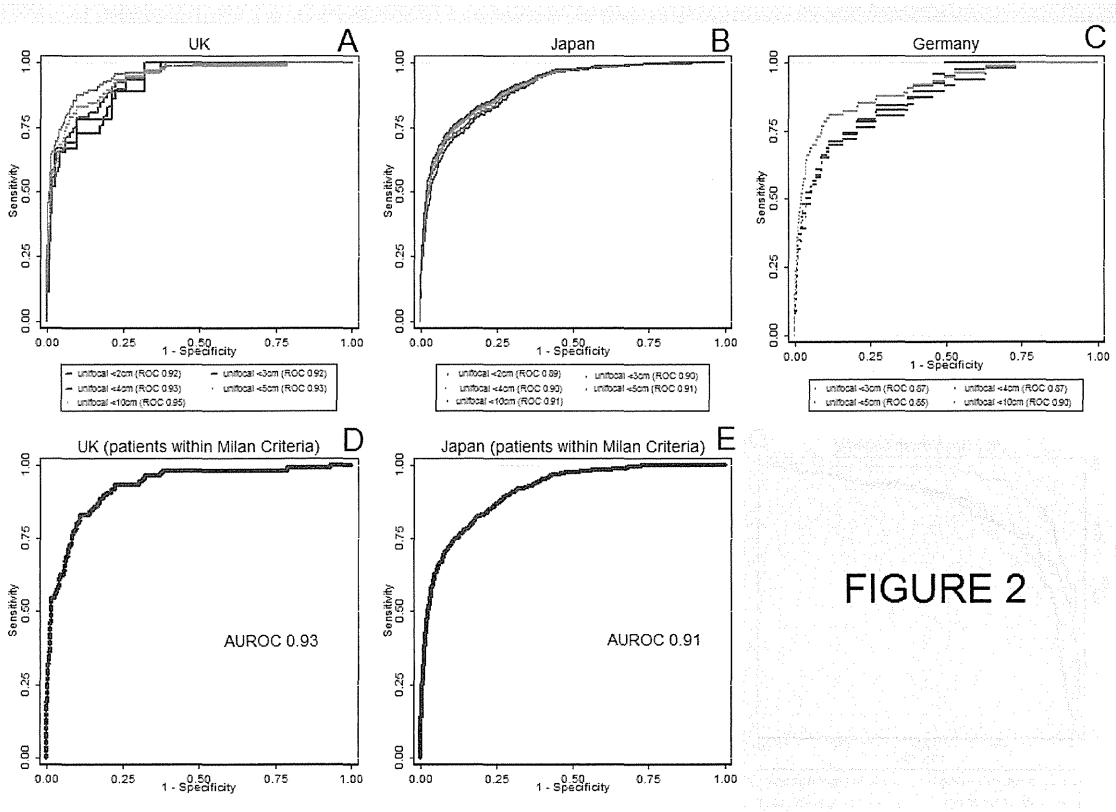

FIGURE 2

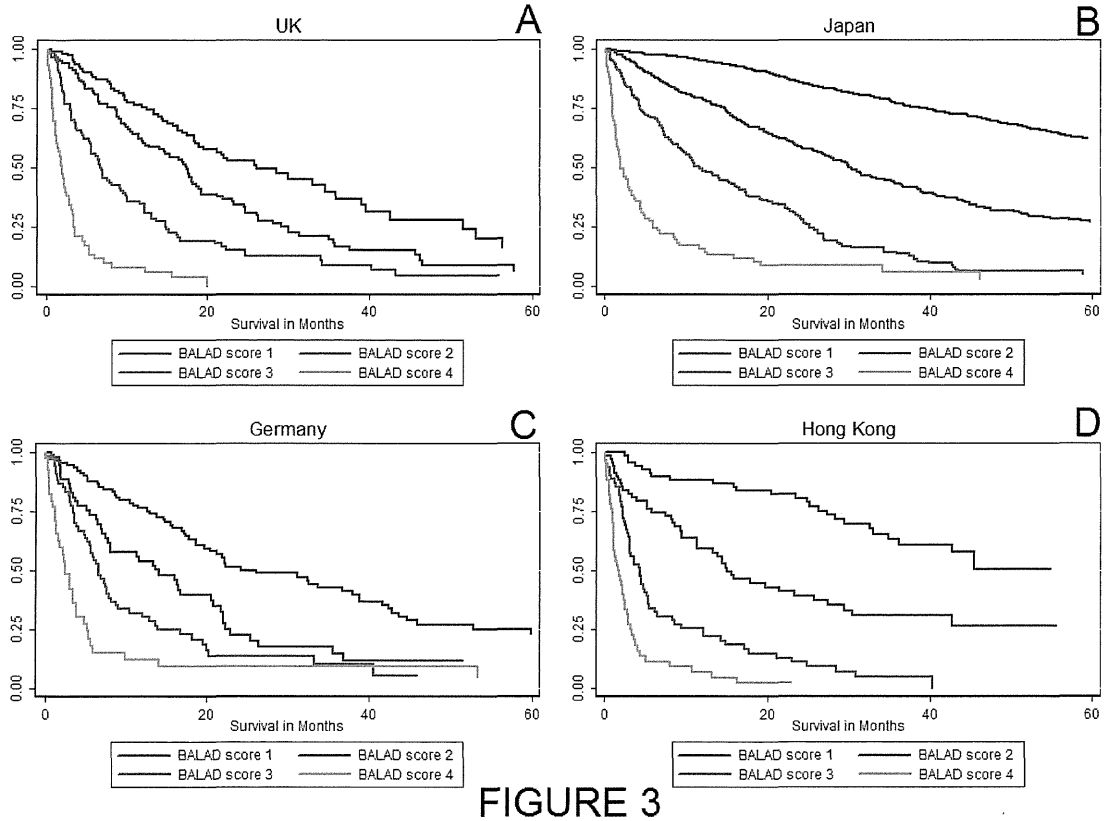

FIGURE 3