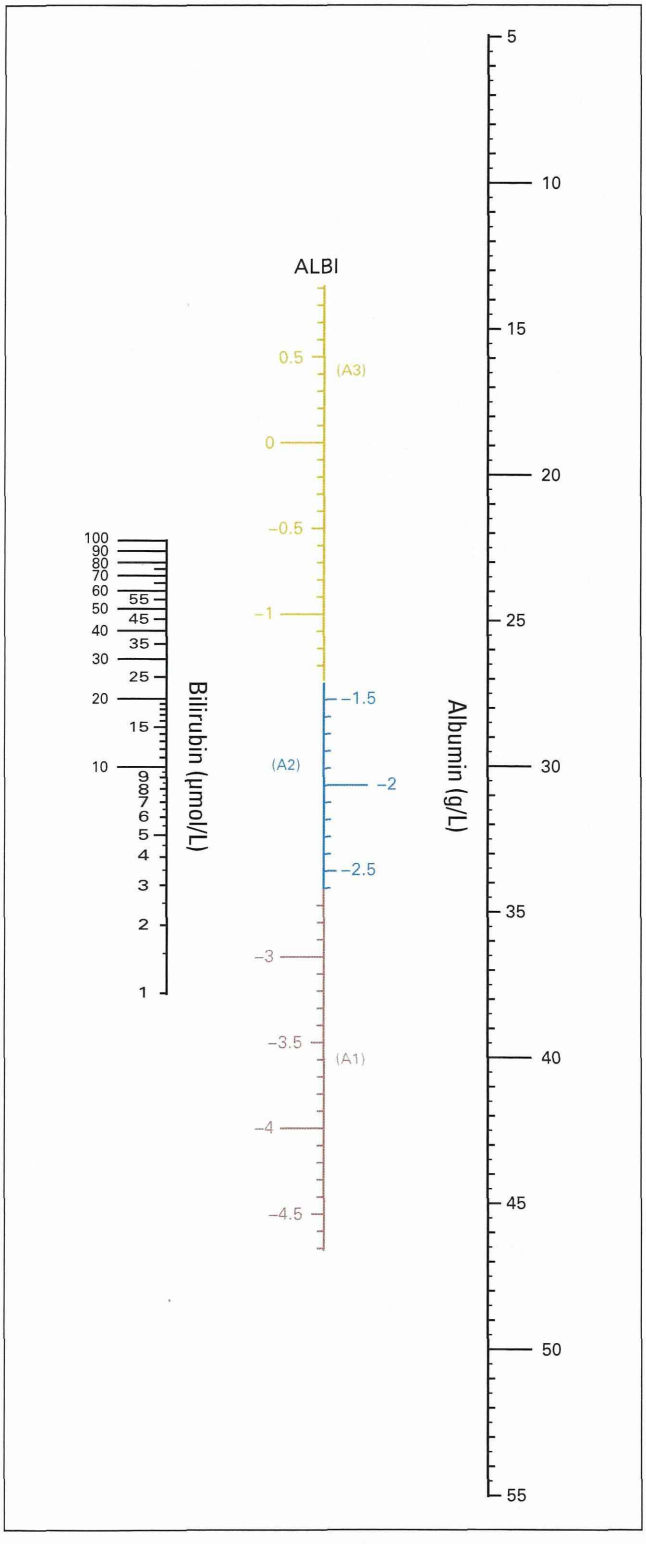

Fig 4. Nomogram for rapid assessment of the Albumin-Bilirubin (ALBI) score. Colors refer to ALBI grades A1, A2, and A3.

regions. In Europe and the United States, for example, when C-P grade A patients were reclassified into ALBI grade 1 or 2, there was a 10-month difference in survival between the two ALBI grades. Our analysis has focused on the impact of liver function on survival, and not on liver disease–related events or deaths, because in practice, it is difficult to specifically attribute the cause of death to the HCC or the underlying liver disease.

Assessment of liver function is particularly important in clinical trials because it is perceived that cirrhosis is a competing cause of death. To isolate the impact of a specific HCC treatment on survival, many HCC treatment studies are limited to patients with C-P grade A. However, our findings suggest that not all C-P grade A patients are the same and that this heterogeneity may have an impact on survival findings. Among C-P grade A patients in clinical trials who received the standard care, sorafenib, the model distinguishes between a good risk group (ALBI grade 1) and a relatively poorer risk group (ALBI grade 2), with a median survival difference of nearly 6 months. Such refinement of liver function assessment might permit retrospective assessment of sorafenib efficacy and survival in these subgroups and determine the most appropriate group for this type of treatment in the future.

One of the strengths of this study is the large number of patients and the generalizability of the results because we have considered high-, medium-, and low-incidence HCC areas and a broad spectrum of etiologies. We specifically excluded patients who underwent liver transplantation because, in these patients, underlying (dys)function is effectively abrogated by the procedure.

The fact that both serum bilirubin and albumin are part of the battery of tests widely referred to as liver function tests suggests that our model is, indeed, measuring liver function, a contention that is supported by our demonstration that the model showed discrimination in patients with uncomplicated cirrhosis. We have examined this group specifically for the aforementioned reason, not to suggest that it would have a role outside the area of HCC and chronic liver disease. Interestingly, in a systematic review of prognostic indicators in cirrhosis, serum albumin and bilirubin were the two most prominent individual prognostic variables in good studies.[23]

It might seem surprising that survival in particular ALBI stages varied across geographical regions, but overall median survival is different across geographical regions and the variation is just as striking in the case of the C-P grades. This may be partly attributable to lead-time bias because it is likely that some countries where extensive public health measures have been implemented, such as Japan, will have apparent survival times greater than those seen in countries such as Hong Kong, where there is less access to primary care. Using sophisticated approaches, our model could be recalibrated for each region, as we have previously shown.[24]

We have previously reported and validated[13,24] an objective serology-based model for survival prediction in HCC. This model, known as BALAD-2, combines bilirubin and albumin with three serum biomarkers (α-fetoprotein [AFP], AFP-L3%, and DCP). This study adds to the plausibility of the BALAD-2 model because bilirubin and albumin seem to represent the impact of the underlying liver function on survival (as shown here), whereas the three biomarkers may represent the impact of the tumor itself on survival.

The Model for End-Stage Liver Disease score might be considered an alternative to C-P grade. However, this system is specifically designed for patients with end-stage cirrhosis,[25-28] and as shown here,

this is not the case in most patients with HCC. Furthermore, serum creatinine is one of the parameters in the Model for End-Stage Liver Disease score and may be less reliable in patients with cancer because of cancer-related cachexia (creatinine levels are related to muscle mass).[29] Although the performance of ALBI is similar to that of C-P, the fact that it is evidence based, much simpler (using the proposed nomogram or heat map), and entirely objective will make it easier to implement. Furthermore, relying on fewer variables, it may be more readily applicable in large-scale international studies. For example, in a recent US study based on the GIDEON registry (a global, prospective, noninterventional study of patients with unresectable HCC undergoing sorafenib treatment),[30] 27% of patients were not evaluable for C-P assessment, largely because of missing international normalized ratio values (a constituent of the C-P grade). A limitation of this study is that we did not have access to the C-P score, only the C-P grade, as classified by investigators at the individual centers. Hence, we can draw no conclusions as to the performance of the ALBI system in relation to specific C-P scores, for example, C-P score 5 to 6.

The ALBI approach will avoid interobserver variation and may highlight distinct prognostic subgroups within C-P grade A. All these advantages are important considerations in the clinical trial setting.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Philip J. Johnson
Provision of study materials or patients: Philip J. Johnson, Sarah Berhane, Chiaki Kagebayashi, Mabel Teng, Helen L. Reeves, James O'Beirne, Anna Skowronska, Winnie Yeo, Frankie Mo, Paul Lai, Mercedes Iñarrairaegui, Stephen L. Chan, Bruno Sangro, Rebecca Miksad, Toshifumi Tada, Takashi Kumada, Hidenori Toyoda
Collection and assembly of data: Philip J. Johnson, Sarah Berhane, Chiaki Kagebayashi, Mabel Teng, Helen L. Reeves, James O'Beirne, Anna Skowronska, Daniel Palmer, Winnie Yeo, Frankie Mo, Paul Lai, Mercedes Iñarrairaegui, Stephen L. Chan, Bruno Sangro, Rebecca Miksad, Toshifumi Tada, Takashi Kumada, Hidenori Toyoda
Data analysis and interpretation: Philip J. Johnson, Sarah Berhane, Shinji Satomura, Richard Fox, Daniel Palmer, Rebecca Miksad
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Simonetti RG, CammàC, Fiorello F, et al: Hepatocellular carcinoma. A worldwide problem and the major risk factors. Dig Dis Sci 36:962-972, 1991
2. Fattovich G, Pantalena M, Zagni I, et al: Effect of hepatitis B and C virus infections on the natural history of compensated cirrhosis: A cohort study of 297 patients. Am J Gastroenterol 97:2886-2895, 2002
3. Child CG, Turcotte JG: Surgery and portal hypertension. Major Probl Clin Surg 1:1-85, 1964
4. Pugh RN, Murray-Lyon IM, Dawson JL, et al: Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 60:646-649, 1973
5. van Dam GM, Gips CH, Reisman Y, et al: Major clinical events, signs and severity assessment scores related to actual survival in patients who died from primary biliary cirrhosis: A long-term historical cohort study. Hepatogastroenterology 46:108-115, 1999
6. Shetty K, Rybicki L, Carey WD: The Child-Pugh classification as a prognostic indicator for survival in primary sclerosing cholangitis. Hepatology 25:1049-1053, 1997
7. Durand F, Valla D: Assessment of prognosis of cirrhosis. Semin Liver Dis 28:110-122, 2008
8. Johnson PJ, Williams R: Cirrhosis and the aetiology of hepatocellular carcinoma. J Hepatol 4:140-147, 1987
9. Johnson PJ, Berhane S, Satomura S, et al: An international collaborative study assessing the role of aetiology and stage in survival in HCC: Implications for screening. Presented at International Liver Congress London, European Association for the Study of the Liver, London, United Kingdom, April 9-13, 2014
10. Llovet JM, Brú C, Bruix J: Prognosis of hepatocellular carcinoma: The BCLC staging classification. Semin Liver Dis 19:329-338, 1999
11. Del Priore G, Zandieh P, Lee MJ: Treatment of continuous data as categoric variables in obstetrics and gynecology. Obstet Gynecol 89:351-354, 1997
12. Royston P, Altman DG, Sauerbrei W: Dichotomizing continuous predictors in multiple regression: A bad idea. Stat Med 25:127-141, 2006
13. Toyoda H, Kumada T, Osaki Y, et al: Staging hepatocellular carcinoma by a novel scoring system (BALAD score) based on serum markers. Clin Gastroenterol Hepatol 4:1528-1536, 2006
14. Toyoda H, Kumada T, Kiriyama S, et al: Impact of surveillance on survival of patients with initial hepatocellular carcinoma: A study from Japan. Clin Gastroenterol Hepatol 4:1170-1176, 2006
15. Chan SL, Mo FK, Johnson PJ, et al: 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, 2011
16. Dyson J, Jaques B, Chattopadyhay D, et al: Hepatocellular cancer: The impact of obesity, type 2 diabetes and a multidisciplinary team. J Hepatol 60:110-117, 2014
17. Johnson PJ, Qin S, Park JW, et al: Brivanib versus sorafenib as first-line therapy in patients with unresectable, advanced hepatocellular carcinoma: Results from the randomized phase III BRISK-FL study. J Clin Oncol 31:3517-3524, 2013
18. Cheng AL, Kang YK, Lin DY, et al: Sunitinib versus sorafenib in advanced hepatocellular cancer: Results of a randomized phase III trial. J Clin Oncol 31:4067-4075, 2013
19. Liver Cancer Study Group of Japan: The General Rules for the Clinical and Pathological Study of Primary Liver Cancer (English ed). Tokyo, Japan, Kanehara & Co, 2003
20. Taktak A, Eleuteri A, Lake S, et al: Evaluation of prognostic models: Discrimination and calibration performance. Comput Intell Med 2007
21. Newson R: Confidence intervals for rank statistics: Somers' D and extensions. Stata J 6:309, 2006
22. Newson RB: Comparing the predictive powers of survival models using Harrell's C or Somers' D. Stata J 10:339, 2010
23. D'Amico G, Garcia-Tsao G, Pagliaro L: Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies. J Hepatol 44:217-231, 2006
24. Fox R, Berhane S, Teng M, et al: Biomarker-based prognosis in hepatocellular carcinoma: Validation and extension of the BALAD model. Br J Cancer 110:2090-2098, 2014
25. Durand F, Valla D: Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. J Hepatol 42:S100-S107, 2005 (suppl)
26. Botta F, Giannini E, Romagnoli P, et al: 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, 2003
27. Bruix J, Sherman M, American Association for the Study of Liver Diseases: Management of hepatocellular carcinoma: An update. Hepatology 53: 1020-1022, 2011
28. Bruix J, Sherman M, Practice Guidelines Committee, American Association for the Study of Liver Diseases: Management of hepatocellular carcinoma. Hepatology 42:1208-1236, 2005
29. Nankivell BJ: Creatinine clearance and the assessment of renal function. Aust Prescr 24:15-17, 2001
30. Miksad RA, Cohn AL, El-Khoueiry AB, et al: Use of staging and scoring systems in hepatocellular carcinoma (HCC): Lessons from U.S. regional analysis of the GIDEON registry. J Clin Oncol 32, 2014 (suppl 3; abstr 323)

Affiliations

Philip J. Johnson, Sarah Berhane, and Daniel Palmer, University of Liverpool, Liverpool; Philip J. Johnson and Daniel Palmer, The Clatterbridge Cancer Centre National Health Service (NHS) Foundation Trust, Bebington; Mabel Teng, Addenbrooke's Hospital, University of Cambridge, Cambridge; Helen L. Reeves, Northern Institute for Cancer Research and the Hepatopancreatobiliary Multidisciplinary Team,

Newcastle upon Tyne NHS Foundation Trust, The Freeman Hospital, Newcastle upon Tyne; James O'Beirne, The Sheila Sherlock Liver Centre, Royal Free Hospital, London; Richard Fox and Anna Skowronska, School of Cancer Sciences, University of Birmingham, Birmingham, United Kingdom; Chiaki Kagebayashi and Shinji Satomura, Wako Life Sciences, Mountain View, CA; Rebecca Miksad, Beth Israel Deaconess Medical Center, Institute for Technology Assessment, Massachusetts General Hospital, and Harvard Medical School, Boston, MA; Winnie Yeo, Frankie Mo, and Stephen L. Chan, State Key Laboratory in Oncology in South China, Sir Y. K. Pao Centre for Cancer, Chinese University of Hong Kong, Hong Kong Cancer Institute; Paul Lai, Prince of Wales Hospital, Chinese University of Hong Kong, Hong Kong, China; Mercedes Iñarrairaegui and Bruno Sangro, Clinica Universidad de Navarra, and Centro de Investigacion Biomedica en Red de Enfermedades Hepaticas y Digestivas, Pamplona, Spain; Toshifumi Tada, Takashi Kumada, and Hidenori Toyoda, Ogaki Municipal Hospital, Ogaki, Gifu, Japan.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Assessment of Liver Function in Patients With Hepatocellular Carcinoma: A New Evidence-Based Approach—The ALBI Grade

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Philip J. Johnson
Honoraria: Bayer
Consulting or Advisory Role: Astellas
Travel, Accommodations, Expenses: Bayer, Wako Life Sciences

Sarah Berhane
No relationship to disclose

Chiaki Kagebayashi
Employment: Wako Life Sciences

Shinji Satomura
Employment: Wako Life Sciences

Mabel Teng
No relationship to disclose

Helen L. Reeves
No relationship to disclose

James O'Beirne
Speakers' Bureau: Gilead Sciences
Travel, Accommodations, Expenses: Bristol-Myers Squibb

Richard Fox
No relationship to disclose

Anna Skowronska
No relationship to disclose

Daniel Palmer
Honoraria: Bayer, Celgene, Astellas Pharma
Consulting or Advisory Role: Bayer, Celgene, Astellas Pharma

Winnie Yeo
Honoraria: Roche
Consulting or Advisory Role: Novartis, GlaxoSmithKline, Boehringer Ingelheim, Eisai, Pfizer

Frankie Mo
No relationship to disclose

Paul Lai
Consulting or Advisory Role: Novartis

Mercedes Iñarrairaegui
No relationship to disclose

Stephen L. Chan
No relationship to disclose

Bruno Sangro
Consulting or Advisory Role: Bayer Schering Pharma, Sirtex Medical, MedImmune, Bristol-Myers Squibb, Transgene, Guerbet
Speakers' Bureau: Bayer Schering Pharma, Sirtex Medical
Travel, Accommodations, Expenses: Bayer Schering Pharma, Sirtex Medical, Gilead Sciences

Rebecca Miksad
Consulting or Advisory Role: Optum Insight, Acceleron Pharma, Onyx
Research Funding: Bayer/Onyx (Inst), NewLink (Inst), Genentech (Inst), Daiichi Sankyo (Inst), ArQule (Inst), Acceleron Pharma (Inst)

Toshifumi Tada
No relationship to disclose

Takashi Kumada
No relationship to disclose

Hidenori Toyoda
No relationship to disclose

Acknowledgment

We gratefully acknowledge the help of all those who contributed to the collection of the data sets used in this study. From Hong Kong, Jane Koh played a major role in data collection and tabulation, and similar roles were fulfilled by Sara Carrillo and Janet Morse in Pamplona and Birmingham, respectively. The US data abstraction team included Atinuke Babalola, Ugonna Nwosu, Olubunmi Oladunjoye, and Bethany Burns. We also thank all those who played roles in managing the patients in this study, including the Newcastle upon Tyne Hospitals National Health Service Foundation Trust and the Queen Elizabeth Hospital University Trust multidisciplinary teams. Additional support for data collection and analysis came from the Biomedical Research Unit at the University of Birmingham and Liverpool Health Partners. Finally, we are indebted to Pfizer and Bristol-Myers Squibb who generously gave us access to data from studies in which they acted as sponsor.

Appendix

Table A1. Univariable Analysis of Whole Japan Cohort

Variable	HR	SE	*z*	$P > z$	95% CI of HR
Sex (male)	1.21	0.085	2.70	.007	1.05 to 1.39
Age	1.00	0.0036	0.17	.865	0.99 to 1.0077
Log_{10} bilirubin	4.91	0.59	13.25	< .001	3.88 to 6.21
INR	1.09	0.055	1.67	.094	0.99 to 1.20
Albumin	0.91	0.0052	−17.39	< .001	0.90 to 0.92
Tumor size, cm	1.14	0.0091	16.53	< .001	1.12 to 1.16
Tumor number	1.32	0.029	12.68	< .001	1.26 to 1.37
Macroscopic vascular invasion	5.40	0.41	22.38	< .001	4.66 to 6.27
TNM stage					
I	1				
II	1.53	0.15	4.26	< .001	1.26 to 1.86
III	2.70	0.27	9.86	< .001	2.22 to 3.29
IV	10.16	1.09	21.55	< .001	8.23 to 12.55

Abbreviations: HR, hazard ratio; INR, international normalized ratio.

Table A2. Multivariable Cox Regression Using Stepwise Forward Selection of Variables (entire Japanese cohort)

Variable	HR	SE	z	P > z	95% CI
Tumor size < 3 cm					
Macroscopic vascular invasion	3.75	1.36	3.64	< .001	1.84 to 7.63
Albumin, g/L	0.90	0.0094	−10.04	< .001	0.88 to 0.92
Tumor number	1.30	0.058	5.89	< .001	1.19 to 1.42
Log_{10} bilirubin	2.74	0.61	4.53	< .001	1.77 to 4.24
Age	1.02	0.0065	3.41	.001	1.01 to 1.03
Sex (male)	1.45	0.16	3.35	.001	1.17 to 1.80
Tumor size 3-5 cm					
Macroscopic vascular invasion	3.43	0.84	5.04	< .001	2.12 to 5.55
Albumin, g/L	0.91	0.013	−6.92	< .001	0.88 to 0.93
Tumor number	1.21	0.057	4.08	< .001	1.10 to 1.33
Log_{10} bilirubin	2.32	0.77	2.55	.011	1.21 to 4.44
Age	1.02	0.0094	2.37	.018	1.0038 to 1.04
Tumor size 5.1-10 cm					
Macroscopic vascular invasion	2.75	0.49	5.67	< .001	1.94 to 3.91
Albumin, g/L	0.96	0.016	−2.63	.009	0.92 to 0.99
Log_{10} bilirubin	2.76	0.92	3.06	.002	1.44 to 5.30
Tumor number	1.13	0.045	2.98	.003	1.04 to 1.22
Tumor size > 10 cm					
Albumin, g/L	0.91	0.022	−4.07	< .001	0.87 to 0.95
Log_{10} bilirubin	10.25	5.52	4.32	< .001	3.56 to 29.48
Stage I					
Log_{10} bilirubin	3.67	1.43	3.34	.001	1.71 to 7.89
Albumin, g/L	0.92	0.016	−5.16	< .001	0.89 to 0.95
Stage II					
Log_{10} bilirubin	2.30	0.51	3.71	< .001	1.48 to 3.56
Albumin, g/L	0.90	0.0097	−9.79	< .001	0.88 to 0.92
Age	1.02	0.0072	2.97	.003	1.01 to 1.04
Sex (male)	1.32	0.16	2.22	.026	1.03 to 1.68
Stage III					
Log_{10} bilirubin	1.76	0.45	2.21	.027	1.07 to 2.91
Albumin, g/L	0.91	0.011	−7.58	< .001	0.89 to 0.94
Stage IV					
Log_{10} bilirubin	3.13	0.76	4.72	< .001	1.95 to 5.02
Albumin, g/L	0.95	0.012	−4.44	< .001	0.92 to 0.97

Abbreviation: HR, hazard ratio.

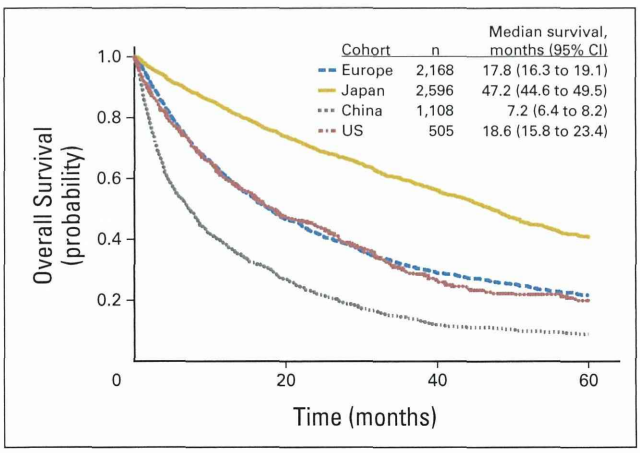

Fig A1. Kaplan-Meier curves showing survival in the Japanese, European, Chinese, and US cohorts.

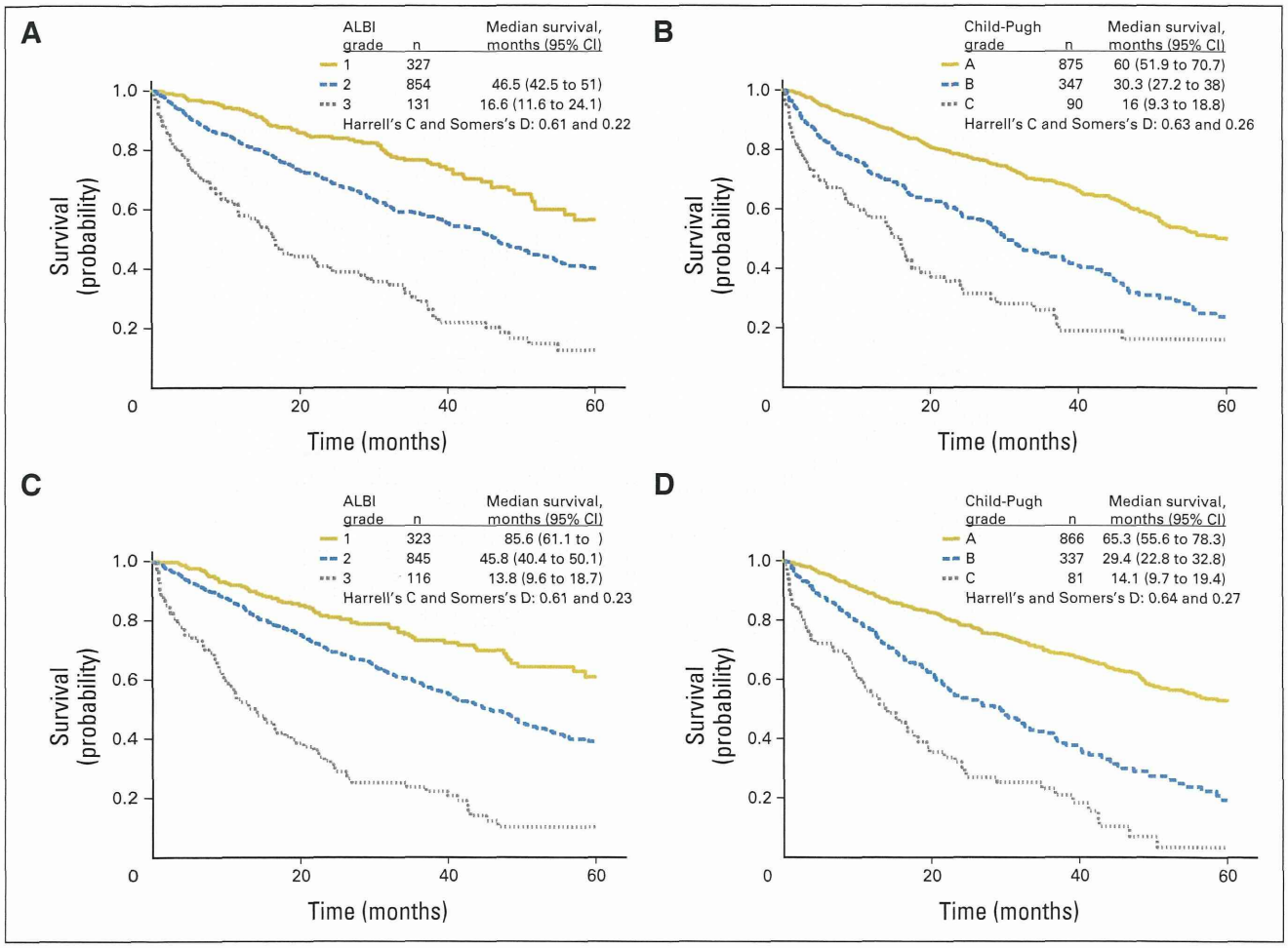

Fig A2. Application of the Albumin-Bilirubin (ALBI) model to the training and validation sets. Kaplan-Meier curves depicting survival according to ALBI and Child-Pugh (C-P) grade in the (A and B) Japanese training set and (C and D) validation set.

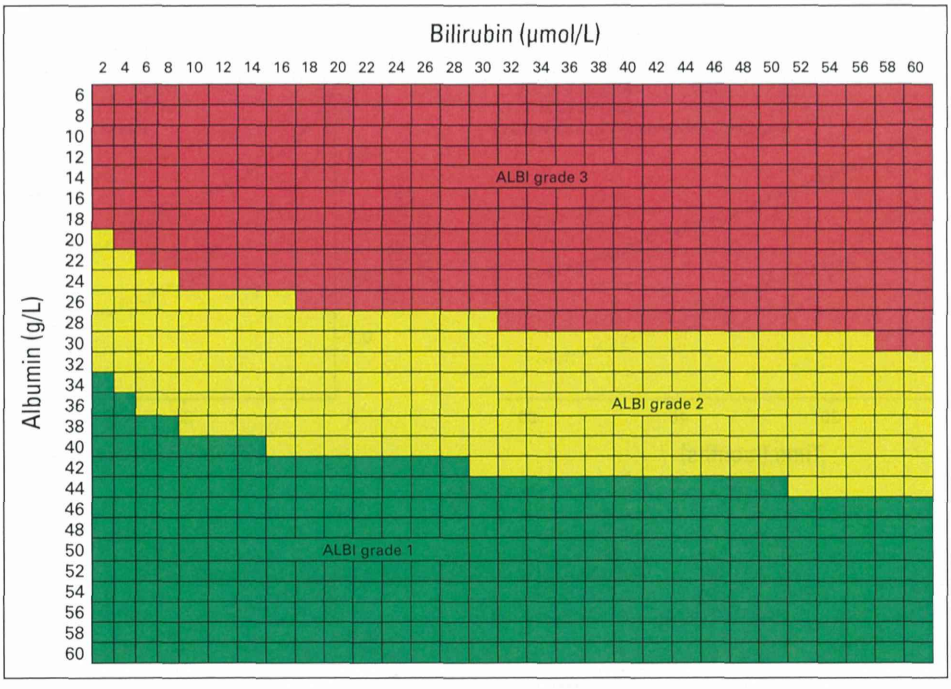

Fig A3. Heat map for rapid assessment of the Albumin-Bilirubin (ALBI) grade.

Epidemiol. Infect., Page 1 of 5. © Cambridge University Press 2014
doi:10.1017/S0950268814000478

SHORT REPORT
Changes in hepatitis C virus genotype distribution in Japan

H. TOYODA[1]*, T. KUMADA[1], K. TAKAGUCHI[2], N. SHIMADA[3] AND J. TANAKA[4]

[1] *Department of Gastroenterology, Ogaki Municipal Hospital, Ogaki, Japan*
[2] *Department of Internal Medicine, Kagawa Prefectural Central Hospital, Takamatsu, Japan*
[3] *Department of Gastroenterology and Hepatology, Shinmatsudo Central General Hospital, Matsudo, Japan*
[4] *Department of Epidemiology, Infectious Disease Control, and Prevention, Hiroshima University Institute of Biomedical and Health Sciences, Hiroshima, Japan*

Received 27 September 2013; Final revision 21 December 2013; Accepted 10 February 2014

SUMMARY

Genotypes are associated with the natural course of hepatitis C virus (HCV) infection and response to antiviral therapy for HCV. HCV genotype 1b has been the dominant genotype in Japan, where the prevention of HCV transmission through blood transfusion or nosocomial infection has been established since 1990. The distribution of HCV genotype was investigated based on patient's birth year in 5515 HCV-infected Japanese individuals at three institutions from different areas of Japan. At all three institutions, the proportion of HCV genotype 1b decreased and was <50% in individuals born after 1970. By contrast, the percentage of HCV genotype 2b increased in subsequent birth cohorts after 1920–1929. Significant changes in HCV genotype distribution were observed across Japan regardless of area.

Key words: Birth year, distribution, genotypes, hepatitis C virus.

Hepatitis C virus (HCV) infection is a major cause of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma. HCV is classified into several genotypes and sub-genotypes with varying prevalence rates in different regions of the world [1]. In Japan, the majority of individuals with HCV infection have genotype 1b [2, 3], which is one of the sub-genotypes resistant to interferon-based antiviral therapy [4]. Before the detection of HCV antibodies was established in 1990, the main modes of HCV transmission had been blood transfusion or nosocomial infection. After the discovery of HCV, prevention of HCV transmission through these routes was established and the number of individuals newly infected with HCV rapidly decreased in Japan [5]. Although decreases in the number of individuals with HCV infection have been reported, changes in the distribution of HCV genotype over time have not been studied. In the present study, we investigate changes in HCV genotype distribution based on the birth year of HCV-infected individuals in Japan. We analysed individuals from three different areas of the country to clarify whether any changes in HCV genotype distribution constitute a nationwide trend.

HCV genotype was analysed in a total of 5515 HCV-infected Japanese individuals at three institutions located in different parts of Japan (Fig. 1): Ogaki

* Author for correspondence: H. Toyoda, M.D., Ph.D., Department of Gastroenterology, Ogaki Municipal Hospital, 4-86 Minaminokawa, Ogaki, Gifu, 503-8502, Japan.
(Email: hmtoyoda@spice.ocn.ne.jp)

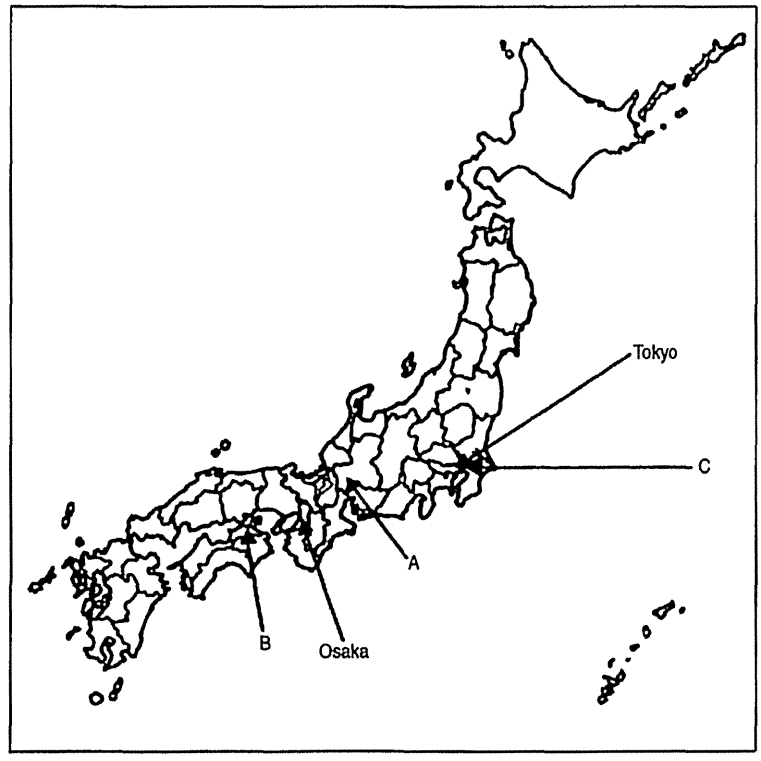

Fig. 1. Location of the three Japanese institutions (liver centres) participating in the study. A, Ogaki Municipal Hospital; B, Kagawa Prefectural Central Hospital; C, Shinmatsudo Central General Hospital.

Municipal Hospital (institution A; Ogaki City, located in central Japan, 3269 individuals), Kagawa Prefectural Central Hospital (institution B; Takamatsu City, located on Shikoku Island, western Japan, 1421 individuals), and Shinmatsudo Central General Hospital (institution C; Matsudo City, near Tokyo, 825 individuals). These individuals received regular follow-up at one of the institutions between 1991 and 2013. HCV infection was confirmed by both positive serum HCV antibody (Architect Anti-HCV; Abbott Laboratories, USA) and serum HCV RNA using a real-time polymerase chain reaction (PCR)-based method (COBAS AmpliPrep/COBAS TaqMan HCV test; Roche Molecular Systems, USA; lower limit of detection, 1·2 $\log_{10}$ IU/ml). HCV genotype was assessed using PCR methods to amplify the core gene sequences using genotype-specific primers [6]. Genotype was classified as 1a, 1b, 2a, 2b, 3a, or mixed. Review and analysis of clinical data including patient's birth year, age, sex, and HCV genotype were approved by the Institutional Review Board of each institution.

Trends for changes in the percentage of each HCV genotype according to birth year were evaluated with the Cochran–Armitage test. All P values were two-tailed, and P <0·05 was considered statistically significant.

All individuals studied were of Japanese ethnicity and no immigrants were included. The HCV-infected individuals comprised of 1790 males and 1479 females with a mean age of 67·1 ± 13·3 years at institution A; 729 males and 692 females with a mean age of 62·0 ± 12·7 years at institution B; and 419 males and 406 females with a mean age of 66·6 ± 13·3 years at institution C. The distribution of HCV genotypes in all patients was: institution A [genotype 1b (64·3%), genotype 2a (27·0%), genotype 2b (7·7%), and other (including genotypes 1a, 3a, and mixed) (1·0%)]; institution B [genotype 1b (62·3%), genotype 2a (26·5%), genotype 2b (10·3%), and other (including genotypes 1a and 3a) (1·0%)]; and institution C [genotype 1b (70·8%), genotype 2a (17·6%), genotype 2b (10·3%), and other (including genotypes 3a and mixed) (1·3%)].

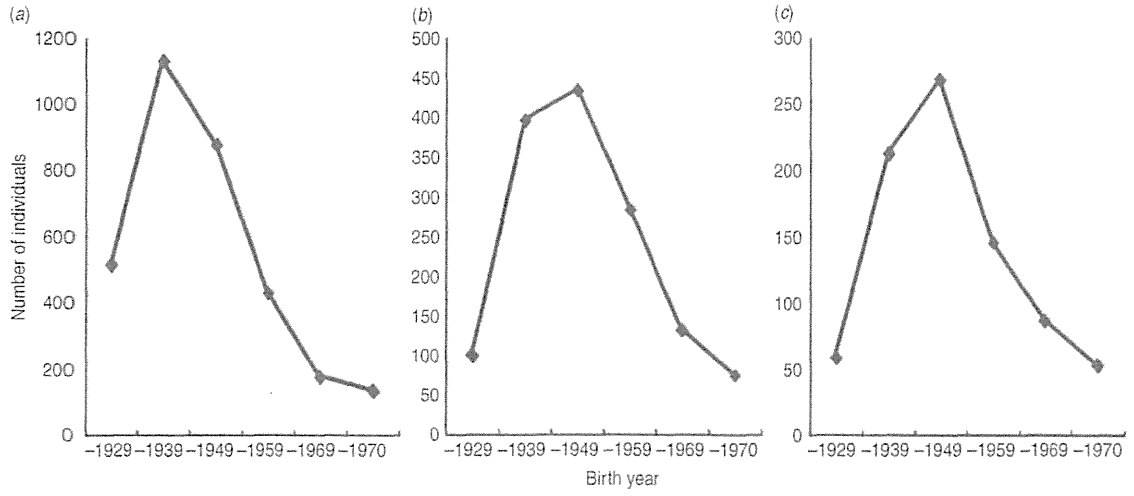

Fig. 2. Changes in the number of hepatitis C virus-infected individuals under follow-up at institutions based on birth year at (*a*) Ogaki Municipal Hospital, (*b*) Kagawa Prefectural Central Hospital, and (*c*) Shinmatsudo Central General Hospital.

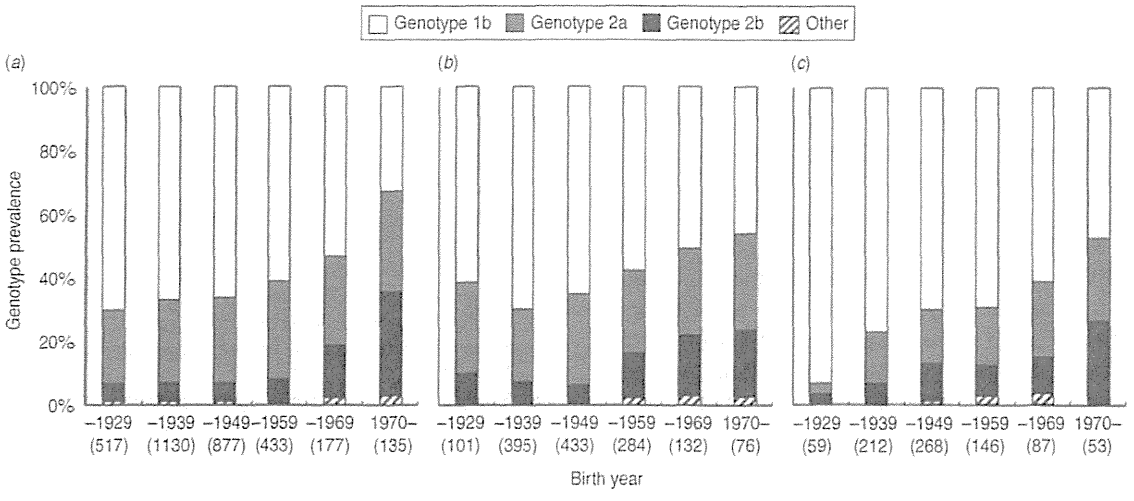

Fig. 3. Changes in hepatitis C virus (HCV) genotype distribution based on birth year of HCV-infected individuals at (*a*) Ogaki Municipal Hospital, (*b*) Kagawa Prefectural Central Hospital, and (*c*) Shinmatsudo Central General Hospital. The number of individuals born during each time period is indicated in parentheses. 'Other' includes genotypes 1a, and 3a, as well as mixed genotypes.

At all three institutions, the number of individuals with HCV infection who were under follow-up at the institutions decreased over time starting with birth years 1930–1939 at institution A and 1939–1940 at institutions B and C (Fig. 2). Figure 3 shows the distribution of HCV genotypes based on patient's birth year at each institution. There is a trend of HCV genotype 1b decreasing over time ($P<0{\cdot}0001$ for all three institutions). The percentage of HCV genotype 1b was <50% in individuals born after 1970. The percentage of HCV genotype 2a was constant except for individuals born before 1929. By contrast, the percentage of HCV genotype 2b increased during the period studied ($P<0{\cdot}0001$ for institutions A and B, and $P=0{\cdot}0004$ for institution C).

This study showed that the HCV genotype distribution changed markedly over time in Japan. The trend was confirmed at three large institutions from different parts of the country and seems to be nationwide. The percentage of individuals with HCV genotype 1, which is a major HCV genotype worldwide [2], is decreasing and is found in <50% of individuals born after 1970. A decrease in the prevalence of patients with HCV genotype 1b has also been reported in the USA and Europe in studies with a smaller number of subjects [7–9].

Several studies have reported that HCV genotypes are different between individuals with HCV acquired through transfusions or medical procedures and individuals with HCV acquired through other transmission route, including intravenous drug use or tattooing [8, 10–13]. In Japan, HCV genotype 1b is associated with the former route of transmission, whereas other genotypes, i.e. genotypes 2a and 2b, are associated with the latter route [13]. With the establishment of methods to prevent HCV transmission during blood transfusions and medical procedures in developed countries, the routes of HCV transmission other than transfusions or medical procedures have become the main routes of HCV infection. Indeed, among 135 individuals with a birth year of 1970 or later at institution A, 29 (21·5%) had a history of intravenous drug use and 33 (24·4%) had a history of tattooing. By contrast, no individuals had a history of blood transfusion. Although the accurate route of HCV transmission was unclear in most individuals in the present study, these changes in the common routes of HCV transmission may be associated with changes in HCV genotype distribution over time. Further studies will be necessary to clarify why the percentage of HCV genotype 2b increased in Japan.

There are several limitations to this study. HCV genotype distribution was analysed based on the birth year of infected individuals instead of the year of HCV infection since there was not enough available information to accurately determine the year of HCV infection. However, results with a large study population from different parts of Japan can be indicative of changes in HCV genotype distribution. The HCV-infected individuals analysed were those who received regular follow-up and may not reflect the HCV genotype distribution of the entire Japanese HCV-infected population. Better nationwide surveillance of HCV that includes data on genotype determination is required to confirm these observed changes in HCV genotype distribution over time.

In conclusion, significant changes in HCV genotype distribution over time were observed in HCV-infected Japanese individuals in various regions of Japan. HCV genotype 1b may no longer be the predominant genotype of HCV in Japan in the future. These changes may also be occurring in other countries throughout the world, especially developed countries. Reappraisal of HCV genotype distribution will be necessary over time. Recent new antiviral drugs for the treatment of HCV, including direct-acting antivirals, mainly target HCV genotype 1 with some exceptions. However, other genotypes may become dominant in various populations of HCV-infected individuals.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Simmonds P, *et al.*** A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology* 1994; **19**: 1321–1324.
2. **McOmish F, *et al.*** Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey. *Journal of Clinical Microbiology* 1994; **32**: 884–892.
3. **Tanaka H, *et al.*** Hepatitis C virus 1b (II) infection and development of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma: a case-control study in Japan. *Journal of Epidemiology* 1998; **8**: 244–249.
4. **Ghany MG, *et al.*** Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009; **49**: 1335–1374.
5. **Moriya T, *et al.*** Epidemiology of hepatitis C virus in Japan. *Intervirology* 1999; **42**: 153–158.
6. **Ohno T, *et al.*** New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. *Journal of Clinical Microbiology* 1997; **35**: 201–207.
7. **Rosen HR, *et al.*** Molecular epidemiology of hepatitis C infection in U.S. veterans liver transplant recipients: evidence for decreasing relative prevalence of genotype 1B. *American Journal of Gastroenterology* 1999; **94**: 3015–3019.
8. **Bourliere M, *et al.*** Epidemiological changes in hepatitis C virus genotypes in France: evidence in intravenous drug users. *Journal of Viral Hepatitis* 2002; **9**: 62–70.
9. **Chlabicz S, *et al.*** Changing HCV genotypes distribution in Poland – relation to source and time of infection. *Journal of Clinical Virology* 2008; **42**: 156–159.
10. **Mathei C, *et al.*** Molecular epidemiology of hepatitis C among drug users in Flanders, Belgium: association of genotype with clinical parameters and with sex- and

drug-related risk behaviours. *European Journal of Clinical Microbiology and Infectious Diseases* 2005; **24**: 514–522.

11. **Roman F, *et al.*** Hepatitis C virus genotypes distribution and transmission risk factors in Luxembourg from 1991 to 2006. *World Journal of Gastroenterology* 2008; **14**: 1237–1243.
12. **Bojovic K, *et al.*** The comparison of chronic hepatitis C treatment outcome between intravenous drug users and non-intravenous drug users. *Biomedicine and Pharmacotherapy* 2013; **67**: 517–520.
13. **Chung H, Ueda T, Kudo M.** Changing trends in hepatitis C infection over the past 50 years in Japan. *Intervirology* 2010; **53**: 39–43.

HEPATOLOGY
Official Journal of the American Association for the Study of Liver Diseases

CORRESPONDENCE

Noninvasive Diagnosis of Portal Hypertension and Esophageal Varices Through the Identification of Liver Blood Flow Markers

To the Editor:

We read with great interest the article entitled "Use of noninvasive markers of portal hypertension and timing of screening endoscopy for gastroesophageal varices in patients with chronic liver disease"[1] in which the authors, Dr. Annalisa Berzigotti and Prof. Jaime Bosch, describe the usefulness of noninvasive tools for the diagnosis of clinically significant portal hypertension (CSPH, hepatic vein pressure gradient [HVPG] ≥10 mmHg) and esophageal varices (EV).

The most recent guidelines[2] on portal hypertension strongly suggest performing upper endoscopy and, where available, HVPG measurement in all patients with liver cirrhosis. Moreover, endoscopy should be repeated every 2-3 years in patients without esophageal varices and more frequently (according to bleeding risk) in patients with EV. As recognized by Berzigotti et al.,[1] this screening and follow-up program leads to significant healthcare costs and patient discomfort since cirrhosis is, nowadays, frequently diagnosed in a very initial stage when varices are still absent. Therefore, in the near future the selection of high-risk patients represents a clinical challenge for the hepatologist in order to reduce futile examinations, the related costs, and the patients' burden.

We strongly agree with the idea of sparing HVPG measurement and endoscopy in patients with less than 20% probability of CSPH based on the combination of noninvasive tests and to perform it in the remaining patients with higher pretest probability. As the authors correctly point out, a number of noninvasive tests based on liver elastography (alone or combined with other parameters) or on spleen stiffness can help to reliably rule out and diagnose CSPH.

Besides the above-mentioned noninvasive tests, we would like to remember the Indocyanine Green Retention Test (ICG-r15), which is a quantitative function test reflecting liver functional reserve and blood flow. Among patients with initial cirrhosis and well-preserved liver function, ICG-r15 correlates with the presence, degree, and complication of portal hypertension, reflecting the modifications of liver blood flow.

We recently evaluated ICG-r15 as a noninvasive marker of CSPH and EV in a population of 96 consecutive patients with compensated liver cirrhosis of different etiologies[3]; in our study an ICG-r15 <10% correctly ruled out the presence of varices in 26 out 27 patients. Therefore, the good diagnostic performance of ICG-r15 (Table 1) makes it a valid tool for the assessment of PH and EV in cirrhosis patients. Although these results have to be validated in a larger or multicentric population and confirmed by longitudinal analysis, this simple and reproducible test allows an initial stratification of cirrhosis patients.

In conclusion, we definitely agree with the authors on the need to spare unnecessary HVPG and upper endoscopies in this clinical setting and suggest that ICG-r15 might be another test to consider besides those based on transient elastography or liver stiffness, particularly in centers where these technologies are not available.

Table 1. Diagnostic Performance of Indocyanine Green 15 Minutes Retention Test for the Rule-Out of Clinically Significant Portal Hypertension and Esophageal Varices

		Prediction of CSPH (HVPG ≥ 10 mmHg)			Prediction of EV		
	No.	AUROC	Sensitivity	−LR	AUROC	Sensitivity	−LR
ICG-r15[3]	96	0.808	95.9%	0.15	0.859	97.8%	0.042

CSPH: Clinically Significant Portal Hypertension; EV: Esophageal Varices; ICG-r15: Indocyanine Green retention at 15 minutes; AUROC: area under ROC curve; −LR: negative likelihood ratio.

ANDREA LISOTTI, M.D.
FRANCESCO AZZAROLI, M.D.
MARCO MONTAGNANI, M.D.
GIUSEPPE MAZZELLA, M.D., PH.D.
Department of Medical and Surgical Science
University of Bologna
S. Orsola-Malpighi Hospital
Bologna, Italy

References

1. Berzigotti A, Bosch J, Boyer TD. Use of noninvasive markers of portal hypertension and timing of screening endoscopy for gastroesophageal varices in patients with chronic liver disease. HEPATOLOGY 2014;59:729-731.
2. de Franchis R, Baveno V Faculty. Revising consensus in portal hypertension: report of the Baveno V consensus workshop on methodology of diagnosis and therapy in portal hypertension. J Hepatol 2010;53:762-768.
3. Lisotti A, Azzaroli F, Buonfiglioli F, Montagnani M, Cecinato P, Turco L, et al. Indocyanine green retention test as a noninvasive marker of portal hypertension and esophageal varices in compensated liver cirrhosis. HEPATOLOGY 2014;59:643-650.

View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep.26897
Potential conflict of interest: Nothing to report.

Postinterferon α-Fetoprotein Elevation and Risk of Hepatocellular Carcinoma Development After Sustained Virological Response: Cause or Results?

To the Editor:

We read with interest the article by Asahina et al.,[1] which clearly demonstrated a higher incidence of hepatocellular carcinoma (HCC) in patients with higher levels of α-fetoprotein (AFP) after interferon-based antiviral therapy. There was a surprisingly high incidence of HCC (almost 50%) in patients who achieved sustained virological response (SVR) but whose postinterferon (IFN) AFP levels were higher than 20 ng/mL.

There are two distinct patterns of HCC development after SVR. In one pattern, HCC develops after the eradication of hepatitis C virus (HCV). This pattern is associated with the residual

potential for hepatocarcinogenesis after SVR, which may be signaled by elevated AFP after IFN. The other pattern involves HCC that was too minute to be detected before and just after IFN treatment, but grew enough to be visualized on imaging studies during post-SVR follow-up. Previous studies of HCC tumor volume doubling time suggest that some minute HCC tumors would take several years to be detected by imaging modalities.[2] Some patients who achieved SVR, therefore, might have had minute, undetectable HCC at the time of SVR. Although the authors described excluding patients with HCC based on imaging studies, such modalities always have limitations in their ability to detect minute HCC (for example, <5 mm in diameter). In particular, the ability of imaging modalities to detect minute HCC was unsatisfactory during the earlier part of the study period (1990s).

Their Fig. 2F suggests a specific feature in the cumulative incidence curves for HCC based on post-IFN AFP levels in patients with SVR. Among patients who did not achieve SVR, the incidence of HCC continued to increase gradually according to the number of years after SVR. This includes patients with high post-IFN AFP levels, whose HCC incidence curves were similar to incidence curves stratified by post-IFN ALT levels. In contrast, in patients with SVR, cumulative HCC incidence curves according to post-IFN AFP levels were different. The incidence of HCC in patients with post-IFN AFP ≥20 ng/mL and with post-IFN AFP ≥10 ng/mL and <20 ng/mL increased rapidly until 3 to 4 years after SVR, with only a few patients developing HCC thereafter. This feature could be due to the detection of preexisting minute HCC after SVR. Was the post-IFN AFP elevation observed in these patients a marker of enhanced hepatocarcinogenesis or an existing HCC?

It will be difficult to determine whether elevated levels of AFP were produced by HCC without visible HCC on imaging studies. It would be interesting to check the fucosylated fraction of AFP (AFP-L3), a more specific marker that has been reported as a marker of minute HCC,[3] if post-IFN AFP-L3 data were available.

HIDENORI TOYODA, M.D., PH.D.
TAKASHI KUMADA, M.D., PH.D.
TOSHIFUMI TADA, M.D.
Department of Gastroenterology
Ogaki Municipal Hospital
Ogaki, Gifu, Japan

References

1. Asahina Y, Tsuchiya K, Nishimura T, Muraoka M, Suzuki Y, Tamaki N, et al. α-Fetoprotein levels after interferon therapy and risk of hepatocarcinogenesis in chronic hepatitis C. HEPATOLOGY 2013;58:1253-1262.
2. Sheu JC, Sung JL, Chen DS, Yang PM, Lai MY, Lee CS, et al. Growth rate of asymptomatic hepatocellular carcinoma and its clinical implications. Gastroenterology 1985;89:259-266.
3. Kumada T, Toyoda H, Tada T, Kiriyama S, Tanikawa M, Hisanaga Y, et al. High-sensitivity Lens culinaris agglutinin-reactive alpha-protein assay predicts early detection of hepatocellular carcinoma. J Gastroenterol [Epub ahead of print].

DOI 10.1002/hep.27064
Potential conflict of interest: Nothing to report.

Could Postinterferon Treatment α-Fetoprotein Levels Truly Predict Hepatocarcinogenesis?

To the Editor:

I read with interest the article by Asahina et al.[1] regarding the levels of α-fetoprotein (AFP) and alanine aminotransferase (ALT) after interferon therapy that could predict the occurrence of hepatocellular carcinoma (HCC) in patients with hepatitis C virus (HCV) infection. The authors made a great effort to evaluate the levels of AFP and ALT in a cohort of 1,818 patients after interferon therapy and found that cutoff values for ALT and AFP for prediction HCC development were 40 IU/L and 6.0 ng/mL, respectively. These findings may be helpful for clinicians to closely follow-up high-risk patients to detect early-stage HCC. However, these data may also be misleading and need further clarification.

The levels of AFP and ALT were measured every 1-6 months. It is unknown which timepoint of AFP and ALT levels were selected for calculation. Serum AFP levels usually fluctuate during serial observation. Some patients may present with an AFP >6 ng/mL before therapy, which decreased to <6 ng/mL within a few months after interferon therapy, and became elevated to >6 ng/mL at long-term follow-up. Should these patients be classified as AFP ≧6 ng/mL decreased group or AFP ≧6 ng/mL unchanged group? The mean follow-up period of this study was 6.1 years. The follow-up period in this study has been as long as 20 years. If the interval of AFP measurement was as short as 1 month, too many unnecessary measurements could have been performed. As shown in the results, postinterferon therapy AFP level ≧6.0 ng/mL had a positive predictive value of only 0.262. Consistent with previous observation, this may evoke inappropriate suspicion of malignancy in 74 out of 100 patients with AFP above this cutoff value.[2] A total of 179 patients developed HCC, accounting for 9.8% of the entire cohort of 1,818 patients. Based on the HALT-C trial, patients with persistent elevation of AFP after interferon therapy, only 2% were noted to develop HCC.[3] We believe that a large proportion of patients still did not have elevated AFP levels on detection of HCC occurrence. Thus, the applicability and cost-effectiveness of this policy merits further investigation.

GIN-HO LO, M.D.
Department of Medical Research
E-DA Hospital
School of Medicine for International Students
I-Shou University
Kaohsiung, Taiwan

References

1. Asahina Y, Tsuchiya K, Nishimura T, Muraoka M, Suzuki Y, Tamaki N, et al. α-Fetoprotein levels after interferon therapy and risk of hepatocarcinogenesis in chronic hepatitis C. HEPATOLOGY 2013;58:1253-1262.
2. Trevisani F, D'Intino PE, Maria A, Accogli E, Caraceni P, Domenicali M, et al. Serum α-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. J Hepatol 2001;34:570-575.
3. Di Bisceglie AM, Sterling RK, Chung RT, Everhart JE, Dienstag JL, Bonkovsky HL, et al. Serum alpha-fetoprotein levels in patients with advanced hepatitis C: results from the HALT-C Trial. J Hepatol 2005;43:434-441.

DOI 10.1002/hep.27065
Potential conflict of interest: Nothing to report.

FULL PAPER

BJC

British Journal of Cancer (2014) 110, 2090–2098 | doi: 10.1038/bjc.2014.130

Keywords: hepatocellular carcinoma; prognosis; prognostic models; biomarkers; AFP; DCP; AFP-L3; bilirubin; albumin

Biomarker-based prognosis in hepatocellular carcinoma: validation and extension of the BALAD model

R Fox[1], S Berhane[2], M Teng[3], T Cox[4], T Tada[5], H Toyoda[5], T Kumada[5], C Kagebayashi[6], S Satomura[6] and P J Johnson[*,2,7]

[1]Cancer Research UK Clinical Trials Unit, School of Cancer Sciences, University of Birmingham, Birmingham B15 2TT, UK; [2]Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool L69 3GA, UK; [3]Department of Oncology, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK; [4]Liverpool Cancer Research UK Centre, Liverpool Cancer Trials Unit, University of Liverpool, Liverpool L69 3GL, UK; [5]Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, Gifu 503-8052, Japan; [6]Wako Life Sciences, Inc., Mountain View, CA 94043-1829, USA and [7]The Clatterbridge Cancer Centre NHS Foundation Trust, Clatterbridge Road, Bebington, Wirral CH63 4JY, UK

Background: The Japanese 'BALAD' model offers the first objective, biomarker-based, tool for assessment of prognosis in hepatocellular carcinoma, but relies on dichotomisation of the constituent data, has not been externally validated, and cannot be applied to the individual patients.

Methods: In this Japanese/UK collaboration, we replicated the original BALAD model on a UK cohort and then built a new model, BALAD-2, on the original raw Japanese data using variables in their continuous form. Regression analyses using flexible parametric models with fractional polynomials enabled fitting of appropriate baseline hazard functions and functional form of covariates. The resulting models were validated in the respective cohorts to measure the predictive performance.

Results: The key prognostic features were confirmed to be Bilirubin and Albumin together with the serological cancer biomarkers, AFP-L3, AFP, and DCP. With appropriate recalibration, the model offered clinically relevant discrimination of prognosis in both the Japanese and UK data sets and accurately predicted patient-level survival.

Conclusions: The original BALAD model has been validated in an international setting. The refined BALAD-2 model permits estimation of patient-level survival in UK and Japanese cohorts.

The key features that influence prognosis in hepatocellular carcinoma (HCC) are now well recognised and can be broadly classified under the headings of tumour-related factors (such as tumour size or multiplicity), those that assess the severity of underlying liver dysfunction (such as conventional liver function tests or the Child–Pugh (C-P) classification (Child and Turcotte, 1964; Pugh *et al*, 1973)) and patient-related factors (such as symptoms or performance status). Several staging systems/prognostic scores that combine a number of these factors have been developed (Okuda *et al*, 1985; Group, 1998; Chevret *et al*, 1999; Leung *et al*, 2002; Kudo *et al*, 2003; Llovet *et al*, 2008) and, to varying degrees, validated and compared (Kudo *et al*, 2004; Marrero *et al*, 2005; Cho *et al*, 2008; Collette *et al*, 2008; Chen *et al*, 2009; Huitzil-Melendez *et al*, 2010; Chan *et al*, 2011). Some simply offer an estimate of prognosis, whereas others aim to indicate the appropriate therapy for specific disease stages (Llovet *et al*, 2008).

In an attempt to develop a more objective staging system, Toyoda *et al* (2006) have described the BALAD model that relies on two liver function tests (Bilirubin and Albumin) and three serological cancer biomarkers (AFP-L3, AFP, and DCP). They have

*Correspondence: Professor PJ Johnson; E-mail: Philip.Johnson@liverpool.ac.uk

Received 22 November 2013; revised 16 January 2014; accepted 28 January 2014; published online 1 April 2014

shown that it is possible to achieve an excellent degree of discrimination between the proposed risk groups using such objective variables. However, the data analysis approach in the BALAD model utilised dichotomisation of the continuous variables, which raises a number of statistical issues.

In the present study, we aimed to validate the original BALAD model (built on a Japanese Cohort) in a geographically and aetiologically distinct HCC patient data set from the UK. We first confirmed that the variables in the BALAD model were identical to those independently identified in a UK data set, and assessed the discrimination achieved within the proposed prognostic groups. We then, in a collaborative Japan/UK study, took the raw data on which BALAD model was initially derived and applied a more sophisticated statistical method that treats the variables in a continuous manner and does not assume a linear relationship between predictors and outcome. The model developed here not only allows classification of patient risk, as with the original BALAD model, but also provides detailed estimation of patient-level survival in the Japanese cohort, and, with calibration, in UK patients.

A major challenge in applying the BALAD model to the UK population is the great difference in survival compared with the Japanese cohort. This problem is due to the difference in the underlying survivor function that describes hazard in relation to time; hazard could be greatest at diagnosis and then decrease over time or, conversely, the hazard at diagnosis may be low and then increase as time accumulates. Indeed, the hazard may be described by a more complicated, non-linear, and not necessarily monotonic function. To account for such differences, the methods applied in this analysis allowed interrogation of the scale and shape of the baseline hazard function.

The derived model is assessed in terms of discrimination and calibration. To assess discrimination, Harrell's *C*-statistic was measured, as described by Taktak *et al* (2007). This measures the proportion of patient pairs for which the model correctly assigns lower risk to the patient who truly survives longest (i.e. is at least risk). A model with good discriminative performance should have a high *C*-statistic. To assess calibration, graphical methods were used. These assessments compare patient level survival with the predicted values.

MATERIALS AND METHODS

The study comprised two cohorts of patients. The first included 2599 Japanese patients previously reported by Toyoda *et al* (2006) and 319 UK patients, all with HCC (Table 1). The Japanese patients were recruited from five institutions in which a total of 3725 patients were initially diagnosed as having HCC between July 1994 and December 2004, and the UK patients from among 724 patients referred to the Queen Elizabeth, Birmingham, UK, between June 2007 and January 2012. The various aetiologies were classified as hepatitis B virus-related, hepatitis C virus-related, alcoholic-related, and 'other'. The 'other' group comprised patients with hemochromatosis, primary biliary cirrhosis, non-alcoholic steatohepatitis, or cryptogenic cirrhosis. The diagnosis of chronic liver disease was made on the basis of liver biopsy and/or typical clinical and imaging features. The study protocol was approved by the institutional ethics review board at each of the institutions.

Age and gender distributions were similar in the two populations, as was the distribution of liver dysfunction as assessed by the C-P classification (Table 1). However, there were striking differences in aetiological attribution, the Japanese patients having predominantly HCV-related HCC and the UK patients having multiple aetiologies. There were also major differences in disease stage (Table 1) and overall survival between the two cohorts. The median survival for those treated with palliative and curative therapy was 22.6 and 60.7 months for Japanese patients, respectively, with analogous figures for the UK of 13.9 and 27.5 months.

In all patients, the three serological cancer biomarkers of HCC (AFP, AFP-L3, and DCP) were measured at the time of diagnosis, and drugs that would influence the serum DCP levels, such as warfarin and vitamin K, were not taken. A standard operating procedure was applied to all blood collection. Samples were collected in the fasting state, before any treatment. Blood was allowed to clot at room temperature for 1–2 h, centrifuged at 3000 **g** for 20 min and the serum collected and stored at −80 °C until processing. Routine liver and renal function was measured by commercially available methods. Albumin was measured by the bromocresol green method in both UK and Japan. The severity of the liver disease was defined according to C-P classification.

Patients were staged by five systems: TNM 5, TNM 6 (Sobin and Fleming, 1997; Greene *et al*, 2002; UICC, 2002; Sobin *et al*, 2011), CLIP (Group, 1998), JIS or BCLC (Llovet *et al*, 2008), or by Milan criteria (Mazzaferro *et al*, 1996). However, for this analysis that focused on prognosis, we also grouped patients on the basis of whether or not the treatment received was curative or palliative. Curative treatments included transplantation, resection, radiofrequency ablation, and percutaneous ethanol injection. Palliative treatments included transarterial chemoembolisation, any form of chemotherapy, and supportive care. Where patients were listed for transplantation but had transarterial chemoembolisation as initial treatment as a 'bridge' to transplantation, they were classified as having potentially curative therapy. For the purpose of this analysis, UK patients who underwent liver transplantation were excluded, as the survival of this group would not be expected to be influenced by the baseline features included in the model (such as bilirubin and albumin).

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) (Kagebayashi *et al*, 2009). Analytical sensitivity of μTASWako i30 is 0.3 ng ml^{-1} AFP, and the percentage of AFP-L3 can be measured when AFP-L3 is over 0.3 ng ml^{-1} (Kagebayashi *et al*, 2009).

Statistical methods

Discrimination. We assessed discriminatory performance using Harrell's *C*-statistic, as described by Taktak *et al* (2007). In brief, this measure reports the number of comparable pairs that are correctly ordered under the risk score. That is, for a pair of comparable patients P_A and P_B, if patient P_A is known to have survived beyond P_B's time of event (death here), then P_A should be subject to a lesser risk than P_B, that is, should be assigned a lower-risk group. This method counts all the correctly ordered pairs from those that are comparable.

Flexible parametric models. Regression analyses utilised flexible parametric models (Royston and Lambert, 2011) that enable fitting of more appropriate baseline hazard functions. The baseline hazard describes risk over time when all covariates take the value zero (rather than the hazard at time zero as sometimes stated), and is described by a restricted cubic spline function (Royston and Lambert, 2011). Here all continuous covariates are centred about their mean, and so the interpretation of the function is the hazard at the mean of all covariates. Traditionally, the baseline hazard is assumed to have a simple constant or monotonic form, as in

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

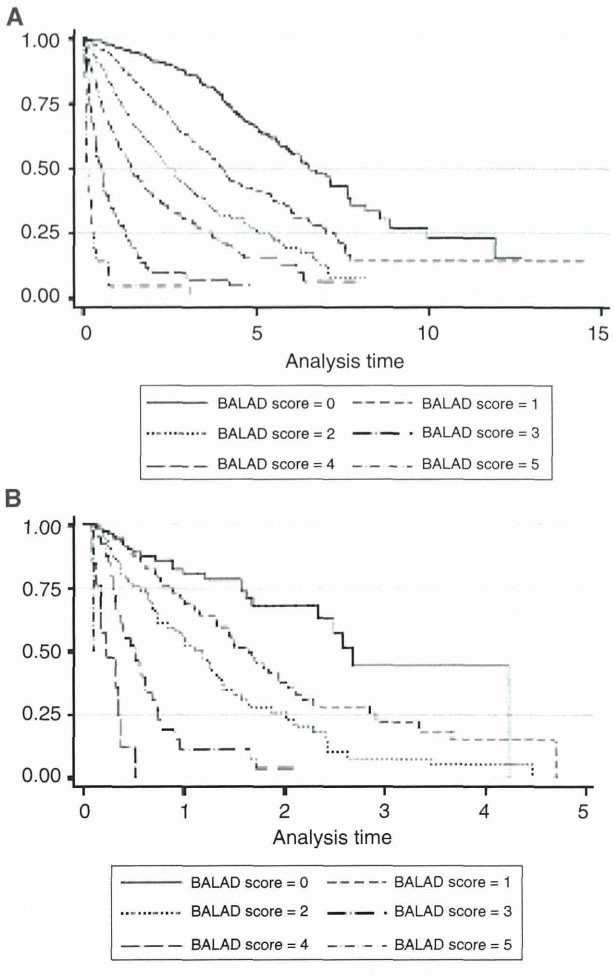

Figure 1. Survival according to the BALAD model. Kaplan–Meier curves showing survival according to the original BALAD model in (**A**) Japanese and (**B**) UK cohorts.

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 ⩽ −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.