Table 1. Demographics and clinical data for the two cohorts

	UK ($n=319$)	Japan ($n=2599$)
Demographics		
Median age (IQR*)	66.4 (59.3–72.9)	67 (61.0–72.0)
Mean age (±s.d.)	65.4 (±9.7)	66.4 (±8.9)
Gender (M:F), %	82.4:17.6	71.7:28.3
Ethnicity		
Caucasian	266 (83.4%)	N/A
Other	53 (16.6%)	2599 (100%)
Aetiology		
Alcohol, %	25.1	N/A
HCV, %	12.9	74.3
HBV, %	9.1	12.4
HCV + HBV, %	0.6	1.7
Other (including those with multiple[a] aetiologies), %	48.3	11.2
Not known, %	4.1	0.4
HCC biomarkers		
AFP, ng ml^{-1}	57 (8.7–1264.3*), $n=319$	29.7 (9–208*), $n=2599$
Log_{10} AFP, ng ml^{-1}	1.76 (0.94–3.1*), $n=319$	1.5 (0.95–2.3*), $n=2599$
L3, %	16.6 (7–51.9*), $n=319$	1.4 (0–18*), $n=2599$
Log_{10} L3, %	1.22 (0.9–1.7*), $n=319$	0.15 (0–1.3*), $n=2599$
DCP, ng ml^{-1}	20.07 (2.6–169.7*), $n=319$	90 (26–797.5*), $n=2599$
Log_{10} DCP, ng ml^{-1}	1.37 (±1.2), $n=319$	1.95 (1.4–2.9*), $n=2599$
Liver function tests		
Albumin, g l^{-1}	38.4 (±5.6), $n=318$	35 (31–39*), $n=2599$
ALP, U l^{-1}	370.5 (259.5–558*), $n=318$	N/A
INR	1.1 (1.0–1.2*), $n=313$	1.1 (1.03–1.2*), $n=2431$
Bilirubin, μmol l^{-1}	17 (11–28*), $n=318$	15.4 (10.3–22.2*), $n=2599$
Child Pugh Score		
A:B:C:NK, %	74.0:22.6:2.8:0.6	67.1:26.3:6.6:0
Tumour characteristics		
Solitary:multifocal:NK, %	44.5:50.8:4.7	52.0:45.4:2.5
Maximum tumour diameter		
<2 cm, %	5.6	26.4
2–5 cm, %	37.6	54.1
>5 cm, %	30.1	13.4
>10 cm, %	12.2	3.5
NK or not specified, %	14.4	2.7
Macrovascular invasion (No:Yes:NK), %	68.3:26.0:5.6	68.1:31.6:0.3
Milan criteria (No:Yes:NK), %	67.7:24.5:7.8	39.1:56.0:4.8
Treatments		
Curative (intended: actual), %	19.3:16.1	66.3 (actual)
Palliative (intended: actual), %	80.7:83.9	33.7 (actual)
Median survival, months	16	47.2

Abbreviations: AFP = alpha-fetoprotein; ALP = alkaline phosphatase; DCP = Des-gamma carboxyprothrombin; F = female; HBV = hepatitis B virus; HCC = hepatocellular carcinoma; HCV = hepatitis C virus; INR = international normalised ratio; M = male; N/A = not applicable; NK = not known; s.d. = standard deviation. For all continuous variables, values are presented either as median (interquartile range*) or mean (±s.d.), the latter for normal distributions where appropriate.
[a]For example, alcoholic and HCV positive.

exponential or Weibull survival models, and Cox modelling does not directly model the baseline hazard function. The model as described here comprises two main components: the baseline hazard, which is described by a spline function consisting of a constant value and a function of log-time, and the covariate vector, which modifies risk based on the subject's covariate values. Each of these components can be recalibrated (Van Houwelingen, 2000) should the model not perform as expected. Given our intention to apply the model in two geographically distinct cohorts, we assessed the baseline hazard function, as clinical insights led us to expect a difference.

Stata version 12 was used for all analyses.

Replication of BALAD results and model derivation. As an exploratory step, we validated the original BALAD model in both the Japanese and UK cohorts. Like Toyoda *et al* (2006), we fitted univariable Cox regression models to verify the set of prognostic parameters and confirmed that statistical significance was maintained when entered into a multivariable model. The steps taken by Toyoda to dichotomise the continuous data were not replicated. The BALAD score was calculated for each cohort and discrimination was assessed by fitting Kaplan–Meier (KM) curves and measuring Harrell's *C*. A 'training' data set, which comprised 50% (Royston and Lambert, 2011) of the Japanese cohort, was used to derive the prognostic model, and the remainder was held back for validation. The random selection of the hold-back sample was stratified by treatment intention (potentially curative and palliative) such that each subset was equally represented in the training and validation data sets. A cohort of UK individuals was also used in the validation process.

New prognostic model. We then fitted flexible parametric survival models to the Japanese data, applying a more rigorous statistical approach in which the continuous form of the covariates was maintained and linearity of predictor–outcome relation was not assumed. Univariable and multivariable models were fitted to identify important prognostic factors with potential prognostic factors chosen from those that were not considered subjective. Martingale residuals were inspected to aid the choice of the appropriate covariate functional form and second-order fractional polynomials were explored, taking powers from the standard power set. Predictors were selected at the $P = 0.05$ level in the multivariable modelling procedure that combined backward elimination with the selection of an FP function. Models were compared using Akaike information criteria (AIC); a 4-point reduction (per additional covariate) is indicative of an improved model. Having identified a preferred prognostic model, we then fitted a model keeping only the serological cancer biomarkers to see if similar performance could be obtained at less 'cost'.

Development of scoring mechanisms. To assign risk groups 'Cox cut-points' were applied by splitting risk predictions, based on the relative part of the model only, in the training data at the 15th, 50th, and 85th percentiles. As a result, individuals were categorised into 1 of 4 levels of risk, ranging from low to high. We then calculated individual risk in the hold-back data and classified patients based on the cut-points established earlier. We refer to this discriminatory model as *BALAD-2d.* By incorporating risk as a function of time, that is, the baseline hazard function, we could estimate the probability of survival for each individual patient. We refer to this patient-level predictor as *BALAD-2p.*

Model validation. BALAD-2d: The prognostic model was validated using graphical methods (Royston and Lambert, 2011). A visual inspection of discrimination between the groups was performed and survival statistics were compared to assess the clinical relevance of the model. We assessed Harrell's *C* of each model in a number of patient subgroups: stage of disease, treatment intention, tumour size, and BCLC (available only in UK patients).

BALAD-2p: Stata's suite of flexible parametric modelling (stpm2) post-estimation tools were used to estimate population average survival for each validation cohort, thus allowing Kaplan–Meier curves depicting actual *vs* predicted survival to be plotted for each risk group. The similarity of the curves is indicative of the performance of the model. To determine if the model is appropriate for the estimation of patient-level survival in the UK validation set, or if recalibration was required, we assessed the similarity of the baseline hazard in each cohort by plotting the function. We also demonstrate the use of the BALAD-2p by example and report the results by graphical means.

RESULTS

Replication of BALAD results. We confirmed that for both the Japanese and UK cohorts the measures of serological cancer biomarkers and bilirubin are associated with increased risk of mortality (results not shown); albumin is associated with a decreased risk. The Kaplan–Meier survival curves according to BALAD scores are shown in Figure 1A and B. For BALAD model in the Japanese and UK cohorts the respective Harrell's *C*-statistics were 0.73 and 0.71, indicating similar discriminative performance. We note that for the BALAD model in the UK cohort there is overlap of the curves in the first 6 months and that there are very few patients in the highest-risk group. Table 2 reports the median survival with 95% confidence intervals; the estimates for the Japanese cohort are quite distinct; however, for the UK patients there is little difference in median survival for some of the groups, indicating that, from a clinical perspective, the BALAD model may have too many levels for use in the UK. The hazard ratio estimates for the BALAD score in the Japanese cohort ranged from 2.24 (95% CI, 1.85–2.72) in the lowest-risk group to 48.48 (95% CI, 30.52–77.02) in the highest-risk group. In the UK cohort, the corresponding figures were 1.93 (95% CI, 1.18–3.17) and 210.42

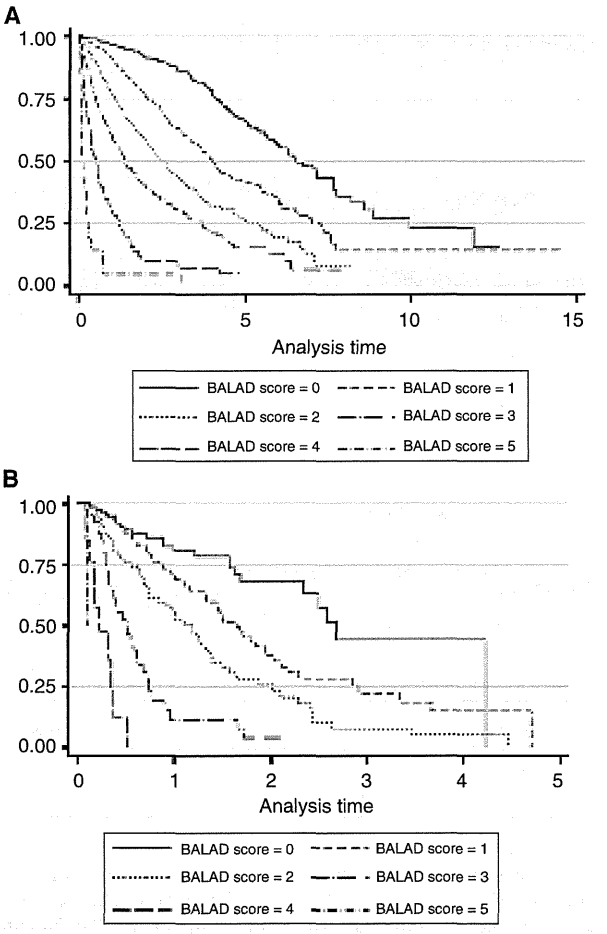


Figure 1. Survival according to the BALAD model. Kaplan–Meier curves showing survival according to the original BALAD model in (**A**) Japanese and (**B**) UK cohorts.

Table 2. Median survival times for BALAD and BALAD-2d in Japanese (validation) and UK cohorts

	Japan				UK			
	Subjects	Median (years)	95% CI		Subjects	Median (years)	95% CI	
BALAD								
0	357	6.7	5.9	8.6	79.0	2.7	2.3	
1	436	4.1	3.8	5.3	88.0	1.6	1.3	1.9
2	261	2.5	2.1	2.9	79.0	1.2	0.7	1.4
3	155	1.4	1.2	2.0	44.0	0.5	0.3	0.6
4	50	0.6	0.4	0.9	13.0	0.2	0.1	0.4
5	12	0.1	0.0	0.3	2.0	0.1	0.1	
Total	1271	3.9	3.6	4.3	305.0	1.4	1.1	1.6
BALAD-2d								
1	172	7.1	6.7		97	2.3	1.7	3.7
2	483	5.9	4.8	7.8	90	1.6	1.2	2.2
3	425	3.1	2.5	3.4	73	0.8	0.7	1.3
4	191	0.8	0.7	1	44	0.3	0.3	0.5
Total	1271	3.9	3.6	4.3	304	1.4	1.1	1.6

Abbreviation: CI = confidence interval.

Table 3. Univariable analysis in Japanese training data

Variable (x)	Transform	HR	95% CI	*P*-value
Gender	NA	1.177	0.974, 1.422	0.092
Major VP	NA	6.095	4.925, 7.542	<0.001
Age (years)	X	1	0.99, 1.01	0.977
INR	x^{-2}	0.254	0.171, 0.378	<0.001
AFP	ln(x)	1.226	1.187, 1.267	<0.001
L3	$x^{1/2}$	1.189	1.156, 1.223	<0.001
DCP	ln(x)	1.271	1.229, 1.315	<0.001
Bilirubin	ln(x)	1.978	1.726, 2.267	<0.001
Albumin	X	0.903	0.889, 0.917	<0.001
Maximum tumour size (mm)	$x^{1/2}$	2.081	1.84, 2.355	<0.001

Abbreviations: AFP = alpha-fetoprotein; CI = confidence interval; DCP = Des-gamma carboxyprothrombin; HR = hazards ratio; INR = international normalised ratio.

(95% CI, 20.87–2121.74). Between these two values, each cohort indicated an increasing trend in risk.

New prognostic model. We split the 2599 Japanese patients into a 'training' set of 1327 patients and a hold-back set of 1272, and, as a result of stratification by treatment intent, each data set was approximately equal in terms of the proportion of curative (and therefore palliative) patients (33.5% training and 33.8% validation).

The univariable analysis confirmed that the variables in the original BALAD model are all highly prognostic (Table 3), and these factors maintained statistical significance in the resulting multivariable model (data not shown). The fractional polynomial transformations identified for the multivariable model were a log transform for DCP and a square-root for bilirubin. The AIC for this model was 2341. An increase in each of the markers, other than albumin, is associated with an increase in risk, and increased albumin has a beneficial effect on prognosis.

The linear predictor resulting from the multivariable analysis considering the 5-serological cancer biomarkers – bilirubin, albumin, AFP-L3, AFP, and DCP – as potential prognostic factors is reported below. This function, the BALAD-2d score, calculates the log cumulative hazard for an individual:

Linear predictor $(xb) = 0.02*(afp\_c - 2.57) + 0.012*(AFP\text{-}L3 - 14.19) + 0.19*(\ln(DCP) - 1.93) + 0.17*((bili(\mu moll)^{1/2}) - 4.50) - 0.09*(alb(gl) - 35.11)$

As part of the modelling procedure AFP was capped at 50 000 units. Both AFP and DCP are modelled as per 1000 units.

The multivariable model incorporating just the three serological cancer biomarkers had considerable overlap between the two lower-risk groups, indicating that the discriminative performance was considerably poorer than the 5-serological-cancer-biomarker model that included bilirubin and albumin (Harrell's *C* of 0.69, AIC 2536) (Figure 2A and B).

BALAD-2d validation in the Japanese cohort. Application of the Cox cut-points for the linear predictor yielded four classes (1–4) of risk. These cut-points were as follows: $xb > 0.24$ (risk 1, low), 0.24 to > -0.91 (risk 2), -0.91 to > -1.74 (risk 3) and $\leqslant -1.74$ (risk 4, high). The KM survival curves depicting actual and predicted survival in the Japanese hold-back sample (Figure 2A) indicate that the risk groups are well discriminated (Harrell's *C* 0.74). The logrank test indicates a statistically significant risk difference ($P < 0.001$) and the differences in survival between the groups are clinically meaningful and distinct (Table 2). Harrell's *C* was approximately equal in the subgroup comparisons detailed in the foregoing. Both BALAD and BALAD-2d models perform better in patients at greater risk.

Recalibration for use in the UK cohort. Figure 3A and B describe the baseline hazard function for each of the cohorts. The baseline hazards are similar in shape but differ in height or magnitude, indicating the need for recalibration (see Supplementary Data for methodology). Adjustment to the constant term in the spline function only was deemed sufficient. Figure 4A shows that for the recalibrated model the overall predicted survival curve approximates the true survival well; here we optimised the fit between 0 and 3 years. Survival is predicted best of all in the higher-risk groups, and there is some overestimation for patients in the lowest-risk group (Figure 4B). The model has an AIC of 827 compared with 1096 prior to recalibration, an improvement of 269 points.

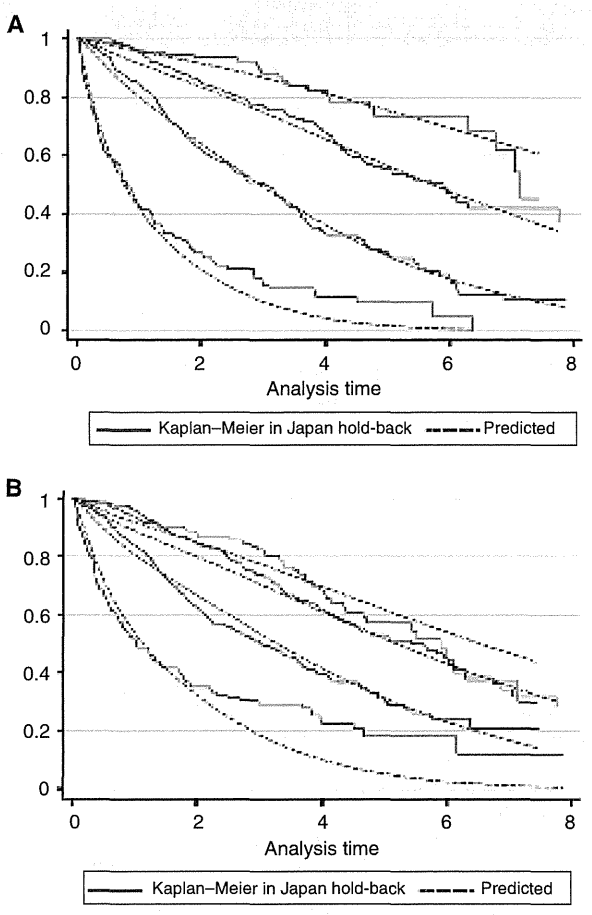


Figure 2. Comparison of five-marker and three-marker BALAD-2d model. Kaplan–Meier curves depicting actual (solid line) and predicted (dashed line) survival using a multivariable model incorporating (**A**) five serological cancer biomarkers and (**B**) three serological cancer biomarkers from the Japanese hold-back sample.

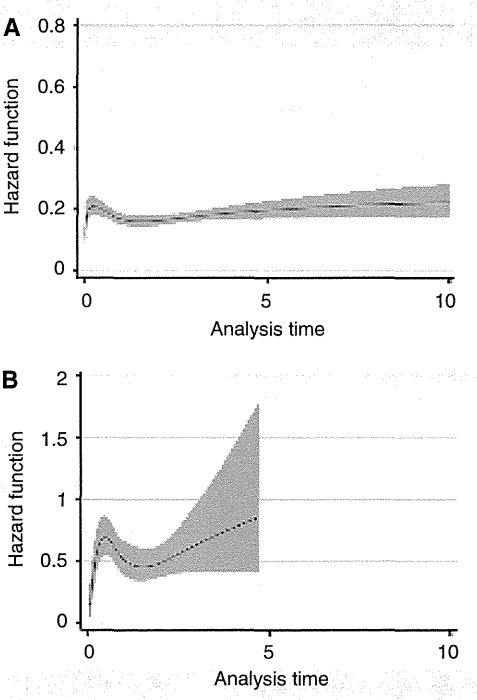


Figure 3. Baseline hazards. Plots illustrating the baseline hazard function with 95% confidence interval, CI (shaded region) for each of the (**A**) Japanese and (**B**) UK cohorts.

Note that the AIC in the Japanese cohort is not comparable to that in the UK.

Figure 5 demonstrates patient-level survival estimation and reports predictions of 2-year survival for increasing levels of albumin and bilirubin; all other parameters are fixed (AFP 34, AFP-L3 16.1, and DCP 1.14). Each curve describes an incrementally different level of albumin, and each point along a curve represents a change in bilirubin. As observed in the regression analysis, increased bilirubin is associated with an increase in risk and albumin is negatively correlated with risk.

DISCUSSION

We have confirmed that the biological factors identified in the original BALAD model, as described by Toyoda *et al* (2006), are highly prognostic. When applied to the UK patients, the BALAD model gave good discrimination, although performance appeared poorer among the UK cohort, particularly in groups 4 and 5, those with the worst survival (Figure 1B). Most likely the reason for the apparent limited degree of discrimination in these groups is related to the very small numbers (only 13 cases in group 4 and 20 in group 5), and the fact that survival for both these groups is only of the order of a few weeks indicates that there is little scope for clear discrimination. The overlap between risk groups is less evident for BALAD-2d, and the Harrell's *C*-test is, in fact, similar, indicating that there is no major difference in discriminatory performance despite the new model using just four risk groups.

One of the concerns about current staging systems is that they encompass factors that are inherently subjective, leading to a potential lack of consistency between observers. For example, one of the most widely used, the CLIP (Group, 1998), system estimates the extent of tumour as $>$ or $<50\%$ of the total liver volume, a measurement that, in clinical practice, is difficult to ascertain with any degree of certainty or consistency. Others such as CUPI (Leung *et al*, 2002) and BCLC (Llovet *et al*, 2008) demand a decision as to whether or not patients are symptomatic. Again, in practice this assessment may be highly variable between observers. Even widely used measures of liver function such as the C-P classification (Child and Turcotte, 1964; Pugh *et al*, 1973), which was developed for patients with cirrhosis rather than HCC, is remarkably subjective. Thus, presence/severity of ascites (one of the constituent variables in the C-P score) is, by some practitioners, based on whether or not subjects have ever developed ascites. Others may include ascites even when it is detectable only by radiological scanning and some may consider ascites to be 'absent' if it is controlled by diuretics. Encephalopathy may be equally difficult to grade, because of many of the early symptoms overlapping with those that may be attributable to the HCC. Such concerns have led to the development of objective measures of liver function such as the MELD score (Malinchoc *et al*, 2000; Botta *et al*, 2003), which is based solely on blood tests.

The second weakness of some of the current staging systems is that, perhaps in the pursuit of simplicity, they handle the relevant data in a categorical manner when it is, in fact, generated as a continuous variable. The loss of information consequent upon this

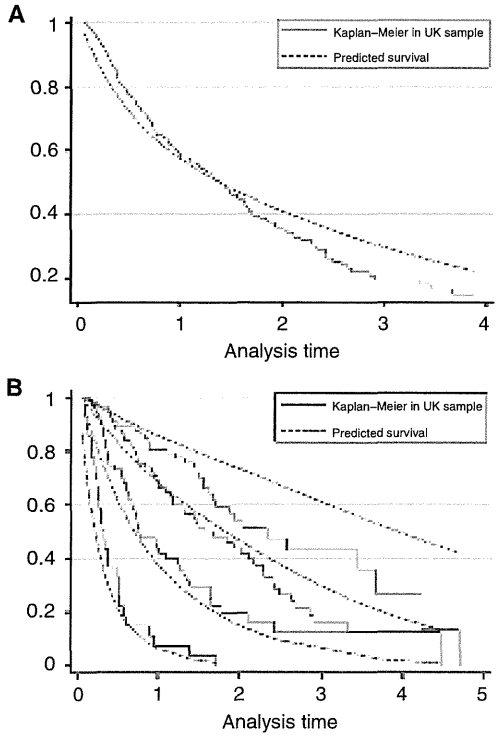


Figure 4. Kaplan–Meier curves of actual *vs* predicted survival in the UK cohort. Kaplan–Meier curves showing actual (solid line) vs predicted (dashed line) survival (**A**) overall and (**B**) by risk group, using the recalibrated model in the UK cohort.

approach is now increasingly well recognised, and it has been suggested that dichotomisation of continuous data in a multiple regression procedure may be associated with considerable loss of statistical power and introduction of bias (Del Priore *et al*, 1997; Royston *et al*, 2006). The most noticeable impact lies in those patients who fall around the 'cut-point', that is, just below or above the value used to define the two levels of the binary variable. They may be classified as having very different risk. Equally, in the C-P classification, a patient with a score based on a serum bilirubin level of 51 $\mu mol\,l^{-1}$ has the same impact as one with a value of 500 $\mu mol\,l^{-1}$ and a serum albumin of 24 $g\,l^{-1}$, a similar impact as a serum albumin of 10 $g\,l^{-1}$. Indeed, it has been shown that dichotomising (at the median) a normally distributed variable is equivalent to losing a third of the data. When the variable in question is exponentially distributed then such a conversion is equivalent to a loss of around half of the data (Royston *et al*, 2006). It has also been shown that in the case of logistic regression the chance of false positives is increased; as the sample size increases, so does the chance of such errors (Austin and Brunner, 2004).

Among the numerous biomarkers (Mann *et al*, 2007) that have been proposed for prognostication in HCC, the three used here have the advantage of being commercially available on a single platform and having regulatory approval in Japan, US, and Europe. All three are well documented to have prognostic significance when used individually, prognosis becoming poorer with increasing levels (Nagaoka *et al*, 2003; Toyoda *et al*, 2007; Nouso *et al*, 2011). AFP is also included in some staging systems such as CLIP (Group, 1998), where a level of $>400\,ng\,ml^{-1}$ is an adverse feature, and in the UK guidelines for liver transplantation levels of $>10\,000$ are a contraindication to transplantation (NHSBT, 2013).

The BALAD model has an advantage over current staging systems in that it is entirely objective. The Toyoda model utilised a common method for determining the value at which continuous covariates were dichotomised, that is, multiple testing in search of the 'optimal' value. Altman *et al* (1994) appropriately refer to this approach as the 'minimum *P*-value approach', as the use of the term optimal is perhaps misleading. There are several issues with this approach. The value, if truly optimal, is likely to be such only in the derivation cohort. Furthermore, the values identified at the univariable level are not necessarily accurate in the multivariable setting. The Type I error rate is greatly increased, and as such there is a greater risk of incorrectly identifying prognostic factors. The extent of this increase is influenced by the number of values tested, and is reported to be in the region of 40% (Altman *et al*, 1994).

In this paper we have undertaken further detailed analyses of the data in its continuous form. The regression analyses were performed using flexible parametric models (Royston and Lambert, 2011), described earlier, with which we were able to examine the baseline hazard function and assess the need for model recalibration for use in cohorts outside Japan. By exploring fractional polynomials (Sauerbrei and Royston, 1999) (FPs) many of the issues associated with dichotomisation and data-driven cut-points are minimised or avoided. If appropriate, more intricate relationships between the outcome and explanatory variables can be fitted. Over-fitting of either the baseline hazard or the covariate functional form is a potential issue but is unlikely if the number of knots and the power-set for the FPs are sensibly chosen. Given the size of the Japanese cohort in particular, throughout our analyses we purposely avoided over-interpretation of *P*-values and considered more the clinical significance of the results.

At present, we cannot be certain of the reason for the markedly differing survival in the two cohorts. However, the most plausible explanation, and one supported by the data presented here, is that the Japanese patients are diagnosed at a much earlier stage, and hence are much more likely to receive potentially curative therapy. Again the most plausible explanation is that the Japanese population at risk of HCC (those with chronic liver disease) is more rigorously screened than that in the UK. Although we cannot rule out the influence of aetiology, we can be confident that ethnicity is unlikely to be responsible as survival in Japan was only 7 months in the decade 1975–85 and has risen steadily thereafter, coincidentally with the introduction of screening (Ikai *et al*, 2010).

Our analyses demonstrated that in both the Japanese and UK cohorts BALAD-2d model has a marginally better level of discrimination compared with the BALAD model despite the former having just four risk levels. Furthermore, visual inspection of the BALAD model suggested that, in the UK cohort at least, six risk levels is too many and that the BALAD-2d model is more appropriate. For risk grouping, such as BALAD or BALAD-2d, it is implicit that patients belonging to the same risk group have equal survival probability. This is of course not necessarily the case; patients at the extremes of each risk group are classified as equal but most likely have quite different chance of survival. BALAD-2p model does not suffer from this limitation.

We addressed the discrepancy in magnitude of the underlying hazard between the UK and Japanese patients through model recalibration. The relatively minor adjustments required are indicative of the transferability of the model, and we have shown that the covariate effects in the Japanese cohort are applicable in UK patients. Had recalibration beyond simple adjustment to the height of the baseline hazard been required (e.g. changes to the shape of the baseline hazard, or even the covariate vector), then the validity of BALAD-2p model would have been questionable. In this case we had no such concerns. Although the BALAD-2p model builds on the BALAD model's concept by including the ability to predict patient-level survival (Figure 5) and as such is a more powerful prognostic device, validation in other regions of the

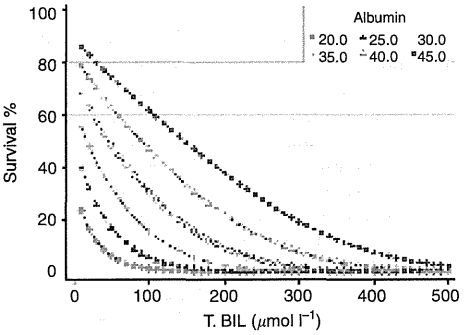


Figure 5. Example of patient-level survival estimation. Reporting predictions of 2-year survival and describing the impact of increasing albumin and bilirubin; all other parameters are fixed (AFP 34, AFP-L3 16.1, and DCP 1.14).

world, especially where the aetiology is related to Hepatitis B, is still required.

ACKNOWLEDGEMENTS

The analysis undertaken in this research was partly funded by Liverpool Health Partners.

REFERENCES

Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994) Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* **86**: 829–835.

Austin PC, Brunner LJ (2004) Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat Med* **23**: 1159–1178.

Botta F, Giannini E, Romagnoli P, Fasoli A, Malfatti F, Chiarbonello B, Testa E, Risso D, Colla G, Testa R (2003) MELD scoring system is useful for predicting prognosis in patients with liver cirrhosis and is correlated with residual liver function: a European study. *Gut* **52**: 134–139.

Chan SL, Mo FK, Johnson PJ, Liem GS, Chan TC, Poon MC, Ma BB, Leung TW, Lai P, Chan AT (2011) Prospective validation of the Chinese University Prognostic Index and comparison with other staging systems for hepatocellular carcinoma in an Asian population. *J Gastroenterol Hepatol* **26**: 340–347.

Chen C-H, Hu F-C, Huang G-T, Lee P-H, Tsang Y-M, Cheng A-L, Chen D-S, Wang J-D, Sheu J-C (2009) Applicability of staging systems for patients with hepatocellular carcinoma is dependent on treatment method–analysis of 2010 Taiwanese patients. *Eu J Cancer* **45**: 1630–1639.

Chevret S, Trinchet J-C, Mathieu D, Rached AA, Beaugrand M, Chastang C (1999) A new prognostic classification for predicting survival in patients with hepatocellular carcinoma. *J Hepatol* **31**: 133–141.

Child CG, Turcotte J (1964) Surgery and portal hypertension. *Major Prob Clin Surg* **1**: 1.

Cho YK, Chung JW, Kim JK, Ahn YS, Kim MY, Park YO, Kim WT, Byun JH (2008) Comparison of 7 staging systems for patients with hepatocellular carcinoma undergoing transarterial chemoembolization. *Cancer* **112**: 352–361.

Collette S, Bonnetain F, Paoletti X, Doffoel M, Bouche O, Raoul J, Rougier P, Masskouri F, Bedenne L, Barbare J (2008) Prognosis of advanced hepatocellular carcinoma: comparison of three staging systems in two French clinical trials. *Ann Oncol* **19**: 1117–1126.

Del Priore G, Zandieh P, Lee M-J (1997) Treatment of continuous data as categoric variables in obstetrics and gynecology. *Obstetr Gynecol* **89**: 351–354.

Greene FL, Page DL, Fleming ID, Balch CM, Fritz AG (2002) *AJCC Cancer Staging Handbook Plus EZTNM*. Springer.

Group C (1998) A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. *Hepatology* **28**: 751–755.

Huitzil-Melendez F-D, Capanu M, O'reilly EM, Duffy A, Gansukh B, Saltz LL, Abou-Alfa GK (2010) Advanced hepatocellular carcinoma: which staging systems best predict prognosis? *J Clin Oncol* **28**: 2889–2895.

Ikai I, Kudo M, Arii S, Omata M, Kojiro M, Sakamoto M, Takayasu K, Hayashi N, Makuuchi M, Matsuyama Y (2010) Report of the 18th follow-up survey of primary liver cancer in Japan. *Hepatol Re* **40**: 1043–1059.

Kagebayashi C, Yamaguchi I, Akinaga A, Kitano H, Yokoyama K, Satomura M, Kurosawa T, Watanabe M, Kawabata T, Chang W (2009) Automated immunoassay system for AFP-L3% using on-chip electrokinetic reaction and separation by affinity electrophoresis. *Anal biochemy* **388**: 306–311.

Kudo M, Chung H, Haji S, Osaki Y, Oka H, Seki T, Kasugai H, Sasaki Y, Matsunaga T (2004) Validation of a new prognostic staging system for hepatocellular carcinoma: the JIS score compared with the CLIP score. *Hepatology* **40**: 1396–1405.

Kudo M, Chung H, Osaki Y (2003) Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). *J Gastroenterol* **38**: 207–215.

Leung TW, Tang AM, Zee B, Lau W, Lai P, Leung K, Lau JT, Yu SC, Johnson PJ (2002) Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system. *Cancer* **94**: 1760–1769.

Llovet JM, Brú C, Bruix J (2008) *Prognosis of hepatocellular carcinoma: the BCLC staging classification*. In:Seminars in liver disease, 2008. © 1999 by Thieme Medical Publishers, Inc. 329–338.

Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PC (2000) A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* **31**: 864–871.

Mann CD, Neal CP, Garcea G, Manson MM, Dennison AR, Berry DP (2007) Prognostic molecular markers in hepatocellular carcinoma: a systematic review. *Eur J Cancer* **43**: 979–992.

Marrero JA, Fontana RJ, Barrat A, Askari F, Conjeevaram HS, Su GL, Lok AS (2005) Prognosis of hepatocellular carcinoma: comparison of 7 staging systems in an American cohort. *Hepatology* **41**: 707–715.

Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, Montalto F, Ammatuna M, Morabito A, Gennari L (1996) Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *New Engl J Med* **334**: 693–700.

Nagaoka S, Yatsuhashi H, Hamada H, Yano K, Matsumoto T, Daikoku M, Arisawa K, Ishibashi H, Koga M, Sata M (2003) The des-γ-carboxy prothrombin index is a new prognostic indicator for hepatocellular carcinoma. *Cancer* **98**: 2671–2677.

NHSBT (2013) *Liver Transplantation; Selection Criteria and Recipient Registration*. NHS Blood and Transplant (NHSBT) Liver Advisory Group.

Nouso K, Kobayashi Y, Nakamura S, Kobayashi S, Takayama H, Toshimori J, Kuwaki K, Hagihara H, Onishi H, Miyake Y (2011) Prognostic importance of fucosylated alpha-fetoprotein in hepatocellular carcinoma patients with low alpha-fetoprotein. *J Gastroenterol Hepatol* **26**: 1195–1200.

Okuda K, Ohtsuki T, Obata H, Tomimatsu M, Okazaki N, Hasegawa H, Nakajima Y, Ohnishi K (1985) Natural history of hepatocellular carcinoma and prognosis in relation to treatment study of 850 patients. *Cancer* **56**: 918–928.

Pugh R, Murray-Lyon I, Dawson J, Pietroni M, Williams R (1973) Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg* **60**: 646–649.

Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* **25**: 127–141.

Royston P, Lambert PC (2011) *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press: USA.

Sauerbrei W, Royston P (1999) Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Society: Series A* **162**: 71–94.

Sobin LH, Fleming ID (1997) TNM classification of malignant tumors, (1997). *Cancer* **80**: 1803–1804.

Sobin LH, Gospodarowicz MK, Wittekind C (2011) *TNM classification of malignant tumours*. Wiley.com.

Taktak AFG, Eleuteri A, Lake SP, Fisher AC (2007) Evaluation of prognostic models: discrimination and calibration performance. *Comput Intell Med (Plymouth)*. Available at http://pcwww.liv.ac.uk/~afgt/CIMED07_1.pdf.

Toyoda H, Kumada T, Osaki Y, Oka H, Kudo M (2007) Role of tumor markers in assessment of tumor progression and prediction of outcomes in patients with hepatocellular carcinoma. *Hepatol Res* **37**: S166–S171.

Toyoda H, Kumada T, Osaki Y, Oka H, Urano F, Kudo M, Matsunaga T (2006) Staging hepatocellular carcinoma by a novel scoring system (BALAD score) based on serum markers. *Clin Gastroenterol Hepatol* **4**: 1528–1536.

UICC U (2002) *TNM Classification of Malignant Tumours*, Sobin LH, Wittekind Ch eds. Wiley-Liss: New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

Van Houwelingen HC (2000) Validation, calibration, revision and combination of prognostic survival models. *Stat Med* **19**: 3401–3415.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)

J Gastroenterol (2014) 49:555–563
DOI 10.1007/s00535-013-0883-1

The Japanese Society of Gastroenterology

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

High-sensitivity *Lens culinaris* agglutinin-reactive alpha-fetoprotein assay predicts early detection of hepatocellular carcinoma

Takashi Kumada · Hidenori Toyoda · Toshifumi Tada · Seiki Kiriyama · Makoto Tanikawa · Yasuhiro Hisanaga · Akira Kanamori · Junko Tanaka · Chiaki Kagebayashi · Shinji Satomura

Received: 19 May 2013 / Accepted: 5 September 2013 / Published online: 23 September 2013

Abstract
Background Prognosis of patients with hepatocellular carcinoma (HCC) remains poor because HCC is frequently diagnosed late. Therefore, regular surveillance has been recommended to detect HCC at the early stage when curative treatments can be applied. HCC biomarkers, including *Lens culinaris* agglutinin-reactive fraction of alpha-fetoprotein (AFP-L3), are widely used for surveillance in Japan. A newly developed immunoassay system measures AFP-L3 % with high sensitivity. This retrospective study aimed to evaluate clinical utility of high-sensitivity AFP-L3 (hs-AFP-L3) as a predictor of early stage HCC in surveillance at a single site.
Methods Of consecutive 2830 patients in the surveillance between 2000 and 2009, 104 HCC-developed and 104 non-HCC patients were selected by eligibility criteria and propensity score matching. Samples were obtained from the HCC patients who had blood drawn annually for 3 years prior to HCC diagnosis.
Results In the present study, hs-AFP-L3 was elevated 1 year prior to diagnosis in 34.3 % of patients. The survival rate of patients with the hs-AFP-L3 $\geq$ 7 % at 1 year prior to diagnosis was significantly lower than that of patients with hs-AFP-L3 $<$ 7 %.
Conclusions Elevation of hs-AFP-L3 was early predictive of development of HCC even at low AFP levels and in absence of ultrasound findings of suspicious HCC. The hs-AFP-L3 should be added to surveillance programs with US because elevated hs-AFP-L3 may be a trigger to perform enhanced imaging modalities for confirmation of HCC.

Keywords Surveillance · A propensity score analysis · High-sensitivity AFP-L3 · DCP · HCC

Abbreviations

HCC	Hepatocellular carcinoma
AFP	Alpha-fetoprotein
AFP-L3	*Lens culinaris* agglutinin-reactive fraction of AFP
hs-AFP-L3	High-sensitivity AFP-L3
US	Ultrasound
DCP	Des-gamma-carboxy prothrombin
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
ALT	Alanine aminotransferase
MRI	Magnetic resonance imaging

T. Kumada (✉) · H. Toyoda · T. Tada · S. Kiriyama · M. Tanikawa · Y. Hisanaga · A. Kanamori
Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, 4-86 Minaminokawa-cho, Ogaki, Gifu 503-8052, Japan
e-mail: hosp3@omh.ogaki.gifu.jp

J. Tanaka
Department of Epidemiology Infectious Disease Control and Prevention, Hiroshima University Institute of Biomedical and Health Sciences, Hiroshima, Japan

C. Kagebayashi · S. Satomura
Diagnostic Division, Wako Pure Chemical Industries Ltd., Osaka, Japan

Introduction

Hepatocellular carcinoma (HCC) is the third most common cause of death from cancer worldwide [1], and poor prognosis is reported because HCC is frequently diagnosed at late stages and is often untreatable. Therefore, surveillance for HCC has been advocated to detect HCC at

early stages when curative treatments can be applied [2, 3]. Global liver associations, including the American Association for the Study of Liver Disease (AASLD), the European Association for the Study of the Liver (EASL), and the Asian Pacific Association for the Study of the Liver (APASL), recommend regular surveillance on patients at high risk for HCC [4–6]. The most common tests used for surveillance are alpha-fetoprotein (AFP) tests and ultrasound (US). EASL and APASL adopt AFP and US in their guidelines, while AASLD recommends only US. Interpretation of US can be challenging when routine screening and comparison to previous imaging results are impossible or when US are performed by different institutes or instruments, whereas HCC biomarker values can be used independently with appropriate cutoff values. The Japan Society of Hepatology (JSH) has recommended not only US but also assays of three biomarkers: AFP, *Lens culinaris* agglutinin-reactive fraction of alpha-fetoprotein (AFP-L3), and des-gamma-carboxy prothrombin (DCP) [7].

However, AFP levels are often elevated even in patients with benign liver diseases. The low specificity of AFP has been a cause of concern for use as a HCC marker [8–10]. In contrast, a rate of AFP-L3 in total AFP (AFP-L3 %) has been reported to be highly specific for HCC in many studies [11–13]; however, accurate measurements of AFP-L3 % have been limited to patients having AFP >20 ng/mL by insufficient analytical sensitivity on a conventional assay system that is a liquid-phase binding assay (LiBASys) [14]. Recently, a micro-total analysis system (μTAS) based lectin-affinity electrophoresis using microfluidics technology has enabled accurate measurements of AFP-L3 % even at low AFP [15]. The high-sensitivity AFP-L3 (hs-AFP-L3) assay has demonstrated improvement in clinical sensitivity and predicting of prognosis in HCC patients with AFP < 20 ng/mL [16–18]. The Liver Cancer Study Group of Japan has reported that 37 % of HCC patients had low AFP (<15 ng/mL) at the HCC diagnosis [19]. They also show that 34 % of patients had tumors with maximum diameter of <2 cm. Early HCC is a distinct clinical entity with a high rate of surgical cure and detection of early HCC results in long-term survival [20]. However, elevated AFP is not always observed in patients with such small tumors. Therefore, the hs-AFP-L3 assay which can measure serum levels at low AFP is expected to improve detection of HCC at the early stage. Moreover, lower cutoff values for hs-AFP-L3 has been considered to improve clinical sensitivity [16–18].

In this study, clinical utility in early prediction of development of HCC in our study cohort under surveillance using hs-AFP-L3 and analyzed retrospectively is reported.

Patients and methods

Patients

The study protocol was approved by the Institutional Ethics Committee of Ogaki Municipal Hospital in January 2009 and was in compliance with the Declaration of Helsinki. Written informed consent for use of stored serum samples for the study was obtained from the enrolled patients.

Between 2000 and 2009, a total of consecutive 2830 patients positive for hepatitis B surface antigen (HBsAg) or anti-hepatitis C virus (HCV) antibody who visited the Department of Gastroenterology and Hepatology at Ogaki Municipal Hospital were prospectively enrolled in our HCC surveillance. Of the 2830, 1214 patients met eligibility criteria: HBsAg- or HCV RNA-positive for more than 6 months, follow-up period of >3 years before HCC diagnosis, availability of sera sampled at least twice at 12-month intervals, maximal tumor diameter <3 cm and 3 nodules or less at diagnosis, and no oral intake of warfarin which is a DCP-inducing agent.

Of these 1214 patients, 114 patients had HCC and 1100 patients had no evidence of HCC during follow-up period. To reduce the confounding effects of covariates between HCC and control patients, we selected patients using propensity score matching. Six covariates including age, gender, etiology (HBV or HCV), Child-Pugh classification, platelet number, and alanine aminotransferase (ALT) except tumor markers were used. We computed the propensity score by using logistic regression with the independent variable including age (<65 years or ≤65 years), sex (female or male), etiology (HBV or HCV), Child-Pugh classification (A, B, or C), platelet count (>150 × $10^3/m^3$ or ≤150 × $10^3/m^3$), and ALT activity (≤40 IU/mL or >40 IU/mL) as shown in previous reported cut-off values according to the previous reports [21, 22]. This model yielded a *c* statistic of 0.832 (95 % confidence interval [CI], 0.797–0.866), indicating a strong ability to differentiate between HCC and control patients. Calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test [23]. The *P* value of the calculated propensity score was 0.647 based on the Hosmer–Lemeshow test and showed an absence of bias. We were able to match 104 HCC developed patients to 104 non-HCC developing patients. Table 1 shows demographics of HCC and non-HCC groups. The median of tumor size was 1.9 cm. The 69 % of HCC patients had single tumor and the 86 % of HCC patients were at TNM stage I and II.

Surveillance and diagnosis

According to Clinical Practice Guidelines for Hepatocellular Carcinoma in Japan [7], we performed US and three

Table 1 Demographics and propensity score matching

Characteristics		HCC ($n = 104$)	Non-HCC ($n = 104$)	P value
Age (years)	Median (range)	67 (37–81)	68 (14–84)	0.980
Gender	Male/female	58 (56 %)/46 (44 %)	58 (56 %)/46 (44 %)	0.889
Etiology	B/C/B + C	14 (13 %)/89 (86 %)/1 (1 %)	14 (13 %)/89 (86 %)/1 (1 %)	1.000
Child-Pugh classification	A/B/C	82 (79 %)/18 (17 %)/4 (4 %)	84 (81 %)/17 (16 %)/3 (3 %)	0.907
ALT (IU/L)	Median (range)	49 (7–361)	46 (12–321)	0.582
Platelet ($\times 10^4/mm^3$)	Median (range)	10.1 (3.2–34.0)	12.1 (2.1–41.4)	0.150
Tumor size (cm)	Median (25 %, 75 % quartile)	1.9 (1.5, 2.3)	NA	NA
Tumor number	Single/Multiple	72 (69 %)/32 (31 %)	NA	NA
TNM stage	I/II/III	49 (47 %)/41 (39 %)/14 (14 %)	NA	NA

biomarker studies (AFP, AFP-L3, and DCP) every 3–4 months and dynamic magnetic resonance imaging (MRI) every 12 months for cirrhosis patients under surveillance. For patients with chronic hepatitis, we performed US and three biomarker studies every 6 months. For diagnostic confirmation of HCC, patients had a dynamic MRI when US suggested progression in nodular lesion, change of echo pattern in nodules, or increased biomarkers: continuous elevation of AFP or increase to AFP 200 ng/mL or more, AFP-L3 15 % or more, or DCP 40 mAU/mL or more. The hs-AFP-L3 assay was not available for the surveillance of those days.

Forty-five patients were diagnosed as HCC histologically (surgical specimen, 39 patients; US-guided needle biopsy specimens, 6 patients). The remaining 59 patients were diagnosed as HCC as typical findings of dynamic MRI including hypervascular in the arterial phase with washout in the portal venous or delayed phase [4].

Treatments

Individual decisions for a primary treatment were generally made on the basis of the guidelines for HCC in Japan [7]. Patients were initially assessed for eligibility for resection. When patients declined or were deemed ineligible for resection, they underwent locoregional ablative therapy (LAT) as a second option or transcatheter arterial chemoembolization (TACE) as a third one. Of the enrolled 104 patients, 99 patients underwent resection ($n = 39$), LAT ($n = 23$), or TACE ($n = 37$: including patients with both LAT and TACE). Five patients did not receive any treatment for HCC. No patient underwent liver transplantation.

Imaging modalities

B-mode US was performed with an Aplio XV or XG ultrasound system (Toshiba Medical System, Tokyo, Japan) equipped with a convex probe (PUT-375BT). MR imaging was performed using a superconducting scanner operating at 1.5 T (Signa Twin Speed; General Electric Medical Systems, Milwaukee, WI). MR images were obtained in the axial plane with a phased-array multicoil for the body. To scan whole livers, the section thickness was 8–10 mm with 2- and 3-mm intersectional gaps, depending on liver size. Breath-hold T1-weighted in-phase and out-of-phase fast spoiled gradient-recalled echo (SPGR, 200/dual echo [4.3/2.1] [TR/TE], 80° flip angle, one signal averaged) MR images were obtained with a field of view of 36–42 cm and a 256 × 192 matrix during a 22-s acquisition time. T2-weighted fat suppression fast spin-echo (2000/85 [TR/TE], two signal averaged) MR images with respiratory synchronization were obtained with a field of view of 36–42 cm and a 352 × 256 matrix. Breath-hold double arterial dynamic fast SPGR images (115/1.2 [TR/TE], 70° flip angle, one signal averaged) were obtained with a field of view of 36–42 cm and 512 × 192 matrix during a 12-s acquisition time. Dynamic MR images were obtained before and after an antecubital intravenous bolus injection of 0.1 mmol/kg of gadopentetate dimeglumine (Magnevist; Bayer in Japan, Tokyo, Japan) followed by 15–20 ml of a sterile normal saline flush. The optimum timing of start of scanning was decided for each case after 1 ml test injection of gadopentetate dimeglumine. The scan times were about 25, 40, and 60 s, and 2–2.5 min after initiation of the contrast injection, representing the early hepatic artery, late hepatic artery, portal vein, and equilibrium phase, respectively. All MR images except T2-weight MR images were obtained using array spatial sensitivity encoding technique (ASSET).

Assays of hs-AFP-L3, AFP, and DCP

For this retrospective study, the measurements of hs-AFP-L3, AFP, and DCP were achieved by using a microchip capillary electrophoresis and liquid-phase binding assay on μTASWako i30 auto analyzer (Wako Pure Chemical Industries, Ltd.) [16]. Analytical sensitivity of the μTAS is 0.3 ng/mL AFP, and percentage of AFP-L3 can be

measured when AFP-L3 is over 0.3 ng/mL. Analytical sensitivity of LiBASys is 0.8 ng/mL AFP, but AFP-L3 % can not be calculated at AFP < 10 ng/mL.

Samples were obtained from 104 HCC patients who had blood drawn annually for 3 years prior to the HCC diagnosis and stored at −80 °C until the measurements. In the HCC patients, stored serum samples at −3 years (over 30 months before, $n = 94$), −2 years (from 18 to 30 months before, $n = 97$), −1 year (from 6 to 18 months before, $n = 103$), and 0 year ($n = 104$) at the time of the HCC diagnosis were measured. In the non-HCC patients, similarly, stored serum samples at −3 years ($n = 99$), −2 years ($n = 104$), and −1 year ($n = 102$), and 0 year ($n = 104$) from the end of follow-up were measured.

Statistical analysis

To evaluate the diagnostic accuracy and predictive values of AFP, hs-AFP-L3, and DCP, sensitivity and specificity were calculated with cutoff values in the guidelines [7]. Furthermore, cutoff values of 5, 7, and 10 % for hs-AFP-L3 were used for this retrospective study according to previous reports [13, 16]. Serial changes of three biomarkers before the diagnosis of HCC were analyzed by Wilcoxon matched pair signed rank test. For the evaluation of prognosis, the long-term survival of patients with HCC was determined by the Kaplan–Meier method and the log-rank test was used to compare the survival rates. The values were considered significant when P value was <0.05. The analyses were performed using JMP10 statistical software (SAS Institute Japan, Japan).

The propensity score matching was performed with SPSS, version 18.0 for Windows (International Business Machines Corporation, Tokyo, Japan).

Results

Dynamic changes of biomarkers

The dynamic changes of hs-AFP-L3, AFP, and DCP in HCC patients at −3, −2, −1, and 0 year before diagnosis are shown in Fig. 1a, b, and c. The levels of hs-AFP-L3 at −1 year were significantly elevated from the levels at −2 years ($P = 0.0001$). The levels of hs-AFP-L3 at −0 year were significantly elevated from the levels at −1 year ($P = 0.0003$, Table 2). AFP and DCP were significantly elevated between −1 and 0 year ($P = 0.0315$

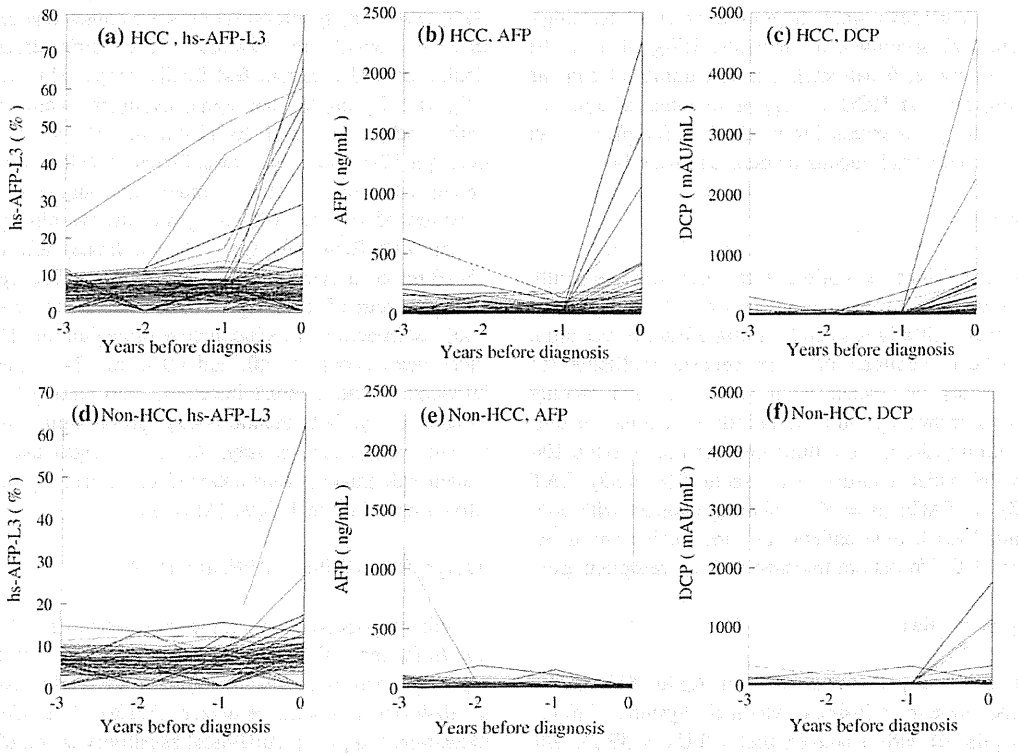


Fig. 1 Dynamic changes of biomarkers: **a** hs-AFP-L3, **b** AFP, and **c** DCP in each HCC patient ($n = 104$), and **d** hs-AFP-L3, **e** AFP, and **f** DCP in each non-HCC patient ($n = 104$)

Table 2 Serial changes of three biomarkers in HCC patients (Wilcoxon matched pair signed rank test)

Analyte	*P* value		
	At −3 year and −2 year	At −2 year and −1 year	At −1 year and 0 year
hs-AFP-L3	0.2935	0.0001	0.0003
AFP	0.4278	0.5359	0.0315
DCP	0.0926	0.6302	<0.0001

and $P < 0.0001$, respectively, Table 2). In non-HCC patients, no significant differences were observed for any markers (Fig. 1d–f). Only hs-AFP-L3 in HCC patients were significantly elevated 1 year prior to HCC diagnosis.

Sensitivity and specificity at diagnosis

Diagnostic sensitivity and specificity were evaluated for the hs-AFP-L3, AFP, DCP, and the combination of biomarkers (Table 3). The sensitivity was calculated by using HCC patient samples at diagnosis ($n = 104$) and the specificity was calculated by using non-HCC patient samples at −3 years ($n = 100$) to ensure that none had developed HCC for the following 3 years. Of the 104 HCC patients, 43 patients (41.3 %) had AFP < 10 ng/mL at which the conventional assay was not able to calculate AFP-L3 %. The sensitivity and specificity for hs-AFP-L3 were 11.5 and 100.0 %, respectively at a cutoff value of 15 %. A cutoff value of 7 % improved the sensitivity to 39.4 %. A combination assay with hs-AFP-L3, AFP, and DCP resulted in sensitivity of 60.6 % at diagnosis.

Sensitivity and specificity for 3 years before diagnosis

We calculated sensitivities using HCC samples at 3, 2, and 1 years prior to diagnosis. Similarly, specificities were calculated by using non-HCC samples (Table 4). The sensitivity and specificity for hs-AFP-L3 at −1 year were 34.3 and 74.7 %, respectively. The sensitivities at −1 year for AFP and DCP were 35.0 and 12.1 %, respectively. In HCC patients, hs-AFP-L3 turned positive at 34 patients (33.3 %) and stayed in positive at 27 patients (26.2 %) for two years till the diagnosis of HCC. In contrast, hs-AFP-L3 turned positive at 25 patients (24.3 %) and stayed in positive at 22 patients (21.4 %) for 2 years till the end of follow-up in non-HCC patients.

Table 3 Sensitivity and specificity at diagnosis

Analyte	Cutoff	Sensitivity (%)	Specificity (%)
hs-AFP-L3	5 %	50.9	51.0
	7 %	39.4	77.0
	10 %	16.3	96.0
	15 %	11.5	100.0
AFP	20 ng/mL	41.4	90.4
	200 ng/mL	12.5	99.0
DCP	40 mAU/mL	34.6	94.0
All biomarkers	7 % + 200 ng/mL + 40 mAU/mL	60.6	76.0

Comparison of tumor characteristics and survival rates

Comparing tumor characteristics at detection of HCC by a level of hs-AFP-L3 at −1 year, the tumor size, the number of tumors, and TNM stage between patients with hs-AFP-L3 ≥ 7 % and <7 % ($P = 0.064$, 0.821, and 0.504, respectively) were not statistically significant. The number of patients receiving curative treatments such as resection and LAT was significantly higher in patients with hs-AFP-L3 < 7 % ($P = 0.020$) (data not shown).

During the follow-up period after the diagnosis that was ranged from 4 to 110 months (median of 39 months), the survival rate of patients with hs-AFP-L3 ≥ 7 % was significantly lower than that of patients with hs-AFP-L3 < 7 % by using values at −1 year ($P = 0.039$) (Fig. 2). There was no statistical significance between patients with DCP ≥ 40 mAU/mL and patients with DCP < 40 mAU/mL ($P = 0.831$). No patients had AFP > 200 ng/mL at −1 year. The survival rate of patients with hs-AFP-L3 ≥ 7 % had a lower tendency than that of patients with hs-AFP-L3 < 7 % at HCC diagnosis ($P = 0.1501$).

Triggers to perform MRI for suspicious HCC and positivity rates for hs-AFP-L3

In this study population, US was performed median of 4 times between −1 year and diagnosis day. The 104 HCC

Table 4 Sensitivity and specificity for three years before diagnosis

Analyte	Year	Sensitivity (%)	Specificity (%)
hs-AFP-L3 ≥7 %	−1	34.3	74.7
	−2	25.3	80.6
	−3	24.5	77.0
AFP ≥20 ng/mL	−1	35.0	86.4
	−2	31.0	83.0
	−3	33.0	86.0
DCP ≥40 mAU/mL	−1	12.1	93.9
	−2	8.4	94.9
	−3	4.3	94.0

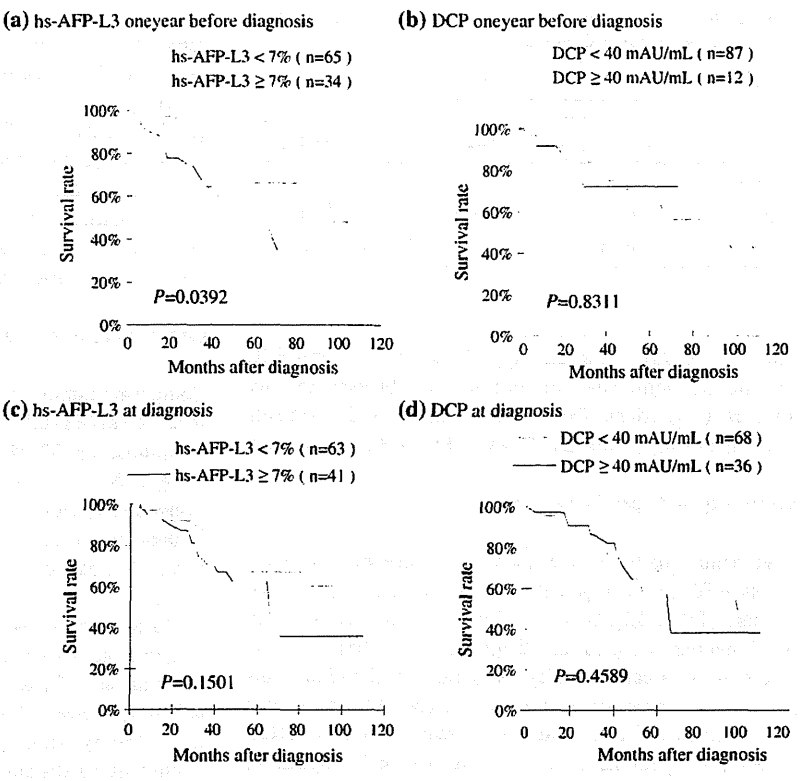


Fig. 2 Survival rates by levels of biomarkers: **a** hs-AFP-L3 and **b** DCP 1 year before, **c** hs-AFP-L3 and **d** DCP at diagnosis

patients were classified into three groups by a trigger to perform MRI for diagnostic confirmation (Table 5). US findings triggered MRI for 86 patients. The 86 patients were classified further by US findings: increase of the tumor number (51/86), increase of the tumor size (18/86), or change of the echo pattern in nodules (17/86). Five patients were monitored by MRI as results of elevated biomarkers. The remaining 13 patients were screened by MRI instead of US because interpretation of US was difficult in patients who were obese or had severe liver atrophy.

In the present retrospective study for hs-AFP-L3, 29.6 % of patients who were diagnosed with HCC by the trigger of US had hs-AFP-L3 ≥ 7 % 1 year prior to the diagnosis day. In the patients who had changes of the echo pattern in nodules, the positivity rate for hs-AFP-L3 at −1 year was 50.0 % and relatively higher compared to the other groups by US.

Table 5 Triggers to perform MRI for suspicious HCC and positivity rates for hs-AFP-L3

Triggers to perform MRI	*n*	hs-AFP-L3 >7 % At −1 year (%)	hs-AFP-L3 >7 % At diagnosis (%)
(a) Ultrasound	86	29.6	36.0
Increase of the tumor number	51	27.7	39.2
Increase of the tumor size	18	16.7	11.1
Change of the echo pattern in nodules	17	50.0	52.9
(b) Biomarkers	5	80.0	60.0
(c) Others	13	46.2	53.8

Discussion

Most studies on HCC biomarkers have focused on the accuracy at the time of diagnosis and the prediction of prognosis. So far there are a few studies which have evaluated early prediction of development of HCC in patients at high risk for HCC by biomarkers.

Taketa et al. [24] have reported that AFP-L3 values elevated above the cutoff value of 15 % with an average of 4.0 ± 4.9 months before the detection of HCC by imaging techniques. Sato et al. [25] also have demonstrated that lectin-reactive AFP elevated 3–18 months before the detection. However, only samples with AFP levels higher than 30 ng/mL were measured in their study. Recent data

indicated that the elevated AFP is not typical at HCC diagnosis for patients under in surveillance in Japan. Therefore, hs-AFP-L3 is expected to be more useful at low levels of AFP. Even though there were some differences in AFP concentration among the studies, they reported that elevation of AFP-L3 prior to diagnosis was associated with development of HCC.

Shiraki et al. [26] detected the small tumor <2 cm in maximum diameter in more than half of the patients. In the study population, they demonstrated clinical utility of lectin-reactive AFP as an early indicator while low AFP was reported limiting of the early recognition of HCC. Shimauchi et al. [27] demonstrated that AFP-L3 and DCP values showed elevated in about half of the patients at 6 months before the recognition of HCC by imaging techniques. These two markers were mutually complementary. In our study, DCP was not significantly elevated 1 year prior to diagnosis.

Lok et al. [28] have reported in a retrospective study of AFP and DCP values in patients in the Hepatitis C Antiviral Long-Term Treatment against Cirrhosis Trial who had blood drawn every 3 months for 12 months prior to HCC diagnosis. They have concluded that the biomarkers are needed to complement ultrasound in the detection of early HCC but neither DCP nor AFP is optimal. For the study, early stage HCC was defined as a single tumor nodule <3 cm in diameter with no evidence of vascular invasion or metastasis, and only 61.5 % of patients presented with early stage HCC. In our study, median of tumor size was 1.9 cm and all patients with <3 cm. Tumor volume doubling time is reported to be 90–132 days [29] and it may take a half year or 1 year for a nodule to develop from <2 cm to >3 cm. Therefore, HCC patients in our study were diagnosed 1 year earlier than the patients in Lok's study. Clinically the tumor size between <2 cm and 3 cm is one of the factor for making decisions of treatments, and it has been reported that survival rate of patients with tumor size <2 cm is higher [20]. Therefore, HCC should be diagnosed at the earlier stage with tumor <2 cm in order to achieve better outcome.

It is well known that AFP-L3 concentration correlates well with AFP; however, AFP-L3 % is not correlated with AFP [24, 30]. AFP-L3 % is a marker that is independent of AFP. Therefore, we have used AFP-L3 % for analysis.

In the present study, hs-AFP-L3 was significantly elevated 1 year prior to HCC diagnosis in 34.3 % of patients at a cutoff value of 7 %. Tamura et al. [16] reported that a cutoff value of 7 % is most appropriate for discriminating HCC from benign liver disease using this assay. Therefore, patients with elevated hs-AFP-L3 value under surveillance should be followed up closely. The specificity of 80 % or less before diagnosis may actually mislead because the non-HCC patients selected by matching with the HCC patients were potentially higher risk group for HCC and would likely develop HCC later.

In previous studies, elevated AFP-L3 has been reported to be correlated to a shorter doubling time of tumor volume, increased hepatic arterial supply, and pathologic features such as infiltrative tumor growth pattern, capsule infiltration, vascular invasion, and intrahepatic metastasis [31, 32]. These findings are often difficult to diagnose by various imaging modalities in small HCCs. Such blood supply changes typically result in change of echo pattern in nodules. In this study, therefore, high positivity rates for hs-AFP-L3 at −1 year in the patients who had such changes of echo pattern may be associated with developing HCC. The survival rate of patients with hs-AFP-L3 > 7 % at −1 year was significantly poorer compared to patients with hs-AFP-L3 < 7 %. However, differences of the detected tumor size and number were not statistically significant between patients with hs-AFP-L3 ≥ 7 % and <7 %. AFP-L3-positive HCC nodules may be aggressive and have high malignancy potential even though the tumor size is small. Therefore, it may be useful in early detection of the aggressive tumor to perform enhanced imaging techniques such as MRI for patients with elevated hs-AFP-L3. Survival rate of patients with the hs-AFP-L3 elevation at HCC diagnosis showed a poorer tendency; however, there were no statistical differences. HCC treatments were done just after the HCC diagnosis. Therefore, HCC tumors in patients with the hs-AFP-L3 elevation 1 year before HCC diagnosis might have 1 year to grow. This 1 year may reflect the difference of survival of two groups. DCP is a good marker for poor prognosis of HCC. However, the difference of overall survival between patients with DCP ≥40 and <40 mAU/mL was not observed due to the early stage (small) HCC without obvious vascular invasion.

AFP is a good marker to distinguish high-risk group for HCC development in the future [22]; however, AFP was not elevated 1 year prior to HCC development. AFP-L3 was elevated 1 year prior to diagnosis of small HCC in 34.3 % of patients.

Interpretation of US can be challenging without comparison to previous imaging results and performance of US can be limited in patients who are obese or have severe background liver cirrhosis. In the present study, sensitivity of the combined three biomarkers was 60.6 % at diagnosis, and measurements of biomarkers are expected to complement to US in surveillance.

In conclusion, elevation of hs-AFP-L3 was early predictive of development of HCC even at low AFP levels and in absence of US findings of suspicious HCC. Prognosis of patients with elevated hs-AFP-L3 was significantly poorer. HCC may be diagnosed earlier to receive curative treatments by the elevated hs-AFP-L3 as a trigger of enhanced imaging techniques. Additional prospective studies are

expected to demonstrate whether routine measurements of hs-AFP-L3 in HCC surveillance can improve overall patient survival.

Acknowledgments This work was supported by Health and Labour Sciences Research Grants (Research on Hepatitis) from the Ministry of Health, Labour and Welfare of Japan.

Conflict of interest All authors declare that the authors report no conflicts of interest.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127:2893–917.
2. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. J Cancer Res Clin Oncol. 2004;130:417–22.
3. McMahon BJ, Bulkow L, Harpster A, Snowball M, Lanier A, Sacco F, et al. Screening for hepatocellular carcinoma in Alaska natives infected with chronic hepatitis B: a 16-year population-based study. Hepatology. 2000;32:842–6.
4. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology. 2011;53:1020–2.
5. Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. J Hepatol. 2001;35:421–30.
6. Omata M, Lesmana LA, Tateishi R, Chen PJ, Lin SM, Yoshida H, et al. Asian pacific association for the study of the liver consensus recommendations on hepatocellular carcinoma. Hepatol Int. 2010;4:439–74.
7. Makuuchi M, Kokudo N, Arii S, Futagawa S, Kaneko S, Kawasaki S, et al. Development of evidence-based clinical guidelines for the diagnosis and treatment of hepatocellular carcinoma in Japan. Hepatol Res. 2008;38:37–51.
8. Di Bisceglie AM, Hoofnagle JH. Elevations in serum alpha-fetoprotein levels in patients with chronic hepatitis B. Cancer. 1989;64:2117–20.
9. Liaw YF, Tai DI, Chen TJ, Chu CM, Huang MJ. Alpha-fetoprotein changes in the course of chronic hepatitis: relation to bridging hepatic necrosis and hepatocellular carcinoma. Liver. 1986;6:133–7.
10. Chu CW, Hwang SJ, Luo JC, Lai CR, Tsay SH, Li CP, et al. Clinical, virologic, and pathologic significance of elevated serum alpha-fetoprotein levels in patients with chronic hepatitis C. J Clin Gastroenterol. 2001;32:240–4.
11. Aoyagi Y, Isemura M, Suzuki Y, Sekine C, Soga K, Ozaki T, et al. Fucosylated alpha-fetoprotein as marker of early hepatocellular carcinoma. Lancet. 1985;2:1353–4.
12. Taketa K, Sekiya C, Namiki M, Akamatsu K, Ohta Y, Endo Y, et al. Lectin-reactive profiles of alpha-fetoprotein characterizing hepatocellular carcinoma and related conditions. Gastroenterology. 1990;99:508–18.
13. Oka H, Saito A, Ito K, Kumada T, Satomura S, Kasugai H, et al. Multicenter prospective analysis of newly diagnosed hepatocellular carcinoma with respect to the percentage of *Lens culinaris* agglutinin-reactive alpha-fetoprotein. J Gastroenterol Hepatol. 2001;16:1378–83.
14. Yamagata Y, Katoh H, Nakamura K, Tanaka T, Satomura S, Matsuura S. Determination of alpha-fetoprotein concentration based on liquid-phase binding assay using anion exchange chromatography and sulfated peptide introduced antibody. J Immunol Methods. 1998;212:161–8.
15. 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 electrophoresis. Anal Biochem. 2009;388:306–11.
16. Tamura Y, Igarashi M, Kawai H, Suda T, Satomura S, Aoyagi Y. Clinical advantage of highly sensitive on-chip immunoassay for fucosylated fraction of alpha-fetoprotein in patients with hepatocellular carcinoma. Dig Dis Sci. 2010;55:2095–101.
17. Hanaoka T, Sato S, Tobita H, Miyake T, Ishihara S, Akagi S, et al. Clinical significance of the highly sensitive fucosylated fraction of alpha-fetoprotein in patients with chronic liver disease. J Gastroenterol Hepatol. 2011;26:739–44.
18. 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.
19. Ikai I, Kudo M, Arii S, Omata M, Kojiro M, Sakamoto M, et al. Report of the 18th follow-up survey of primary liver cancer in Japan. Hepatol Res. 2010;40:1043–59.
20. Takayama T, Makuuchi M, Hirohashi S, Sakamoto M, Yamamoto J, Shimada K, et al. Early hepatocellular carcinoma as an entity with a high rate of surgical cure. Hepatology. 1998;28:1241–6.
21. Kumada T, Toyoda H, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y, et al. Incidence of hepatocellular carcinoma in hepatitis C carriers with normal alanine aminotransferase levels. J Hepatol. 2009;50:729–35.
22. 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.
23. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2000.
24. Taketa K, Endo Y, Sekiya C, Tanikawa K, Koji T, Taga H, et al. A collaborative study for the evaluation of lectin-reactive a-fetoproteins in early detection of hepatocellular carcinoma. Cancer Res. 1993;53:5419–23.
25. Sato Y, Nakata K, Kato Y, Shima M, Ishii N, Koji T, et al. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. N Engl J Med. 1993;328:1802–6.
26. Shiraki K, Takase K, Tameda Y, Hamada M, Kosaka Y, Nakano T. A clinical study of lectin-reactive alpha-fetoprotein as an early indicator of hepatocellular carcinoma in the follow-up of cirrhotic patients. Hepatology. 1995;22:802–7 Oncol Rep. 2000; 7:249–256.
27. Shimauchi Y, Tanaka M, Kuromatsu R, Ogata R, Tateishi Y, Itano S, et al. A simultaneous monitoring of *Lens culinaris* agglutinin A-reactive alpha-fetoprotein and des-gamma-carboxy prothrombin as an early diagnosis of hepatocellular carcinoma in the follow-up of cirrhotic patients. Oncol Rep. 2000;7:249–56.
28. 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.
29. Okazaki N, Yoshino M, Yoshida T, Suzuki M, Moriyama N, Takayasu K, et al. Evaluation of the prognosis for small hepatocellular carcinoma based on tumor volume doubling time. A preliminary report. Cancer. 1989;63:2207–10.

30. Taketa K. Alpha-fetoprotein: reevaluation in hepatology. Hepatology. 1990;12:1420–32.
31. Kumada T, Nakano S, Takeda I, Kiriyama S, Sone Y, Hayashi K, et al. Clinical utility of *Lens culinaris* agglutinin-reactive alpha-fetoprotein in small hepatocellular carcinoma: special reference to imaging diagnosis. J Hepatol. 1999;30:125–30.
32. Tada T, Kumada T, Toyoda H, Kiriyama S, Sone Y, Tanikawa M, et al. Relationship between Lens culinaris agglutinin-reactive alpha-fetoprotein and pathologic features of hepatocellular carcinoma. Liver Int. 2005;25:848–53.

岐阜県医師会医学雑誌　第27巻

第34回岐阜県医師会メディカルセミナー　―消化器疾患のトピックス―

C型肝炎は克服できる時代に　―最新の抗ウイルス治療―

岐阜県総合医療センター
杉原潤一

２０１４年５月
岐　阜　県　医　師　会

岐阜県医師会医学雑誌　第27巻

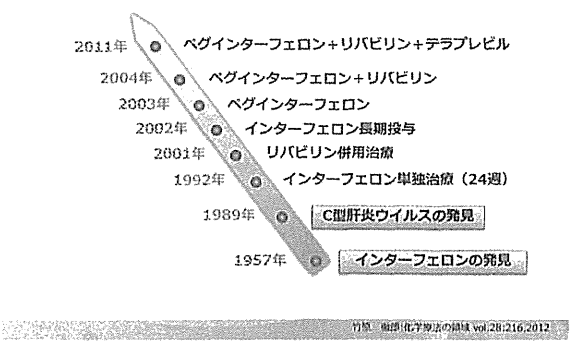
第34回岐阜県医師会メディカルセミナー　―消化器疾患のトピックス―
2013年1月20日

C型肝炎は克服できる時代に　―最新の抗ウイルス治療―

岐阜県総合医療センター
杉原潤一

C型肝炎に対しては，1992年よりC型肝炎ウイルス（HCV）の完全排除を期待しうる治療としてインターフェロン（IFN）療法が開始された．そのなかで，ペグインターフェロン（Peg-IFN）単独治療，Peg-IFN＋リバビリン併用治療，IFN自己注射などが登場し，セロタイプ2型の高ウイルス量症例においては約90%が，また従来難治性とされていたセロタイプ1型の高ウイルス量症例でも約50～60%が完治できるようになってきた．そして2年前からは新しい抗ウイルス治療であるPeg-IFN＋リバビリン＋テラプレビル3剤併用治療が可能となった（図1）．さらに本年になり，副作用の少ないPeg-IFN＋リバビリン＋シメプレビル3剤併用治療が可能となり，副作用の軽減とともにその治療成績（ウイルス排除率）の向上が期待されている．そこで当センターの治療成績や日本での各種治験の成績を中心に，C型肝炎に対する抗ウイルス治療の変遷と治療成績の進歩について述べてみたい．

図1
C型慢性肝炎に対する抗ウイルス治療の変遷



1．ペグインターフェロン＋リバビリン2剤併用治療

2004年から，週1回ですむペグインターフェロン（Peg-IFN）とリバビリンの2剤併用治療が開始された．これにより従来のインターフェロン単独治療（24週）に比して，飛躍的に治療成績が向上した．当時の治療ガイドライン（表1）によれば，Peg-IFN＋リバビリン2剤併用治療の適応は，セロタイプ2型，高ウイルス量症例，および今まで最も難治であったセロタイプ1型，高ウイルス量症例である．

1）セロタイプ2型，高ウイルス量症例に対する治療成績

①ウイルス陰性化時期に関与する因子

治療開始後のウイルス陰性化時期が評価可能な96例の陰性化時期は，4週以内56例，8週37例，12週3例であった．4週以内陰性化（EVR）群と8週以降陰性化（LVR）群を比較すると，年齢はそれぞれ48.5歳，57.3歳，性別（男性：女性）はそれぞれ29：27，16：24であり，LVR

表1
C型慢性肝炎に対する初回治療ガイドライン

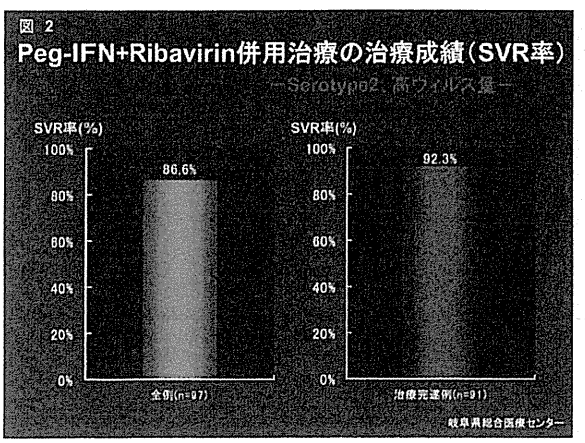
	Genotype 1	Genotype 2
高ウイルス量 1 Meq/mL 5.0 Log IU/mL 300 fmol/L以上	Peg-IFNα2b:Peg-Intron＋Ribavirin:Rebetol（48～72週間） Peg-IFNα2a:Pegasys＋Ribavirin:Copegus（48～72週間） IFNβ:Feron＋Ribavirin:Rebetol(48-72週間)	Peg-IFNα2b:Peg-Intron＋Ribavirin:Rebetol（24週間） IFNβ:Feron＋Ribavirin:Rebetol(24週間)
低ウイルス量 1 Meq/mL 5.0 Log IU/mL 300 fmol/L未満	IFN(24週間) Peg-IFNα2a:Pegasys (24-48週間)	IFN(8-24週間) Peg-IFNα2a:Pegasys (24-48週間)

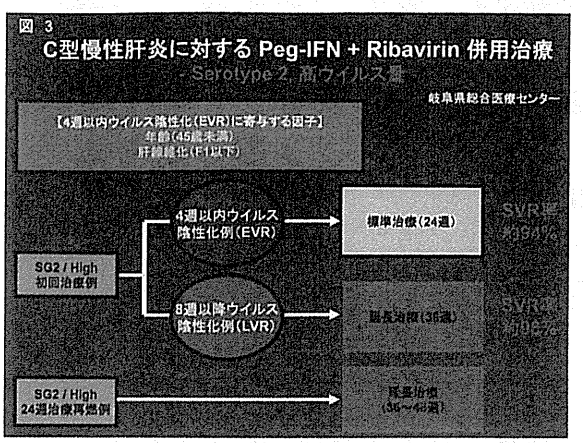
群では年齢が有意に高く，女性が多い傾向がみられたが，ウイルス量やGenotype，血液生化学所見には差はみられなかった．多変量解析（治療前因子）では，EVRすなわちウイルスの早期陰性化に最も関与している因子は，年齢（45歳未満）であった．

②治療成績

中止例などを除いた治療効果判定可能な治療完遂91例について検討すると，標準治療（24週）は55例，延長治療（32～36週）は36例で，治療完遂例のSVR率（ウイルス駆除率）は92.3%と極めて高率であった（図2）．EVR群全体のSVR率は94.5%であり，うち標準治療のSVR率は94.1%，延長治療のSVR率は100%であった．一方LVR群では，標準治療のSVR率は50.0%であったのに対し，延長治療のSVR率は93.5%と極めて高率であった．延長治療のなかでも32週治療のSVR率は75.0%，36週治療のSVR率は96.4%であり，とくに36週延長治療例のSVR率（96.4%）は標準治療例のSVR率（50.0%）に比して有意に高率であった．また初回治療にて8週でウイルスが陰性化し，24週の標準治療でウイルスが再燃した69歳男性症例は，再治療でも8週陰性化であったが36週の延長治療によりSVRが得られている．

このようにセロタイプ2，高ウイルス量症例に対するPeg-IFN＋リバビリン併用治療においては，4週以内陰性化（EVR）が得られれば24週の標準治療で極めて良好なSVR率が得られる．一方，8週以降陰性化（LVR）例については，24週の標準治療では海外の成績や国内の学会発表の成績をみてもそのSVR率はおおむね50～60%にとどまっている．従って高いSVR率を得るためには，8週以降陰性化（LVR）例に対しては36週の延長治療が望ましいと考えられる．また再治療症例に対しても36週の延長治療は有用と思われる（図3）．





2）セロタイプ1型，高ウイルス量症例に対する治療成績

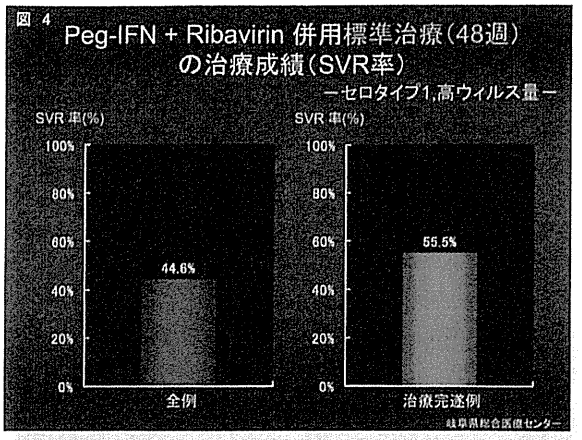
①ウイルス陰性化時期

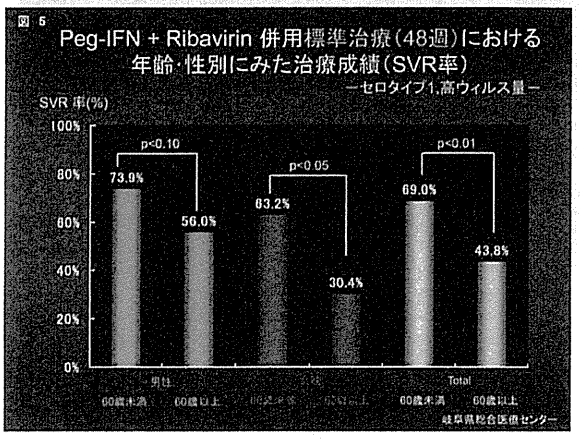
治療開始後の累積ウイルス陰性化率は，4週後12.2%，8週後30.5%，12週後46.3%，16週後61.0%，20週後70.7%，24週後76.8%であった．8週以内の早期ウイルス陰性化に関与する因子（多変量解析）は，2週後のウイルス減少率，治療前ウイルス量，性別であった．さらに12～20週陰性化に関与する因子は4週後のウイルス減少率，20週までのリバビリン服薬率，治療開始時のALT値であった．

②治療成績

・標準治療（48週）

治療全例のSVR率は44.6%であり，治療完遂例のSVR率は55.5%であった（図4）．年齢別にSVR率をみると，60歳未満では69.0%であるのに対し，60歳以上では43.8%と有意に低率であった．さらに性別でSVR率を検討すると，男性では60歳未満73.9%，60歳以上56.0%，女性では60歳未満63.2%，60歳以上では30.4%であり，とくに高齢女性





においてSVR率が低率であった（図５）．ウイルス陰性化時期別に標準治療のSVR率をみると，８週以内の早期ウイルス陰性化例においてはSVR率は91.3%と極めて高率であったが，12～20週陰性化例では57.1%，24週以降では０%であり，SVR率はやはりウイルス陰性化時期と密接な関連がみられる．

・延長治療（72週）

治療開始後のウイルス陰性化時期が12週以降の症例を中心として，72週の延長治療を施行した．年齢別にSVR率をみると，60歳未満では60.5%，60歳以上では50.0%であり，高齢でもSVR率が改善した．さらに性別でSVR率を検討すると，男性では60歳未満69.2%，60歳以上50.0%，女性では60歳未満56.0%，60歳以上では50.0%であり，高齢女性においても標準治療に比してSVR率が改善した．またウイルス陰性化時期別に延長治療のSVR率をみると，12～20週以内ウイルス陰性化例では68.0%，24～36週陰性化例では23.1%，40週以降陰性例では０%であり，延長治療ではとくにウイルス陰性化時期が12～36週の症例において標準治療に比してSVR率の向上が認められた（図６）．このようにセロタイプ１，高ウイルス量症例では，８週以内の早期ウイルス陰性化例は標準治療で極めて良好なSVR率（約90%）が得られ，また12～20週陰性化例では延長治療（72週）によってSVR率の向上（約70%）が認められた．しかしながら24～36週陰性化例では延長治療を行ってもSVR率は低率（約25%）にとどまり，40～48週陰性化例では延長治療でもSVRは得られなかった（図７）．

３）IFN治療効果に影響を及ぼす因子

IFN治療効果に影響を及ぼす因子としては，すでに述べたように患者側因子（年齢，性），ウイ