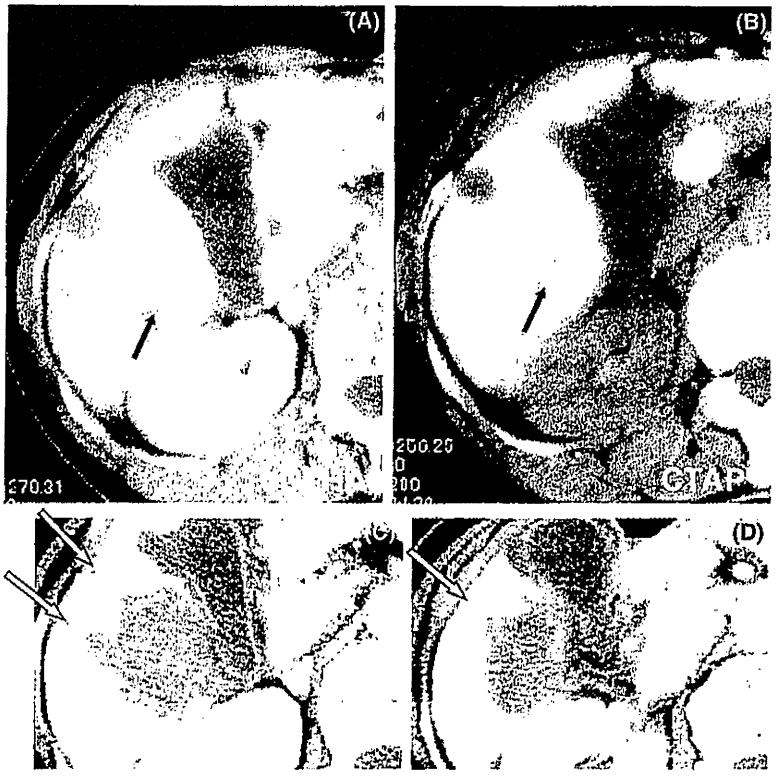

Fig. 5 Images in a female patient who had hepatocellular carcinoma (HCC) in segment VI. Before treatment, scans obtained on computed tomography during hepatic arteriography (CTHA) (a) and computed tomography during arterial portography (CTAP) (b) showed a nodular HCC (*arrow*) measuring 2.5 cm. Three applicators were placed in parallel in the HCC in liver segment VI, and then the tumor was ablated in one procedure (total ablation time 13 min 42 s, total applied energy 35.3 kJ). After the procedure, computed tomography (CT) images showed a necrotic area of 46 mm in diameter including the nodular HCC (c [*arrows* show applicator for insertion paths], d [*arrow* shows applicator for insertion paths])

Table 2 Maintenance of the therapeutic effect (TE) (overall assessment of the TE; intention-to-treat (ITT) analysis)

	This clinical study		Patients who underwent local therapy [24]	
	10 weeks	24 weeks[a]	3 months	6 months
Complete response (CR) (no. of patients)	85/88 (96.6 %)	82/87 (94.3 %)	4468/5394 (82.8 %)	4318/5378 (80.3 %)
Other (no. of patients)	3/88 (3.4 %)[b]	5/87 (5.7 %)[c]	926/5394 (17.2 %)	1060/5378 (19.7 %)

[a] One patient who died was omitted from the 24-week assessment

[b] Includes 3 patients who developed local recurrence within 10 weeks

[c] Includes 5 patients who developed local recurrence within 24 weeks

of each RFA session in the present clinical study. As shown in Fig. 6a, the complete necrosis (Class V TN) rate with the CelonPOWER System was 97.8 % (88/90 patients), which was higher than the rate of 86.2 % (50/58 patients) with the Cool-tip RF System. These results thus confirm the non-inferiority of the CelonPOWER System ($p < 0.001$; Fisher's exact test based on binomial distribution). As shown in Fig. 6b, the percentage of patients in whom treatment was completed in a single session was 77.8 % (70/90 patients) in the present study with the CelonPOWER System, compared with 51.7 % (31/60 patients) in the Cool-tip RF System study [10].

Safety

The overall safety assessment was performed for the entire clinical study period, i.e., inclusive of the treatment period and the follow-up period. Of the 91 patients included in the safety analysis, no procedure was rated as unsafe, although 2 procedures (2.2 %) were rated as somewhat unsafe, one with an abdominal wall burn and one with biliary peritonitis owing to bile leakage; 78 procedures (85.7 %) were rated as safe overall and 11 procedures (12.1 %) were rated as safe. There was no device failure. In the patient with biliary peritonitis, three 30-mm electrodes had been

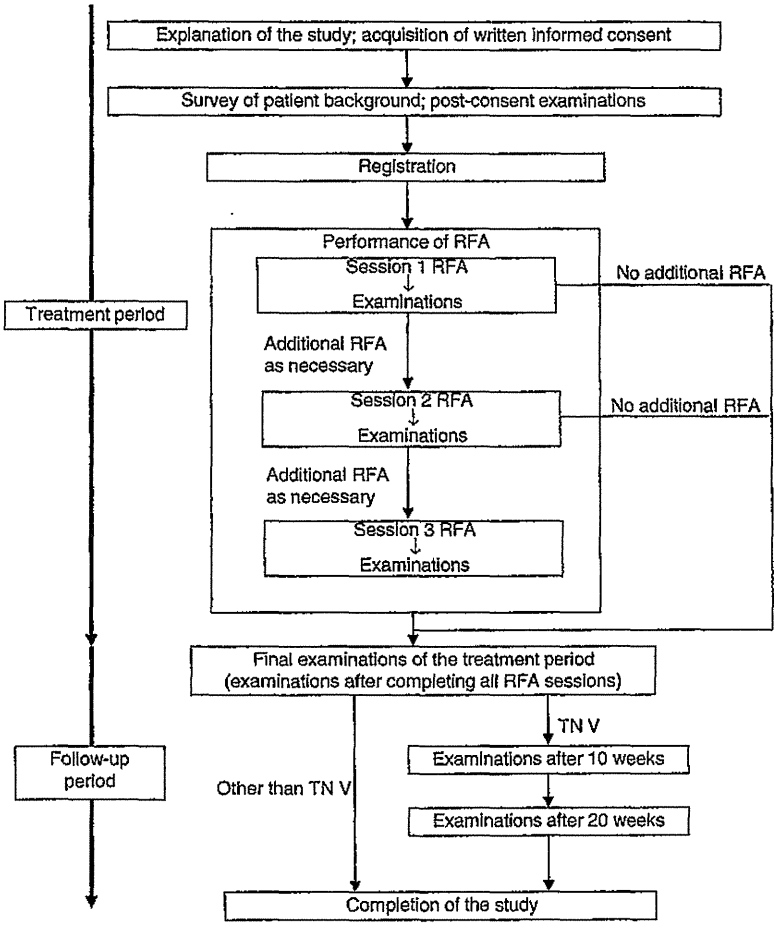

Fig. 4 Clinical study procedure. *TN* Tumor necrosis

Efficacy

Of the 90 patients who completed this clinical treatment study, 88 showed Class V TN (97.8 %). The 2 patients (2.2 %) who did not show 100 % TN both had primary liver cancers and were categorized as Class IV TN. The CNR was 100 % in patients with metastatic liver cancer (7/7 patients) and 97.6 % in patients with primary liver cancer (81/83 patients). The Japanese package insert for the Cool-tip RF System [21] states that the CNR obtained by that system was 86.2 % (50/58 patients). Assuming a 5 % non-inferiority margin, the lower limit of the confidence interval (one-sided 97.5 %) was 92.2 %, and the p value was <0.001 for the exact test based on binomial distribution.

The initial success rate (Class V TN after 1 session) was 77.8 % (70 of 90 patients), while Class V TN was seen in 16 (17.8 %) patients following a second session. The remaining 4 (4.4 %) patients underwent a third RFA session, and 2 were rated as Class V TN following that session.

We used 1 applicator in 20 patients, 2 simultaneously in 54 patients, and 3 simultaneously in 16 patients. We used 30-mm electrodes in all the patients, except in 3 of the 16 patients in whom 3 electrodes were used simultaneously; in these 3 patients we used 3 40-mm electrodes. A representative case in which 3 applicators were used is shown in Fig. 5.

Of the 88 patients who proceeded to the follow-up phase, excluding the single out-of-hospital fatality, examination at 24 weeks showed that CR was obtained in 94.3 % (82/87). The cumulative local recurrence rate at the end of 24 weeks in the follow-up period was 5.7 % (5/87 patients; ITT analysis) (Table 2).

Figure 6a, b shows a comparison of the treatment results of the Cool-tip RF System clinical trial [21] and the number of patients analyzed for the CNR and the efficacy

Table 1 Patient background factors and lesion characteristics

Patients ($n = 91$)		Lesions ($n = 112$)	
Background factors	*N* (%)	Characteristics	*N*
Sex		Maximum dimension (cm)	
M	61 (67.0)	<1.0	22
F	30 (33.0)	1.1–2.0	69
Age (years)		2.1–3.0	17
31–40	1 (1.1)	3.1–4.0	4
41–50	4 (4.4)	Mean ± SD	
51–60	9 (9.9)	1.6 ± 0.7	
61–70	32 (35.2)	Subsegment	
71–80	34 (37.4)	S1	0
81–90	11 (12.1)	S2	6
Cancer		S3	9
Primary	84 (92.3)	S4	8
Metastatic	7 (7.7)	S5	18
Underlying disease		S6	20
Cirrhosis	63 (69.2)	S7	18
Chronic hepatitis	22 (24.2)	S8	33
None	6 (6.6)		
Child-Pugh classification			
Grade A	83 (91.2)		
Grade B	8 (8.8)		
Number of treated lesions			
1	73 (80.2)		
2	15 (16.5)		
3	3 (3.3)		
Previous treatment of primary disease			
Yes	40 (44.0)		
No	51 (56.0)		

markers), or others. In addition, ITT analysis was performed in regard to the cumulative local recurrence rate and the overall assessment of the TE.

Assessment of safety

The following safety endpoints were assessed in all 91 patients in whom the study was conducted: overall safety assessment, adverse events, device-related adverse events, device failure, laboratory test values, and blood pressure.

Statistical analysis

Statistical analysis was performed using a one-sided significance level of 2.5 % for the primary endpoint. In principle, a two-sided significance level of 5 % was used for the other endpoints to avoid data dispersion. The CNR (the primary endpoint) was calculated as the percentage of the total number of patients who achieved Class V TN, and its exact one-sided 97.5 % confidence interval was calculated. For the secondary endpoints, the variables and their ratios were compiled, and the basic statistics for the mean and standard deviation were calculated.

Results

Patients

Written informed consent was obtained from 104 patients, including the 96 patients in the study. The study was conducted in 91 of these patients, and treatment was completed in 90 patients. Eighty-eight of the 90 patients (excluding 2 TN4 patients) were followed up. Five patients discontinued the study during the follow-up period, leaving 83 patients who completed the follow-up period. Three patients were excluded because of unacceptable enrollment dates, so the final number of patients eligible for the efficacy analysis was 80.

controls the high-frequency power (Fig. 3b). This function makes it possible to prevent unexpected rapid increases in electrical resistance resulting from tissue necrotization.

Patients

This clinical study was carried out based on the HCC treatment algorithm in the Scientific Data-based Clinical Practice Guidelines for Liver Cancer-2005 Version [18]. We enrolled adult male and female patients aged 20 years or older with primary or metastatic small liver cancers who had provided written informed consent. Target tumors were defined as nodular, numbering up to 3 lesions, each of which was 3 cm or less in diameter, or solitary lesions up to 4 cm in diameter. Exclusion criteria included a Child-Pugh grade of C, or platelet count below 50000/μl. Informed consent was obtained from 104 patients, of whom 96 were initially enrolled, but 5 withdrew consent before the trial started. The trial was therefore carried out in a total of 91 patients (112 treated lesions) with intention-to-treat (ITT) analysis, and 90 patients were eligible for the analysis of efficacy.

Patient details

Table 1 summarizes the data on the background characteristics of the 91 patients and 112 treated lesions treated in the study (73 patients had 1 lesion, 15 had 2, and 3 patients had 3 lesions; Table 1). The cohort consisted of 61 men and 30 women, and the mean age (±SD) was 69 ± 10 years; 84 patients had primary liver cancer, while 7 had metastatic liver cancer.

Study design

This prospective multicenter, collaborative, single-group, open-label study was conducted at 5 institutions between December 2008 and December 2009. The study protocol was approved by each center's institutional review board. The trial treatment period lasted from the acquisition of written informed consent through completion of the final treatment (maximum 3 treatments), in addition to a follow-up period from the day after the final examinations of the treatment period until the completion of examinations performed 24 weeks later. The non-inferiority of the CelonPOWER System was evaluated relative to the results obtained with a Cool-tip RF System in 2002–2003 [10].

Study methodology

Figure 4 shows the study procedures. During the treatment period, the following procedures were performed, in the order listed: registration of eligible patients, RFA treatment and examinations including computed tomography (CT) imaging, laboratory tests, and blood pressure measurement. The efficacy was evaluated from the extent of the necrotic area (tumor necrosis; TN) induced by ablation as measured on conventional and dynamic CT imaging. Additional ablation, up to a maximum of 3 sessions, was performed as necessary. The laboratory tests consisted of RBC count, WBC count, hemoglobin level, hematocrit, platelet count, prothrombin time (PT) activity, total bilirubin, albumin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), blood urea nitrogen (BUN), and creatinine.

In the follow-up phase, at 10 ± 2 weeks (70 ± 14 days) and 24 ± 2 weeks (168 ± 14 days) following the day of the final RFA session, we performed CT imaging, laboratory tests, blood pressure measurement, measurement of alpha-fetoprotein (AFP), and measurement of protein induced by vitamin K absence or antagonist II (PIVKA-II). The CT images and tumor marker data were employed to assess the continuity of the therapeutic effect (TE) of the RFA treatment.

RFA procedure

The procedure with the CelonPOWER System device was similar to the procedure with the existing monopolar RFA devices. In all cases, the procedure was performed percutaneously under ultrasound guidance and local anesthesia.

Assessment of efficacy

TN was assessed using 5 grades, in accordance with the Criteria for Direct Effects of Liver Cancer Treatment (1994) [19]. Class V tumor necrosis (100 % TN) of liver cancer following the final RFA session was defined as "complete necrosis," and the percentage of patients achieving Class V TN was defined as the "complete necrosis rate" (CNR), the primary endpoint. The TN classification was used for short-term (during treatment) evaluation, and this was the only evaluation reported for the Cool-tip RF System in the marketing authorization holder's application for Japanese government approval. However, now the government demands not only short-term evaluation, but also long-term evaluation, for which such parameters as TE, overall response, and complete response (CR) are used.

The secondary endpoints of our study were the number of RFA sessions, the TE, and the overall assessment of the TE. The assessment of the immediate TE and the overall assessment of TE were performed in accordance with the General Rules for the Clinical and Pathological Study of Primary Liver Cancer (2008) [20]. The TE was classified as either CR (total necrosis and normalization of all tumor

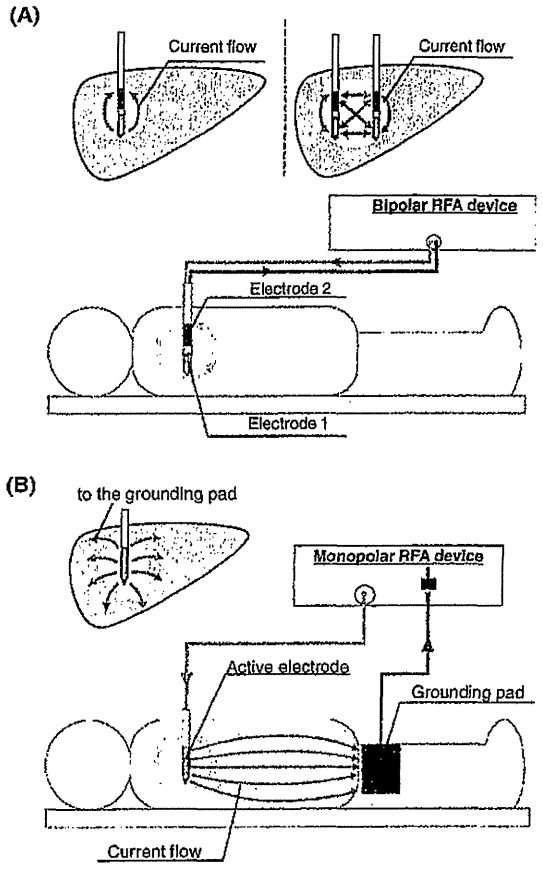

Fig. 2 Differences in the electrical flow routes of a the monopolar and b the bipolar (CelonPOWER System) radiofrequency ablation (*RFA*) systems. With the bipolar system (CelonPOWER System), the electrical current flows between the two electrodes, and for this reason the current pathway is limited to the treatment area, allowing lower power to be concentrated in a specific area and yet yielding effects equivalent to those obtained by higher energy monopolar devices, the power of which is dispersed throughout the body to the dispersion grounding pads placed under the patient

frequency of the power generator is 470 kHz, with a maximum output of 250 W. All the needles for RFA are 1.8 mm in width (15 G) but there are 3 different lengths: 20, 30, and 40 mm. The Cool-tip RF System needles are 1.5 mm in width (17 G).

Bipolar applicators

Each applicator is needle-shaped and has two electrodes near its tip. The electrical current flows between the two electrodes on the single probe, limiting the current pathway to within the treatment area. A grounding pad is unnecessary (Fig. 2a). The applicators are cooled by the internal circulation of chilled water.

Multipolar application

When simultaneously using multiple applicators (up to 3 can be employed simultaneously), it is possible to treat relatively large cancers that could not be sufficiently ablated by means of one insertion of a single applicator. The high-frequency electrical current flows sequentially between the electrodes of the applicators (6 electrode pair combinations when there are 2 applicators, 15 electrode combinations when there are 3 applicators) (Fig. 3a).

Resistance controlled automatic power (RCAP)

RCAP is a function that monitors the change of electric resistance between the electrodes, and automatically

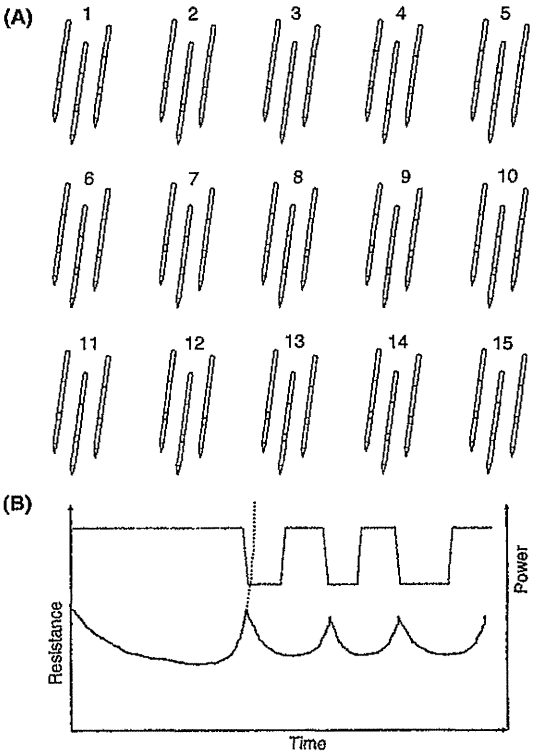

Fig. 3 When 3 applicators are employed, the high-frequency electrical current flows sequentially between 15 combinations of electrode pairs (a), and an image is generated of the automated control of the output by the resistance controlled automatic power (RCAP) function (b). RCAP is a function by which the degree of change in the electrical resistance among the electrodes (increase/decrease in slope) is monitored, and the high-frequency power output is automatically controlled

even if the primary tumor is completely resected, there is a very high recurrence rate in the residual liver [3, 4].

Radiofrequency ablation (RFA) is a minimally invasive method that can yield radical localized therapeutic results, and it has become a standard treatment for small liver cancers 3 cm or less in diameter [5].

Three different RFA systems have been introduced in Japan, all consisting of monopolar devices. One of the main problems with monopolar RFA devices is that the electrical current flows between the electrodes and the grounding pad that is used in these devices. The current flows in a wide area of the body, which may cause systemic symptoms, such as heat retention and perspiration. In addition, because the applicator is distant from the grounding pad, its low energy efficiency requires a long ablation time. Moreover, energy concentration can occur owing to an unanticipated current pathway between the applicator and grounding pad, posing a risk of burns at the grounding pad patch site and at non-treatment sites [3, 6–9].

A bipolar system, in contrast to the monopolar systems, features as its principal characteristic an electrical current flowing between two electrodes on a single probe. With a bipolar system, the current pathway is limited to only within the treatment area, thus eliminating the need for a grounding pad. A bipolar RFA system also overcomes such disadvantages of a monopolar system as the occurrence of heat retention and other side effects, low energy efficacy, and thermal injuries at electrode pad sites caused by an electrical current flowing in the body. The simultaneous use of multiple applicators with a bipolar system makes it possible to achieve a sufficiently large thermocoagulation volume with a single ablation procedure. That is, one ablation is usually sufficient for a wide area and this enables a short ablation time. In addition, ablation can be achieved even if the electrodes are not inserted directly into the tumor. The use of the bipolar system with multiple applicators with a wide ablation area maximizes the effectiveness of the bipolar system.

The purpose of this study was to evaluate the safety and efficacy of a bipolar RFA device, the CelonPOWER System, in order to obtain the clinical data necessary for an application for its regulatory approval in Japan. The study and protocol were designed in compliance with Japanese good clinical practice (GCP) based on the advice from the Pharmaceuticals and Medical Devices Agency (PMDA) of the Japanese regulatory authority. In designing this study, we were requested by the PMDA to compare this device with an existing RFA device (that had been already approved in Japan) and we selected the data from the 2002 to 2003 clinical study of the Cool-tip RF System as valid control data. The study of the Cool-tip RF System was also conducted to obtain marketing approval in Japan [10]. This study was sponsored by Olympus Medical Systems Corp.

Patients, materials, and methods

Device

Celon AG Medical Instruments (Teltow, Germany) developed a bipolar RFA device (CelonPOWER System) in order to overcome the disadvantages of monopolar RFA devices. Unlike a monopolar RFA system, the prime characteristic of this new device is its bipolar feature, i.e., two electrodes are located on the same needle (Fig. 1a, b), allowing electricity flow only between the electrodes at the treatment target site, eliminating both the need for a grounding pad and the danger of burns (Fig. 2a, b).

The bipolar characteristics of the CelonPOWER system ensure the return of power to the device, and the simultaneous use of multiple applicators yields an extensive ablated area in a single treatment, which can reduce treatment time and the burden on the patient. This eliminates the need for repeated reinsertion of single monopolar needles to perform overlapping ablation. Another advantage of the bipolar device is that electric current is immediately retrieved, preventing it from flowing to unintended sites. The CelonPOWER System was awarded the Conformite Europeenne (CE) mark in 2003, and since then its use has spread mainly in Europe [11–17].

The CelonPOWER System consists of a high-frequency power generator, a water pump, and computerized applicators for regulation of the current frequency. The basic

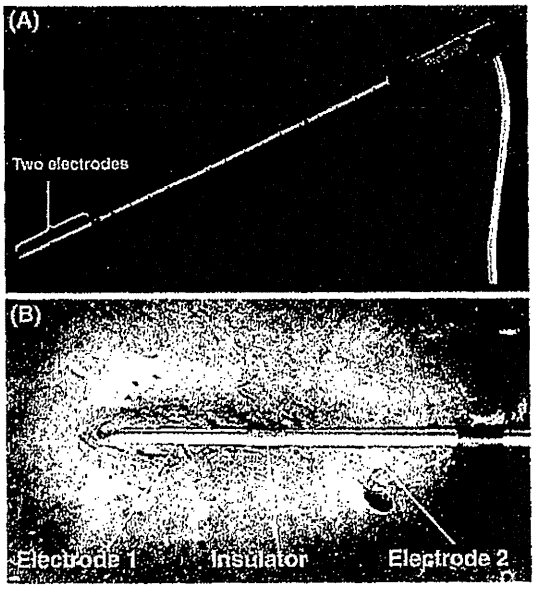

Fig. 1 In the CelonPOWER System, each applicator is needle-shaped and has two electrodes near its tip

J Gastroenterol (2013) 48:874–883
DOI 10.1007/s00535-012-0685-x

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

Clinical effectiveness of bipolar radiofrequency ablation for small liver cancers

Yukio Osaki · Kenji Ikeda · Namiki Izumi · Satoyoshi Yamashita · Hiromitsu Kumada · Shinji Hatta · Kiwamu Okita

Received: 2 May 2012 / Accepted: 7 September 2012 / Published online: 10 October 2012

Abstract
Background Radiofrequency ablation (RFA) is minimally invasive and can achieve a high rate of cure of liver cancer. This study was conducted to evaluate the efficacy and safety of a bipolar RFA device (CelonPOWER System) in the treatment of Japanese liver cancer patients.
Methods The study was a multicenter, single-group, open-label trial. The indications for RFA were based on the Japanese guidelines for the management of liver cancer. The subjects had a Child-Pugh classification of A or B, and the target tumors were defined as nodular, numbering up to 3 lesions, each of which was 3 cm or less in diameter, or solitary lesions up to 4 cm in diameter. To test for the non-inferiority of the CelonPOWER System, this system was compared with the Cool-tip RF System, which has already been approved in Japan, in terms of the complete necrosis rate (CNR).
Results The CNR obtained with the CelonPOWER System was 97.8 % (88/90 patients). The CNR obtained with the Cool-tip RF System was 86.2 % (50/58 patients), confirming the non-inferiority of the CelonPOWER System ($p < 0.001$, Fisher's exact test based on binomial distribution). Throughout the treatment and follow-up periods, there were no adverse events regarding safety that were uniquely related to the CelonPOWER System and there were no cases of device failure.
Conclusions The CelonPOWER System was confirmed to be an effective and safe RFA device. It could become extensively used as a safe next-generation RFA device, reducing the physical burden on patients.

Keywords Small hepatocellular carcinoma · Radiofrequency ablation (RFA) · Bipolar RFA · Conformite Europeenne (CE) mark · Non-inferiority to monopolar RFA

Y. Osaki
Department of Gastroenterology and Hepatology, Osaka Red Cross Hospital, Osaka, Japan

K. Ikeda · H. Kumada
Department of Hepatology, Toranomon Hospital, Tokyo, Japan

N. Izumi
Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

S. Yamashita · K. Okita (✉)
Department of Hepatology, Center for Liver Disease, Social Insurance Alliance Shimonoseki Kohsei Hospital, Kami-Shinchi Cho 3-3-8, Shimonoseki, Yamaguchi 750-0061, Japan
e-mail: k.okita@kousei-h.jp; k.okita@www.kousei-h.jp

S. Hatta
Olympus Medical Systems Corp, Tokyo, Japan

Introduction

According to a report of the Japanese Ministry of Health, Labor and Welfare in 2010, the number of deaths due to malignancies, including hepatocellular carcinoma (HCC), which is the most common type of primary liver cancer, has tended to increase annually [1]. In the 2007 report of the Japanese Ministry of Health, Labor and Welfare, the mortality of liver cancer was the 3rd highest among malignant diseases, following gastric cancer and lung cancer [2]. HCC appears in cirrhotic liver, and cirrhotic liver often results from alcohol abuse or chronic hepatitis B virus (HBV) or HCV infection. The presence of liver cirrhosis limits HCC treatment options, because surgery and systemic chemotherapy impair residual liver function and can induce fatal liver failure. In addition,

Table 1. Patient Characteristics and Demographics

Characteristics	Entire Cohort				Propensity Score Matched Cohort		
	All Patients (n = 1,615)	Entecavir (n = 472)	Control (n = 1,143)	P	Entecavir (n = 316)	Control (n = 316)	P
Age (y)†	42 (13.5)	47 (12.4)	39 (13.1)	<0.001	46 (12.1)	46 (13.5)	0.907
Gender (male:female)	1,035:580	315:157	720: 423	0.171	210:106	210:106	1.000
Alcohol consumption (>200kg)	355 (22)	97 (20.5)	288 (25.1)	0.013	62 (20)	105 (33)	<0.001
Cigarette smoking	443 (27)	157 (33.2)	286 (25.0)	0.005	110 (35)	110 (35)	1.000
Preexisting cirrhosis	311 (19)	116 (25)	195 (17)	0.001	79 (25)	85 (29)	0.324
HBV genotype	–	–	–	<0.001	–	–	0.843
A	53 (3.3)	12 (2.5)	41 (3.6)	–	8 (2.5)	9 (2.8)	–
B	254 (15.7)	66 (14.0)	188 (16.4)	–	49 (15.5)	50 (15.8)	–
C	1,135 (70.3)	344 (72.9)	791 (69.2)	–	225 (71.2)	226 (71.5)	–
D	1 (0.06)	0	1 (0.09)	–	0	0	–
F	1 (0.06)	0	1 (0.09)	–	0	0	–
H	2 (0.1)	2 (0.4)	0	–	0	0	–
Unclassified / missing	169 (10.4)	48 (10.2)	121 (10.5)	–	34 (10.7)	31 (9.8)	–
Baseline HBeAg positive	617 (38)	219 (46)	398 (35)	<0.001	135 (43)	133 (42)	0.936
Baseline HBV DNA (log copies/mL)	6.0 (4.3-7.7)	6.7 (5.3-8.0)	5.8 (4.0-7.5)	<0.001	6.3 (5.2-7.9)	6.6 (4.5-7.8)	0.795
Baseline AST level (IU/L)	35 (22-63)	53 (35-95)	28 (20-50)	<0.001	45 (32-70)	49 (27-98)	0.956
Baseline AST level (x ULN)	1.1 (0.7-1.9)	1.6 (1.1-2.9)	0.8 (0.6-1.5)	<0.001	1.4 (1.0-2.1)	1.5 (0.8-3.0)	0.989
Baseline ALT level (IU/L)	42 (22-88)	70 (42-163)	33 (20-68)	<0.001	61 (39-109)	60 (28-144)	0.110
Baseline ALT level (x ULN)	1.1 (0.7-2.4)	1.9 (1.2-4.3)	0.9 (0.6-1.8)	<0.001	1.7 (1.0-3.3)	1.6 (0.8-3.7)	0.086
Baseline GGTP level (IU/L)	28 (16-59)	39 (24-72)	24 (14-52)	<0.001	34 (23-64)	34 (18-68)	0.088
Baseline total bilirubin level (mg/dL)	0.7 (0.5-0.9)	0.7 (0.5-1.0)	0.6 (0.5-0.9)	<0.001	0.7 (0.5-1.0)	0.7 (0.5-0.9)	0.210
Baseline serum albumin level (g/L)	4.2 (3.9-4.5)	3.9 (3.6-4.1)	4.4 (4.1-4.6)	<0.001	3.9 (3.7-4.2)	4.0 (3.8-4.3)	0.084
†Platelet count (10^5/mm^3) (SD)	19.1 (6.3)	16.9 (5.6)	20.0 (6.4)	<0.001	17.5 (5.2)	17.2 (6.0)	0.349
Follow-up duration (yrs)	5.4 (3.1-13.2)	3.2 (2.1-4.3)	9.5 (4.4-16.1)	<0.001	3.3 (2.3-4.3)	7.6 (3.4-13.7)	<0.001
Person-years of follow-up	13,986	1561	12381	–	1064	2978	–
No. of HCC cases	156	12	144	–	6	72	–
Incidence rates per 1000 person-years	11.15	7.69	11.63	–	5.63	24.1	–
Progression of cirrhosis within 5 year	21 (1.3)	0	21 (1.8)	0.001	0	10 (3.2)	0.001
HBV DNA <400 copies/mL at 1 year	–	421 (89)	NA	–	288 (90)	NA	–
Emergence of drug-resistant mutants during ETV treatment	–	4 (0.8)	NA	–	2 (0.6)	NA	–

HBeAg, hepatitis B e antigen; HBV, hepatitis B virus; AST, aspartate aminotransferase; GGTP, gamma glutamyltransferase (ULN=33 IU/L); ALT, alanine aminotransferase (ULN=42 IU/L for men and 27 IU/L for women); HCC, hepatocellular carcinoma; ETV, entecavir.

*P <0.05.

**P <0.001, comparison of entecavir-treated group and control group.

†Data displayed as mean ± standard deviation. ‡All other values are expressed as median (25th to 75th percentile) or number (percentage of total, %).

matched control group were 4.0% at year 2, 7.2% at year 3, 10.0% at year 4, and 13.7% at year 5. Log-rank test revealed a statistically significant difference between the incidence of HCC in the ETV group and the control group over time ($P < 0.001$) (Fig. 2). We then used Cox proportional regression analysis to estimate the effects of ETV treatment on HCC risk. Factors that were associated with HCC at year 5 in the propensity score matched cohort were age, gender, alcohol consumption (>200 kg), the presence of cirrhosis, HBeAg positivity, baseline viral load, ALT, γ-GTP, total bilirubin, serum albumin, and platelet counts (Table 2). For ETV treatment effect, we estimated the hazard ratio of HCC development, adjusting for multiple baseline variables (age, gender, alcohol consumption, smoking, preexisting cirrhosis, HBeAg, HBV DNA, ALT, albumin, γ-GTP, total bilirubin, and platelet count) in the propensity matched cohort. Progression of cirrhosis within 5 years was used as a time-dependent covariate in the proportional hazard regression but it did not show a statistically significant hazard to HCC development.

Subanalyses Showing HCC Suppression Effect Between ETV and LAM. PS matching of the LAM-treated patients without rescue therapy (n = 492) with ETV-treated patients resulted in a matched cohort of 182 patients (Supporting Table 3). The rate of non-rescued LAM-treated group having undetectable HBV DNA at 1 year after treatment was lower when compared with the ETV-treated group. The LAM-treated group also had a higher drug-resistant mutation rate. Comparisons of HCC incidence among the ETV-treated group, nonrescued LAM-treated group, and control showed that the HCC suppression effect was greater in ETV-treated ($P < 0.001$) than nonrescued LAM-treated ($P = 0.019$) when compared with the

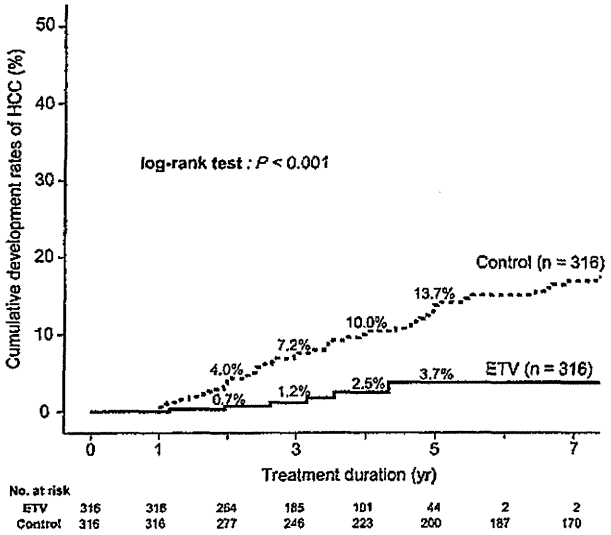

Fig. 2. Comparison of HCC cumulative incidence rates between the entecavir-treated group and the nontreated control group after propensity score matching. The log-rank test revealed a statistically significant difference between the ETV and the control group in the incidence of HCC at 5 years time (log-rank test: $P < 0.001$).

control group (Fig. 3). The difference of effect between ETV and LAM was also significant ($P = 0.043$). The treatment effect was seen in cirrhosis patients but not in noncirrhosis patients. The result showed ETV's superiority to LAM in suppressing HCC.

Effect of ETV on the Reduction of HCC Development by Preexisting Cirrhosis and Risk Scores. To further examine the ETV treatment effect, we compared the ETV and the control groups by preexisting cirrhosis and published risk scores. Viral response rates (HBV DNA < 400 copies/mL) of 1-year post-ETV treatment was 87% in the noncirrhosis patients and 91% in the cirrhosis patients (LC). ALT normalization was 94% and 90% in the chronic hepatitis and cirrhosis patients, respectively. The treatment effect was not inferior by cirrhosis status. Among those who developed HCC, 97 out of 144 patients in the control group and 9 out of 12 patients in the ETV group had cirrhosis. Interactions between preexisting cirrhosis and ETV treatment were not observed ($P = 0.177$).

Cumulative HCC incidence rates by risk scores are compared between the two cohorts in Fig. 4A-G. Figure 4A,B shows the risk scores developed by Yang et al.[10] Figure 4C,D shows the risk scores developed by Yuen et al.[11] Figure 4E-G shows the risk scores developed by Wong et al.[12] All three risk score scales showed that ETV significantly reduced HCC incidence in patients with a higher risk (risk score ≥ 12, $P = 0.006$; risk score ≥ 82, $P = 0.002$; medium risk, $P = 0.062$; high risk, $P < 0.001$). Interactions between risk scores and ETV treatment were not observed (Yang et al.: $P = 0.713$, Yuen et al.: $P = 0.267$, Wong et al.: $P = 0.265$).

Discussion

Our study suggests that long-term ETV therapy would significantly suppress the development of HCC in HBV-infected patients when compared with HBV-infected patients in the control group. The treatment effect was more prominent among patients at high risk of HCC than those at low risk.

Table 2. Factors Associated with HCC Development as Determined by Cox Proportional Hazard Regression Analysis at 5-Year (Propensity Score Matched Cohort)

Variable	Univariate HR (95% CI)	P	Multivariate Adjusted HR (95% CI)	P
Age (per year)	1.05 (1.02-1.07)	<0.001	1.06 (1.03-1.09)	<0.001
Gender (M)	2.81 (1.25-6.32)	0.012		
Alcohol consumption (>200kg)	2.71 (1.49-4.92)	0.001	2.21 (1.18-4.16)	0.013
Cigarette smoking	1.53 (0.84-2.80)	0.164		
Preexisting cirrhosis	12.0 (5.57-25.9)	<0.001	4.28 (1.88-9.73)	0.001
HBV genotype (C)	2.73 (0.98-7.65)	0.056		
HBeAg (positive)	2.64 (1.41-4.94)	0.002	2.26 (1.18-4.34)	0.014
HBV DNA (≥5.0 log copies/mL)	4.66 (1.44-15.1)	0.010		
ALT (≥45 IU/L)	2.29 (1.10-4.77)	0.027		
GGTP (≥50 IU/L)	3.79 (2.02-7.09)	<0.001		
Total bilirubin (≥1.5 mg/dL)	5.51 (2.87-10.6)	<0.001		
Serum albumin (<3.8 g/L)	4.44 (2.42-8.14)	<0.001		
Platelet count ($<1.5\times10^5$/mm^3)	14.8 (5.84-37.7)	<0.001	5.64 (2.13-15.0)	0.001
*Progression of cirrhosis within 5 years	1.80 (0.25-13.2)	0.562		
ETV treatment	0.23 (0.09-0.55)	0.001	0.37 (0.15-0.91)	0.030

Asterisks (*) indicate time-dependent covariates.

†Adjusted for age, gender, alcohol, cigarette, cirrhosis, genotype, HBeAg, HBV DNA, ALT, albumin, GGTP, total bilirubin, and platelet counts

Abbreviations: ETV, entecavir; HR, hazard ratio; CI, confidence interval; HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; ALT, alanine aminotransferase; GGTP, gamma glutamyltransferase.

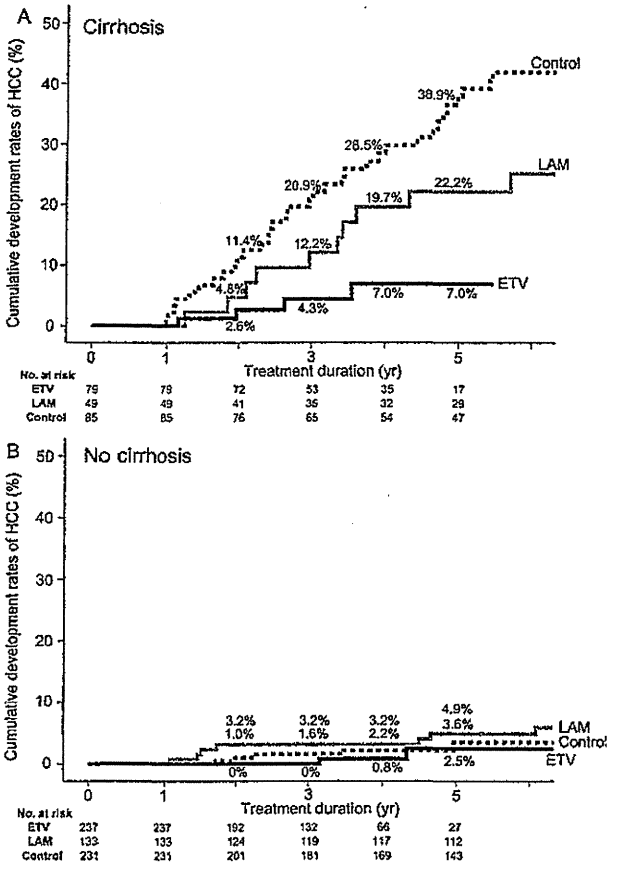

Fig. 3. Comparison of HCC cumulative incidence rates between the entecavir (ETV)-treated group, lamivudine (LAM)-treated, and the nontreated control group after PS matching stratified by cirrhosis. The log-rank test revealed a statistically significant difference in the incidence of HCC at 5 years time in cirrhosis patients: ETV versus control group ($P < 0.001$); LAM versus control ($P = 0.019$); ETV versus LAM ($P = 0.043$). The differences were not seen in the noncirrhosis patients: ETV versus control ($P = 0.440$); LAM versus control ($P = 0.879$); ETV versus LAM ($P = 0.126$).

HBV has been previously shown to influence HCC development. Ikeda et al.[20] reported that the cumulative HCC incidence rates among Japanese HBV patients were 2.1% at 5 years, 4.9% at 10 years, and 18.8% at 15 years among NA-naïve patients. Other studies, both from Japan and other countries, have reported a 5-year cumulative HCC incidence rate of 3.3% among chronic HBV, and 21.2% to 59% among cirrhosis patients.[21,22] The incidence of HCC varies significantly by country and ethnic group,[4] which seems to be attributable to diverse exposure to HCC risk factors.

Carcinogenicity related to HBV infection is somewhat complex and multifactorial when compared with carcinogenicity related to HCV infection. Known HCC risk factors among HBV-infected patients include older age, male gender, cirrhotic status, diabetes mellitus, family history, alcohol consumption, AST, HBsAg, HBeAg, and genotype C.[20,23,25] Chen et al.[5] found a dose-response relationship between pretreatment serum HBV DNA levels and the development of HCC. Baseline ALT is another risk factor for HCC, as elevated ALT levels indicate an active immune response against HBV, resulting in repetitive hepatocyte injury.[5] Our study corroborates these findings on these factors influence on HCC development.

The potential ability of ETV to reduce the risk of HCC is an additional example of a long-term NA treatment effect. Some studies have shown that ETV has low incidence of HCC but these studies did not have a control arm.[9] A meta-analysis and a systematic review showed that NAs can reduce liver complications, including HCC.[26,27] Other studies have begun to show that control of sustained viral loads through drugs such as NAs is important in preventing long-term complications. Chen et al.[28] showed that greater decreases in serum HBV DNA levels ($<10^4$ copies/mL) during follow-up were associated with a lower risk of HCC.

Our comparison among the PS-matched ETV-treated group, nonrescued LAM-treated patients, and the control showed that ETV is superior to LAM in HCC suppression. Kurokawa et al.[29] showed that treatment with lamivudine for an average of 5 years reduced the incidence of HCC in HBV-infected cirrhosis patients, who showed sustained viral response at a median HBV DNA of <4.0 log copies/mL. Unfortunately, only 48% of the patients in this study achieved sustained viral response, while 51% developed lamivudine-resistant tyrosine-methionine-aspartate-aspartate mutation (YMDD mutation) during follow-up.[29] Patients with drug resistance were reported to have a 2.6 times greater chance of developing long-term complications.[26] A systematic review of 21 studies showed that HCC occurred more (2.3% versus 7.5%, $P < 0.001$) in nonresponding patients or in patients with viral breakthrough compared with those who experienced remission.[28] On-treatment drug resistance could subject patients to a variable viral status. Suppression of HCC by NAs requires NAs that do not lead to drug resistance. Compared with other NAs, ETV shows minimal drug resistance. Our results showed that ~90% of the ETV-treated patients had sustained viral suppression at year 1, and that drug resistance was minimal (0.8%) during the median follow-up period of 3.2 years.

We found that the effect of ETV treatment in reducing the risk of HCC was more prominent among high-risk patients. This phenomenon was observed by examining the combination of parameters associated with the recently developed risk scores (Fig. 4). The published risk scores were developed mainly to create

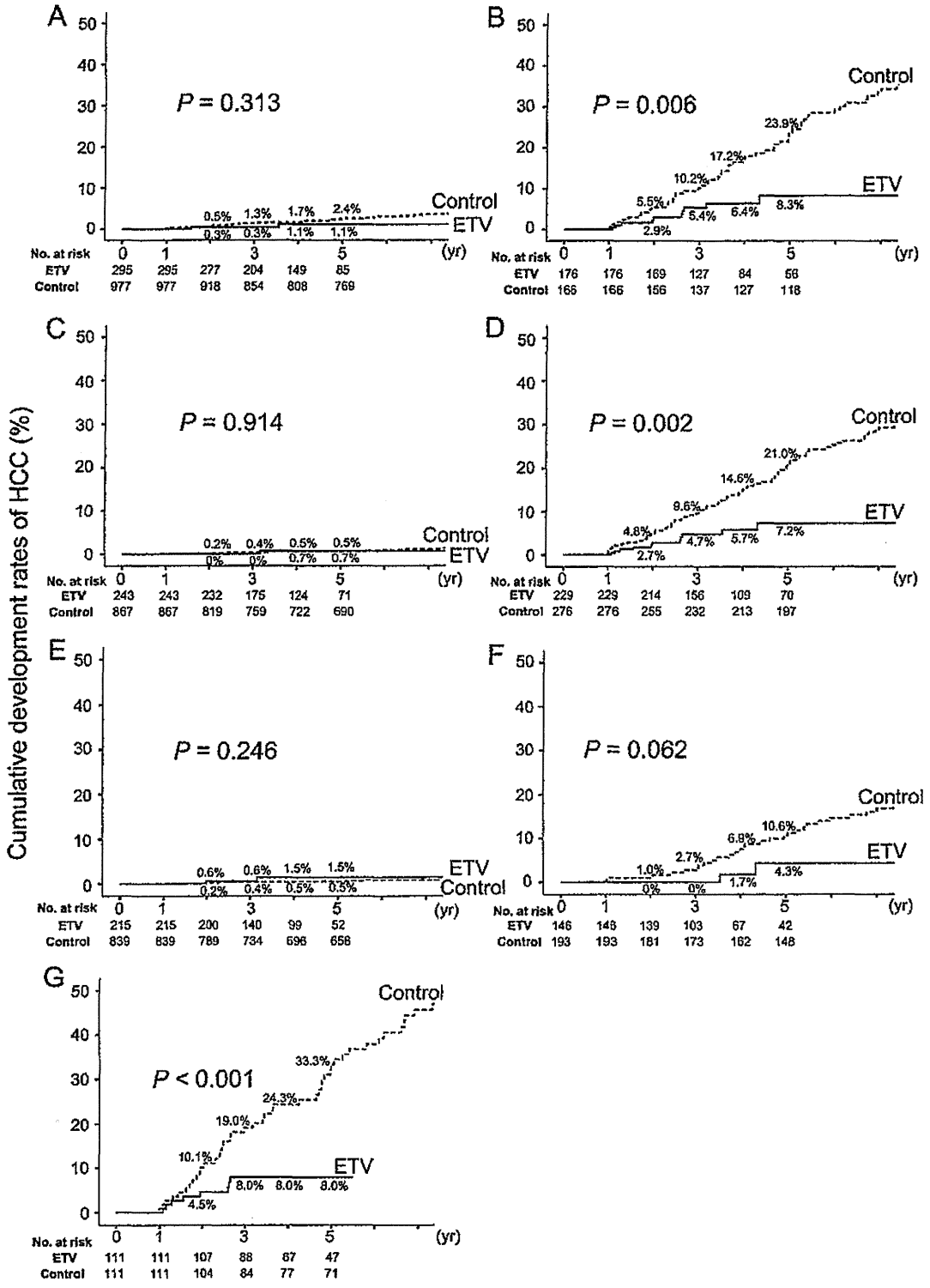

Fig. 4. Cumulative incidence of HCC by risk score scales: comparison between entecavir-treated and nontreated control patients: Risk score cutoff points were based on those presented in articles by the following: A,B (Yang et al.[10]): low-risk score cutoff point < 12; high-risk score cutoff point ≥ 12. C,D (Yuen et al.[11]): low-risk score cutoff point < 82; high-risk score cutoff point ≥ 82. E-G (Wong et al.[12]): low-risk score cutoff point < 4; medium-risk cutoff point 4-19; high-risk score cutoff point ≥ 20. A statistically significant difference in HCC incidence was seen between the ETV group and the control group in the higher-risk groups when observed the incidence of HCC over time (log-rank test $P = 0.006$ for risk score ≥ 12; $P = 0.002$ for risk score ≥ 82; $P = 0.062$ for patients with medium risk; $P < 0.001$ for patients at high risk for HCC).

easy-to-use nomograms based on clinical characteristics to predict the risk of HCC in patients with HBV. These scales have been validated, and can accurately estimate the risk of HCC up to 10 years. The cutoff scores used in these studies were based on their sensitivity to detect HCC derived and validated with non-treated HBV cohorts. The importance of our study using these risk scales in our cohorts was to see the change in risk with the initiation of therapy. We found that the ETV treatment effect to reduce the risk of HCC was more prominent among cirrhosis and high-risk patients despite the lack of interactions between ETV treatment and preexisting cirrhosis or risk factors. The lower treatment effect among lower-risk patients was somewhat not surprising. HCC development among low-risk patients is generally rare, and therefore, the treatment effect may not have occurred in large enough numbers during the treatment period allotted in our study to be able to detect a difference. In addition, HCC development differs greatly by cirrhotic status and risk factors in the control group. The treatment effect of ETV to reduce HCC is probably more likely reflected among cirrhosis or high-risk patients. A study with a longer observation period and higher patient numbers might be necessary to examine this ETV treatment effect among low-risk patients. The development of a scoring system to predict treatment effect of HBV patients with different risk levels will be useful in determining the most appropriate timing of treatment initiation in clinical settings.

Study Limitations. There were several limitations to our study. First, because our patients were recruited from one hospital, they might not have been representative of the general Japanese HBV population. Second, our control group included historically observed patients who entered the cohort long before the ETV group, resulting in treatment differences during the time gap. However, we used PS matching and a similar follow-up period between the two cohorts to minimize this bias. Third, our study was an observational study with patients having large demographic differences. Although we used a PS to match ETV-treated and control groups, our sample size did not take into account other unobserved confounding factors such as HCC family history, stage of cirrhosis, and comorbidities when determining associating factors for carcinogenesis in HBV. Finally, the observation period of the ETV group was relatively short, and patients in the ETV-treated cohort at 5 years consisted of only less than ~25% of the initial recruited patients. Because of this limitation, we censored patients who were followed for more than 5 years. The observed treatment effect would require confirmation over a longer period and a more complete follow-up.

Conducting a long-term study to examine the effect of antiviral therapy with HCC as the endpoint would be time-consuming and challenging. Such a study would require a large sample size and would, therefore, be costly. In addition, the increases in choices of therapy over time would make it difficult to conduct a long-term study using a single therapy. Owing to ethical issues, it would be difficult to recruit or follow a naïve, untreated cohort over an extended period of time. Because of these challenges, most studies have examined the relationship between antiviral treatment and the risks of HCC involved older drugs, lacked a control group, or were of relatively short duration. Consequently, the association between antiviral treatment and carcinogenesis is inferential and requires additional confirmatory studies.

In conclusion, in our study we observed the effect of HCC risk among HBV-infected patients treated by ETV by comparing them with a group of NA-naïve patients. We followed these Japanese patients for a relatively long period of time and compared them with a large pool of untreated control patients. In this long-term study among Japanese patients, ETV significantly reduced the incidence of HCC among chronic HBV-infected patients, and was more prominent among patients at higher risk for HCC.

Acknowledgment: We thank Fujio Matsuo, MSc, Executive Director, Statistical/Analysis Department, Statcom Company Limited and Yasuo Ohashi, PhD, Professor, Department of Biostatistics, School of Public Health, University of Tokyo for their review and advice on statistical analyses.

References

1. Lee WM. Hepatitis B virus infection. N Engl J Med 1997;337:1733-1745.
2. Lai CL, Ratziu V, Yuen MF, Poynard T. Viral hepatitis B. Lancet 2003;362:2089-2094.
3. Merican I, Guan R, Amarapuka D, Alexander MJ, Chutaputti A, Chien RN, et al. Chronic hepatitis B virus infection in Asian countries. J Gastroenterol Hepatol 2000;15:1356-1361.
4. Chen CJ, Yu MW, Liaw YF. Epidemiological characteristics and risk factors of hepatocellular carcinoma. J Gastroenterol Hepatol 1997;12:S294-S298.
5. Chen CJ, Yang HI, Jun S, Jen CL, You SL, Lu SN, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 2006;295:65-73.
6. Liaw YF, Sung JJY, Chow WC, Farrell G, Lee CZ, Yuen H, et al. Lamivudine for patients with chronic hepatitis B and advanced liver disease. N Engl J Med 2004;351:1521-1531.
7. Matsumoto A, Tanaka E, Rokuhara A, Kiyosawa K, Kumada H, Omata M, et al. Efficacy of lamivudine for preventing hepatocellular carcinoma in chronic hepatitis B: a multicenter retrospective study of 2795 patients. Hepatol Res 2005;32:173-184.

8. Yokosuka O, Takaguchi K, Fujioka S, Shindo M, Chayama K, Kobashi H, et al. Long-term use of entecavir in nucleoside-naïve Japanese patients with chronic hepatitis B infection. J Hepatol 2010;52:791-799.
9. Chang TT, Lai CL, Yoon SK, Lee SS, Coelho HSM, Carrilho FJ, et al. Entecavir treatment for up to 5 years in patients with hepatitis B e antigen-positive chronic hepatitis B. HEPATOLOGY 2010;51:422-430.
10. Yang HI, Yuen MF, Chan HLY, Han KH, Chen PJ, Kim DY, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. Lancet Oncol 2011;12:568-574.
11. Yuen MF, Tanaka Y, Fong DYT, Fung J, Wong DKH, Yuen JCH, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol 2009;50:80-88.
12. Wong VWS, Chan SL, Mo F, Chan TC, Loong HH, Wong GL, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. J Clin Oncol 2010;28:1660-1665.
13. Yang HI, Sherman M, Su J, Chen PJ, Liaw YF, Iloeje UH, et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. J Clin Oncol 2010;28:2437-2444.
14. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984; 79:516-524.
15. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. Ann Intern Med 2002;137:693-695.
16. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. J Am Stat Assoc 1985;39:33-38.
17. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-2281.
18. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 1988;16:1141-1154.
19. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496-509.
20. Ikeda K, Saitoh S, Suzuki Y, Kobayashi M, Tsubota A, Koida I, et al. Disease progression and hepatocellular carcinogenesis in patients with chronic viral hepatitis: a prospective observation of 2215 patients. J Hepatol 1998;28:930-938.
21. Kato Y, Nakata K, Omagari K, Furukawa R, Kusumoto Y, Mori I, et al. Risk of hepatocellular carcinoma in patients with cirrhosis in Japan. Cancer 1994;74:2234-2238.
22. Lo KJ, Tong MJ, Chien MC, Tsai YT, Liaw YF, Yang KC, et al. The natural course of hepatitis B surface antigen-positive chronic active hepatitis in Taiwan. J Infect Dis 1982;146:205-210.
23. Yang HI, Lu SN, Liaw YF, You SL, Sun CA, Wang LY, et al. Hepatitis B e antigen and the risk of hepatocellular carcinoma. N Engl J Med 2022;347:168-174.
24. Kao JH, Chen PJ, Lai MY, Chen DS. Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. Gastroenterology 2000;118:554-559.
25. Chen CJ, Yang HI, Iloeje UH. Hepatitis B virus DNA levels and outcomes in chronic hepatitis B. HEPATOLOGY 2009;49:S72-S84.
26. Zhang Q-Q, An X, Liu YH, Li SY, Zhong Q, Wang J, et al. Long-term nucleos(t)ide analogues therapy for adults with chronic hepatitis B reduces the risk of long-term complications: a meta-analysis. Virol J 2011;8:72.
27. Papatheodoridis GV, Lampertico P, Manolakopoulos S, Lok A. Incidence of hepatocellular carcinoma in chronic hepatitis B patients receiving nucleos(t)ide therapy: a systematic review. J Hepatol 2010;53: 348-356.
28. Chen CF, Lee WC, Yang HI, Chang HC, Jen CL, Iloeje UH, et al. Changes in serum levels of HBV DNA and alanine aminotransferase determine risk for hepatocellular carcinoma. Gastroenterology 2011;141: 1240-1248.
29. Kurokawa M, Hiramatsu N, Oze T, Yakushijin T, Miyazaki M, Hosui A, et al. Long-term effect of lamivudine treatment on the incidence of hepatocellular carcinoma in patients with hepatitis B virus infection. J Gastroenterol 2012;47:577-585.

Hepatology Research

Hepatology Research 2013; 43: 596–604 doi: 10.1111/j.1872-034X.2012.01115.x

Original Article

Fibrosis score consisting of four serum markers successfully predicts pathological fibrotic stages of chronic hepatitis B

Kenji Ikeda,[1,2] Namiki Izumi,[3] Eiji Tanaka,[7] Hiroshi Yotsuyanagi,[4] Yoshihisa Takahashi,[5] Junichi Fukushima,[6] Fukuo Kondo,[5] Toshio Fukusato,[5] Kazuhiko Koike,[4] Norio Hayashi[8] and Hiromitsu Kumada[1,2]

[1]*Department of Hepatology, Toranomon Hospital,* [2]*Okinaka Memorial Institute for Medical Research,* [3]*Department of Gastroenterology, Musashino Red Cross Hospital,* [4]*Department of Gastroenterology, Tokyo University of Medicine,*[5]*Department of Pathology, Teikyo University School of Medicine,* [6]*Department of Pathology, NTT Medical Center Tokyo, Tokyo,* [7]*Department of Gastroenterology, Shinshu University of Medicine, Matsumoto, and* [8]*Department of Gastroenterology, Kansai-Rosai Hospital, Hyogo, Japan*

Aim: In order to evaluate and judge a fibrotic stage of patients with chronic hepatitis B, multivariate regression analysis was performed using multiple fibrosis markers.

Method: A total of 227 patients from seven hepatology units and institutes were diagnosed by needle biopsy as having chronic liver disease caused by hepatitis B virus. Twenty-three variables and their natural logarithmic transformation were employed in the multivariate analysis. Multiple regression function was generated from data of 158 patients in one hospital, and validation was performed using the other data of 69 patients from six other hospitals.

Results: After stepwise variable selection, multivariate regression analysis finally obtained the following function: $z = 1.40 \times ln$ (type IV collagen 7S) (ng/mL) $- 0.017 \times$ (platelet count) ($\times 1000^3/mm^3$) $+ 1.24 \times ln$ (tissue inhibitor of matrix metalloproteinase-2) (ng/mL) $+ 1.19 \times ln$ (α-2-macroglobulin) (mg/dL) $- 9.15$. Median values of fibrosis scores of F1 ($n = 73$), F2 ($n = 42$), F3 ($n = 31$) and F4 stages ($n = 12$) were calculated as 0.95, 2.07, 2.98 and 3.63, respectively. Multiple regression coefficient and coefficient of determination were 0.646 and 0.418, respectively. Validation with patient data from other institutions demonstrated good reproducibility of fibrosis score for hepatitis B (FSB), showing 1.33 in F1 ($n = 27$), 2.20 in F2 ($n = 20$), 3.11 in F3 ($n = 20$) and 5.30 in F4 ($n = 2$), respectively.

Conclusion: A concise multiple regression function using four laboratory parameters successfully predicted pathological fibrosis stage of patients with hepatitis B virus infection.

Key words: chronic hepatitis, hepatitis B virus, liver cirrhosis, liver fibrosis, multiple regression analysis, stage

INTRODUCTION

WHEN HEPATITIS B virus (HBV)-related chronic liver disease is found by biochemical and virological examination, liver biopsy can establish the definitive diagnosis of chronic hepatitis and its fibrotic staging. Although these pathological procedures are reliable and informative both in diagnosis and treatment, they sometimes require medical invasion and financial costs, including the risk of bleeding from needle puncture, some pain experienced during the procedure and hospital stays of a few days. The pathological examination is, therefore, rarely performed repeatedly in a short period of time, unless disease activity is severe or progression of liver disease is highly suspected. Recently, many authors described the usefulness of ultrasonographic elastography and multiple resonance imaging technology in the estimation of staging of chronic hepatitis and cirrhosis.[1–5] These ways of estimation using the imaging apparatuses seem truly useful for current patients, but they cannot evaluate and compare with past fibrotic states of patients retrospectively. Moreover,

Correspondence: Dr Kenji Ikeda, Department of Hepatology, Toranomon Hospital, 2-2-2 Toranomon, Minato-ku, Tokyo, 105-8470, Japan. Email: ikedakenji@tora.email.ne.jp
Received 6 May 2012; revision 17 September 2012; accepted 4 October 2012.

the same apparatus for elastometry will not be available for repeated measurement for a follow-up examination, for example, several years later.

In spite of the accuracy of biopsy and convenience of elastography in chronic liver disease, clinical diagnosis based on biochemistry and hematology is still indispensable for the daily practice of many patients with HBV-related liver disease. Recently, several studies were published about estimation of hepatitis stages, using one or more serum biomarkers. Discriminant functions or multivariate analyses demonstrated that approximately 60–90% of patients with chronic hepatitis B were correctly classified as having mild hepatitis and severe hepatitis with advanced fibrosis.[2,6–13] Up to the present time, however, the usefulness of the discriminant functions are less valuable for a few reasons. First, these functions were made for the purpose of discrimination of severe hepatic fibrosis from mild fibrosis, and four histological classifications (F1–F4) were neglected in almost of the studies. Second, some studies analyzed both hepatitis B and hepatitis C virus infection, although the significance and actual values of each liver function test in the evaluation of the severity of liver disease were not similar among each viral hepatitis and alcoholic liver disease. Third, biochemical markers for liver fibrosis (e.g. hyaluronic acid, type IV collagen, procollagen III peptide)[14–16] were not always included in those previous studies.

We tried to generate a function estimating fibrotic stages of HBV-related chronic hepatitis, which were objectively diagnosed by liver biopsy. The purpose of this study is, therefore, to make a reliable multiple regression function and to obtain practical coefficients for significant variables also using fibrosis markers.

METHODS

Patients

A TOTAL OF 273 Japanese patients with chronic hepatitis B were recruited for the study from seven hospitals in Japan: Toranomon Hospital, Hiroshima University Hospital (K. Chayama, M.D.), Ehime University Hospital (M. Onji, M.D.), Musashino Red Cross Hospital (N. Izumi, MD), Shishu University Hospital (E. Tanaka, M.D.), Showa University Hospital (M. Imawari, M.D.) and Osaka University Hospital (T. Takehara, M.D.). Inclusion criteria for this study were: (i) positive hepatitis B surface antigen for more than 6 months; (ii) persistent or intermittent elevation in aspartate aminotransferase (AST)/alanine aminotransferase (ALT) levels; and (iii) liver biopsy showing chronic hepatitis (F1–F4). We excluded those patients with overt alcoholic liver disease or fatty liver, association of other types of liver disease (e.g. hepatitis C, primary biliary cirrhosis, autoimmune hepatitis), or those associated with hepatocellular carcinoma or other malignancy. Among the patients, 244 patients fulfilled the conditions for the study: complete demographic data, basic laboratory data of hematology and biochemistry, required liver biopsy specimens, and sufficient amount of frozen sera. Also, we excluded additional 17 patients with eventual histological diagnosis as F0 stage.

Finally, a total of 227 patients who were diagnosed as having chronic hepatitis or cirrhosis (F1–F4) were analyzed for the following hematological, biochemical and histopathological examination. There were 172 males and 55 females aged 16–70 years (median, 39 years).

All the patients presented written informed consent in individual hospitals and medical centers, and the study was approved in each ethical committee.

Hematological and biochemical examination

Hematological and standard biochemical evaluation had been performed in each medical institution: white blood cells, red blood cells, hemoglobin, platelets, total bilirubin, AST, ALT, AST/ALT ratio (AAR), γ-glutamyl transpeptidase (γ-GTP), total protein, albumin and γ-globulin.

Special biochemical examinations including "fibrosis markers" were carried out using stored frozen sera at −20°C or lower: α-2-macroglobulin, haptoglobin concentration, haptoglobin typing, apolipoprotein A1, hyaluronic acid, tissue inhibitor of matrix metalloproteinase (TIMP)-1, TIMP-2, procollagen III peptide and type IV collagen 7S.

Histological diagnosis of chronic hepatitis and cirrhosis

All the 227 cases fulfilled required standards of histological evaluation: sufficient length of specimen, hematoxylin–eosin staining, and at least one specimen with fiber staining. Four independent pathologists (Y. T., J. F., F. K. and T. F.), who were not informed of patients' background and laboratory features except for age and sex, evaluated the 227 specimens regarding the stages of fibrosis and activity. Pathological classification of chronic hepatitis staging was based on Desmet *et al.*[17]

Before judgment of histological staging of individual specimens, the pathologists discussed the objective and reproducible judgment of pathological diagnosis of

hepatitis. They made a panel about obvious criteria using typical microscopic pictures for each stage, and it was always referred to during the procedure of pathological judgment. When inconsistent results were found in the diagnosis of hepatitis stage among the pathologists, the final judgment accepted majority rule among them.

Statistical analysis

Non-parametric procedures were employed for the analysis of background characteristics and laboratory data among patients in each stage, including Mann–Whitney *U*-test, Kruskal–Wallis test and χ^2-test.

The normality of the distribution of the data was evaluated by a Kolmogorov–Smirnov one-sample test. Because certain variables partly did not conform to a normal distribution, natural logarithmic transformation of bilirubin, AST, ALT, γ-GTP, α-2-macroglobulin, hyaluronic acid, type IV collagen 7S and TIMP-2 were also analyzed in the following calculation. The natural logarithmic transformation of the results yielded a normal distribution or symmetrical distribution for all the analyzed factors. After the procedures, the following multiple regression analysis became rationally robust against deviations from normal distribution. In order to avoid introducing into the model any variables that were mutually correlated, we checked the interaction between all pairs of the variables by calculating variance inflation factors. Of the highly correlated variables, less significant factors were removed from the viewpoint of multicollinearity.

Multivariate regression analysis was performed using 158 patient data from Toranomon Hospital (training dataset) to generate a training data of predicting function. We used a stepwise method for selection of informative subsets of explanatory variables in the model. Multiple regression coefficient and coefficient of determination were also taken into account in the selection of variables. Next, we validated the obtained predictive function using the remaining 69 patient data from the other six liver institutions (validation dataset).

A *P*-value of less than 0.05 with two-tailed test was considered to be significant. Data analysis was performed using the computer program SPSS ver. 19.[18]

For evaluation of the efficiency and usefulness of obtained function for fibrosis estimation, we compared various fibrosis scores for hepatitis B and C, including AAR,[19] AST-to-platelet ratio index (APRI),[20] FIB-4,[21] FibroTest[22] and discrimination function of cirrhosis from hepatitis in Japanese patients.[23]

RESULTS

Pathological diagnosis

FOUR PATHOLOGISTS INDEPENDENTLY judged the fibrotic stages and inflammatory activity for 227 specimens of chronic hepatitis/cirrhosis caused by HBV. One hundred patients (44.1%) had a fibrosis stage of F1, 62 (27.3%) F2, 51 (22.5%) F3 and 14 (6.2%) F4. In the subgroup of the 158 patients in the training group, judgment as F1 was made in 73 cases, F2 in 42, F3 in 31 and F4 in 12. Of the 69 patients in the validation group, judgment as F1 was made in 27, F2 in 20, F3 in 20 and F4 in two.

According to hepatitis activity classification, A0 was found in five (2.2%), A1 in 100 (44.1%), A2 in 107 (47.1%) and A3 in 15 (6.6%).

Laboratory data of each hepatitis stage in the training group

There were 124 men and 34 women with a median age of 39 years ranged 16–70 years. Laboratory data of 158 patients in the training group are shown in Table 1. Although several individual items were well correlated with the severity of hepatic fibrosis, significant overlap values were noted among F1–F4 stages: platelet count, γ-globulin, α-2-macroglobulin, haptoglobin, hyaluronic acid, TIMP-2 and type IV collagen 7S.

Significant variables serving staging of hepatitis

Univariate analyses using trend analysis with the Cochran–Armitage method showed that the fibrotic stage of chronic hepatitis B (FSB) was significantly correlated with platelet count (Spearman: $r = -0.45$, $P < 0.001$), γ-GTP ($r = 0.19$, $P = 0.017$), γ-globulin ($r = 0.29$, $P < 0.001$), α-2-macroglobulin ($r = 0.32$, $P < 0.001$), hyaluronic acid ($r = 0.36$, $P < 0.001$), TIMP-2 ($r = 0.16$, $P = 0.043$), procollagen III peptide ($r = 0.30$, $P < 0.001$) and type IV collagen 7S ($r = 0.55$, $P < 0.001$).

Regression function generated from training patient group

After stepwise variable selection, multivariate regression analysis finally obtained the following function: $z = 1.40 \times ln$ (type IV collagen 7S) (ng/mL) $- 0.017 \times$ (platelet count) ($\times 1000^3/mm^3$) $+ 1.24 \times ln$ (TIMP-2) (ng/mL) $+ 1.19 \times ln$ (α-2-macroglobulin) (mg/dL) $-$ 9.15. Median values of the fibrosis score of F1 ($n = 73$), F2 ($n = 42$), F3 ($n = 31$) and F4 stages ($n = 12$) were calculated as 0.95, 2.07, 2.98 and 3.63, respectively

Table 1 Demography and laboratory data of 158 patients in training group

	F1 ($n=73$)	F2 ($n=42$)	F3 ($n=31$)	F4 ($n=12$)
Demographics				
Men : women	58:15	33:9	23:8	10:2
Age (median, range)	36 (16–70)	39.5 (18–66)	39 (25–64)	43 (32–59)
Laboratory data (median, range)				
WBC (×1000/mm^3)	5.4 (2.5–10.6)	5.1 (2.4–8.7)	4.9 (3.0–8.7)	4.1 (3.7–6.6)
Hemoglobin (g/dL)	15.3 (10.3–18.8)	15.4 (12.5–17.9)	15.2 (11.5–17.2)	14.45 (12.1–18.2)
Platelet (×1000/mm^3)	204 (124–341)	173 (82–308)	155 (96–220)	130 (86–230)
Albumin (g/dL)	4.1 (3.2–4.9)	4.0 (3.2–5.1)	4.0 (3.3–4.9)	3.95 (3.4–4.6)
Bilirubin (mg/dL)	0.8 (0.2–1.7)	0.8 (0.3–2.3)	0.9 (0.4–5.4)	0.85 (0.6–2.3)
AST (IU/L)	48 (16–450)	55 (17–588)	54 (17–1446)	76.5 (27–396)
ALT (IU/L)	102 (10–839)	90 (12–886)	85 (19–2148)	89 (18–809)
γ-GTP (IU/L)	37 (7–247)	55 (8–687)	44 (14–564)	69 (33–262)
γ-Globulin (g/dL)	1.29 (0.78–2.11)	1.495 (0.62–3.20)	1.43 (0.90–2.30)	1.735 (0.92–2.47)
γ-Globulin (%)	17.3 (10.8–26.1)	19.3 (8.5–35.6)	19.9 (12.9–28.6)	22.55 (13.9–30.2)
α-2-Macroglobulin (mg/dL)	226 (116–446)	276 (148–495)	261 (202–565)	286.5 (166–425)
Haptoglobin (mg/dL)	77 (<5–318)	59 (<5–238)	61 (<5–151)	48.5 (<5–145)
Apolipoprotein A-I (mg/dL)	134 (89–212)	143 (78–250)	133 (87–189)	125 (73–169)
Hyaluronic acid (μg/L)	16 (<5–130)	32.5 (<5–204)	38 (<5–418)	49 (24–335)
TIMP-1 (ng/mL)	168 (93–271)	172 (116–314)	157 (119–365)	192 (145–365)
TIMP-2 (ng/mL)	80 (41–135)	80.5 (35–121)	92 (38–251)	85.5 (70–123)
Procollagen III peptide (U/mL)	0.75 (0.53–1.90)	0.835 (0.45–1.20)	0.89 (0.58–2.50)	1.05 (0.71–2.20)
Type IV collagen 7S (ng/ml)	4.0 (2.7–7.7)	4.6 (2.6–9.6)	5.6 (2.3–15.0)	7.2 (4.2–14.0)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; γ-GTP, γ-glutamyl transpeptidase; TIMP, tissue inhibitor of matrix metalloproteinase; WBC, white blood cells.

(Fig. 1). The multiple regression coefficient and coefficient of determination were 0.646 ($P<0.001$) and 0.418 ($P<0.001$), respectively.

Because the generated regression function was obtained by multivariate analysis with stepwise variable selection, several variables were removed from the function due to multicollinearity among them. Mutual correlation among the fibrosis predictors are shown in Table 2.

A 28-year-old man of F1 fibrotic stage (Fig. 2a) had a serum type IV collagen concentration of 4.4 ng/mL, platelet 221×10^3 count/mm^3, TIMP-2 75 ng/mL and α-2-macroglobulin 226 mg/dL. The regression function provided a fibrosis score of 0.99. Another man aged 46 years had F3 fibrosis on histological examination (Fig. 2b). His type IV collagen was 5.3 ng/mL, platelet 137×10^3 count/mm^3, TIMP-2 92 ng/mL and α-2-macroglobulin 255, and the regression function calculated his fibrosis score as 3.10.

Validation of discriminant function

Validation data of 69 patients (Table 3) were collected from the other six institutions in Japan. When applying

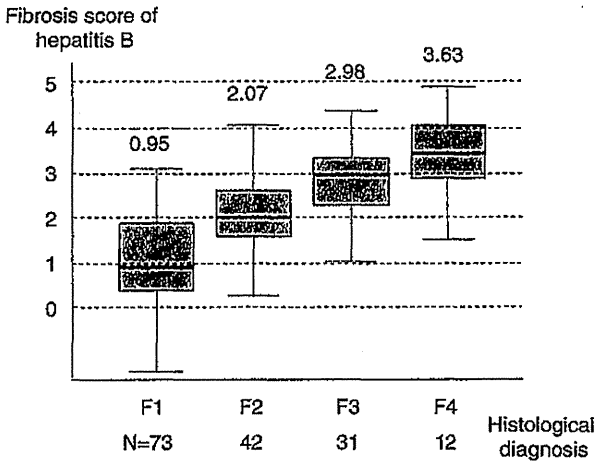

Figure 1 Box and whisker plots of fibrotic score of each histological fibrosis group in the training dataset. The fibrosis score of hepatitis B was generated by the function, $z = 1.40\times ln$ (type IV collagen 7S) (ng/mL) − $0.017\times$ (platelet count) (×1000^3/mm^3) + $1.24\times ln$ (tissue inhibitor of matrix metalloproteinase-2) (ng/mL) + $1.19\times ln$ (α-2-macroglobulin) (mg/dL) − 9.15.

Table 2 Correlation coefficients (Spearman's ρ) among fibrosis predictors used in multivariate analysis

	Platelet	gamma-globulin	*ln* (α-2-macroglobulin)	*ln* (hyaluronate)	*ln* (P-III-P)	*ln* (IV collagen)	*ln* (TIMP-2)
Platelet ($\times 10^3/mm^3$)	1.000	−0.214 ($P = 0.008$)	−0.260 ($P = 0.001$)	−0.384 ($P < 0.001$)	−0.045 ($P = 0.58$)	−0.297 ($P < 0.001$)	0.094 ($P = 0.24$)
γ-Globulin (g/dL)		1.000	0.276 ($P = 0.001$)	0.349 ($P < 0.001$)	0.342 ($P < 0.001$)	0.414 ($P < 0.001$)	0.268 ($P = 0.001$)
ln (α-2-macroglobulin) (mg/dL)			1.000	0.281 ($P < 0.001$)	0.141 ($P = 0.078$)	0.171 ($P = 0.032$)	-0.079 ($P = 0.32$)
ln (hyaluronic acid) (mg/L)				1.000	0.373 ($P < 0.001$)	0.493 ($P < 0.001$)	0.089 ($P = 0.27$)
ln (procollagen III peptide) (U/mL)					1.000	0.600 ($P < 0.001$)	0.145 ($P = 0.071$)
ln (type IV collagen) (mg/L)						1.000	0.358 ($P < 0.001$)
ln (TIMP-2) (mg/L)							1.000

TIMP, tissue inhibitor of matrix metalloproteinase.

the regression function for the validation set, the fibrosis score demonstrated good reproducibility, showing 1.33 in patients with chronic hepatitis of F1 ($n = 27$), 2.20 of F2 ($n = 20$), 3.11 of F3 ($n = 20$) and 5.30 of F4 ($n = 2$), respectively (Fig. 3). Although F4 fibrosis stage consisted of only two patients and the score 5.30 was regarded as of rather higher value, the scores of other stages of fibrosis were concordant with histological fibrosis.

Comparisons of efficacy with various fibrosis scores (Fig. 4)

In order to evaluate the efficacy and usefulness of the obtained FSB, we compared it with previously reported fibrosis scores using training data. AAR, APRI and FibroTest showed only slight correlation with actual histological stage. FIB-4 demonstrated an increasing trend of the score associated with histological fibrosis, but significant overlapping scores were found in F1–F4. Spearman's correlation coefficients of AAR, APRI, FIB-4 and FibroTest were 0.199 ($P = 0.012$), 0.265 ($P = 0.001$), 0.412 ($P < 0.001$) and 0.330 ($P < 0.001$), respectively. Our FSB showed a Spearman's correlation coefficient of 0.625 ($P < 0.001$), and was a much higher value than the others. The dichotomous discrimination function for cirrhosis and hepatitis C in Japanese patients[23] showed good differentiation also in patients with hepatitis B virus.

DISCUSSION

RECOGNITION OF SEVERITY of chronic hepatitis is essential in managing patients with chronic HBV infection: estimation of length of infection, existence of any previous hepatitis activity, presumption of current fibrotic stage, and prediction of future fibrosis progression and hepatocarcinogenesis. Differential diagnosis of cirrhosis from chronic hepatitis is especially important in the evaluation of chronic HBV infection. Identification of liver cirrhosis often leads to an important change in management of the patient: need for fiberscopic examination for esophageal varices, ultrasonographic exploration for the association of liver cancer, and prediction of hepatic decompensation. Guidelines published by the American Association of Study of Liver Disease[24] recommend liver biopsy for HBV carriers with aminotransferase elevation or for any candidates of antiviral therapy, because hepatic fibrosis sometimes shows unexpectedly far advancement to cirrhosis, and because it is very difficult to evaluate and translate the liver function tests or ultrasonographic findings compared to chronic hepatitis type C.

Table 3 Demography and laboratory data of 69 patients in training group

	F1 ($n=27$)	F2 ($n=20$)	F3 ($n=20$)	F4 ($n=2$)
Demographics				
Men : women	18:9	15:5	13:7	2:0
Age (median, range)	36 (13–64)	45 (14–64)	36.5 (24–59)	32 (25–39)
Laboratory data (median, range)				
WBC (×1000/mm^3)	5.0 (2.8–8.7)	5.8 (2.8–11.6)	5.3 (3.2–8.1)	3.85 (2.7–5.0)
Hemoglobin (g/dL)	14.8 (12.4–17.4)	15.0 (12.4–16.9)	14.4 (11.1–16.4)	14.4 (12.5–16.3)
Platelet (×1000/mm^3)	204 (86–322)	180 (90–275)	147 (90–276)	130 (67–183)
Albumin (g/dL)	4.4 (2.8–5.2)	4.2 (3.5–5.1)	4.3 (3.4–4.9)	4.45 (4.0–4.9)
Bilirubin (mg/dL)	0.9 (0.4–6.4)	0.8 (0.2–1.6)	0.75 (0.4–1.7)	1.15 (1.1–1.2)
AST (IU/L)	52 (17–575)	50.5 (21–272)	65 (22–284)	248.5 (51–446)
ALT (IU/L)	84 (16–1101)	101.5 (19–554)	86.5 (16–1113)	453.5 (74–833)
γ-GTP (IU/L)	42 (14–332)	54 (16–205)	52.5 (13–191)	193 (57–329)
γ-Globulin (g/dL)	1.30 (1.04–1.59)	1.35 (1.18–2.53)	1.62 (1.16–1.97)	1.545 (1.51–1.58)
γ-Globulin (%)	17.9 (14.3–22.1)	19.6 (15.5–30.8)	22.0 (16.5–24.6)	20.15 (19.3–21.0)
α-2-Macroglobulin (mg/dL)	287 (160–687)	270 (89–452)	272.5 (211–463)	389 (313–465)
Haptoglobin (mg/dL)	58 (<5–229)	74 (<5–154)	56.5 (<5–198)	<5 (<5–<5)
Apolipoprotein A-I (mg/dL)	146 (95–216)	137 (87–162)	120 (88–170)	100.5 (74–127)
Hyaluronic acid (μg/L)	27 (<5–113)	36 (10–1050)	59 (14–439)	331 (225–437)
TIMP-1 (ng/mL)	168.5 (83–302)	176 (127–408)	182 (104–303)	390.5 (283–498)
TIMP-2 (ng/mL)	76 (25–143)	86.5 (28–154)	77.5 (32–141)	100.5 (91–110)
Procollagen III peptide (U/mL)	0.71 (0.27–2.20)	0.88 (0.63–2.80)	0.995 (0.60–2.10)	1.75 (1.50–2.00)
Type IV collagen 7S (ng/ml)	3.6 (2.7–17.0)	5.25 (3.3–13.0)	5.7 (3.0–16.0)	15.5 (15.0–16.0)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; γ-GTP, γ-glutamyl transpeptidase; TIMP, tissue inhibitor of matrix metalloproteinase; WBC, white blood cells.

Recently, non-invasive estimation of severity of liver fibrosis has been reported in patients with HBV-related chronic hepatitis.[2,6–13] However, these studies were principally aimed at differentiation of advanced fibrotic stages of F3 or F4 from mild fibrotic stages of F1 or F2. Those discrimination functions were insufficient to recognize the stepwise progression of viral hepatitis from F1–F4. This dichotomy (mild or severe) of chronic hepatitis B seemed less valuable in the study of disease progression, disease control abilities of antiviral drugs and estimation of histological improvement after anti-inflammatory drugs. A histology-oriented, practical and reliable formula is therefore required for the diagnosis and investigation of chronic hepatitis B.

This study aimed to establish non-invasive evaluation and calculation of liver fibrosis for patients with chronic hepatitis B virus infection. Although it was retrospectively performed as a multicenter study of eight institutions, judgment of histological diagnosis was independently performed by four pathologists in another hospital, who were informed only of the patient's age, sex and positive HBV infection. Objective judgment of the histological staging and grading in sufficient biopsy specimens could be obtained.

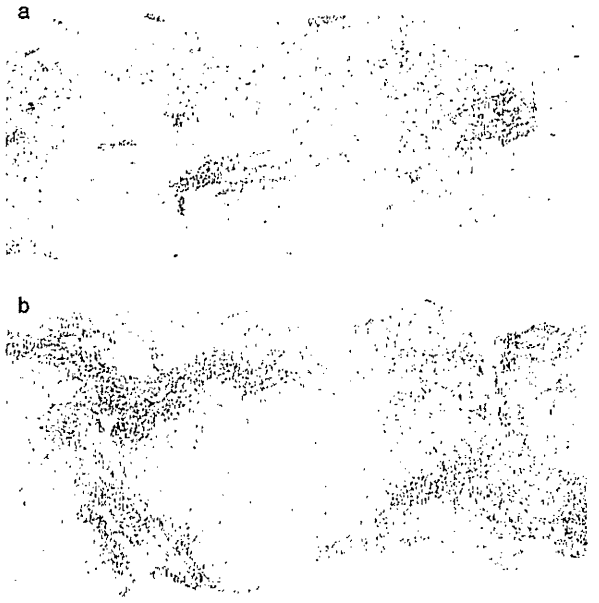

Figure 2 Case presentations of the training set. (a) A 28-year-old man with F1 fibrosis. Final regression function provided his fibrosis score as 0.99. (b) A 45-year-old man with F3 fibrosis. His regression coefficient was calculated as 3.10. Silver stain, ×40.

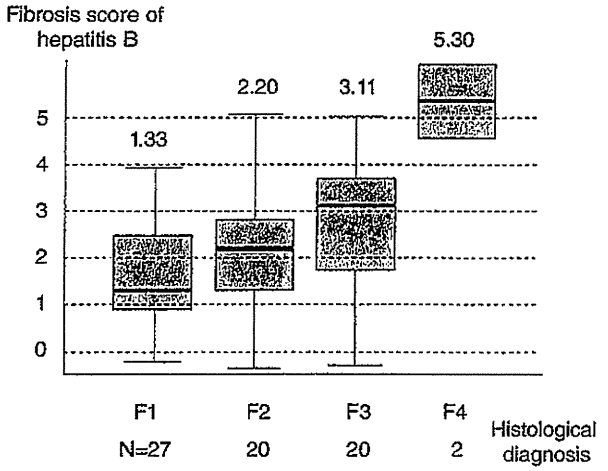

Figure 3 Box and whisker plots of fibrotic score of each group of histological fibrosis in the validation dataset. The fibrosis score of hepatitis B was generated by the function, $z = 1.40 \times ln$ (type IV collagen 7S) (ng/mL) – $0.017 \times$ (platelet count) ($\times 1000^3/mm^3$) + $1.24 \times ln$ (tissue inhibitor of matrix metalloproteinase-2) (ng/mL) + $1.19 \times ln$ (α-2-macroglobulin) (mg/dL) – 9.15.

As many as 227 patients with chronic hepatitis B were analyzed in this study, who had been diagnosed as having chronic hepatitis or cirrhosis by liver biopsy performed in experienced liver units in Japan. To obtain the most suitable equation approximating histological fibrotic stage, multivariate analysis was performed using two demographic parameters (age and sex) and 21 hematological and biochemical markers with or without logarithmic transformation. They included many kinds of fibrosis markers: α-2-macroglobulin, haptoglobin concentration, haptoglobin typing, apolipoprotein A1, hyaluronic acid, TIMP-1, TIMP-2, procollagen III peptide and type IV collagen 7S Multiple regression analysis finally generated a first-degree polynomial function consisting of four variables: type IV collagen 7S, platelet count, TIMP-2 and α2-macroglobulin. A constant numeral (–9.15) was finally adjusted in the regression equation in order to obtain fitted figures for a fibrotic stage of F1–F4. From the magnitude of the standardized partial regression coefficient of individual variable in the function, platelet count demonstrated the most potent contribution toward the prediction of liver fibrosis. Type IV collagen 7S and *ln* (TIMP-2) proved to be the second and third distinctive power in the model, respectively.

The FSB was sufficiently fitted to actual fibrotic stages with certain overlapping as is usually found in histological ambiguity judged by pathologists. Because judgment of fibrosis in chronic hepatitis often shows a transitional histological staging, pathological examination cannot always make a clear-cut diagnosis discriminating F1–F4. Considering the limitation of the pathological difficulty in differentiating the four continuous disease entities, the obtained regression function showed satisfactory high accuracy rates in the prediction of liver disease severity. The FSB can provide one or two decimal places (e.g. 3.2 or 3.24) and the utility of the score is possibly higher than the mere histological stage of F1–F4. The reproducibility was confirmed by the remaining 67 patients' data obtained from the other six hospitals. Although the validation data were collected from a different geographic area and different chronological situation, the FSB showed similar results in prediction of histological staging.

The FSB seemed a very useful quantitative marker in evaluating fibrotic severity of hepatitis B patients without invasive procedures and without any specialized ultrasonography or magnetic resonance imaging. The FSB also has an advantage of measurement, in which old blood samples are available for retrospective assessment of varied clinical settings: for example, old sera from 20 years prior to the time of initial liver biopsy, or paired sera before and after long-term antiviral therapy. These kinds of retrospective assessments of fibrotic staging will be valuable in estimating a long-term progression of liver disease, in evaluating efficacy of long-term medication or other medical intervention, or in making a political judgment from the viewpoints of socioeconomic efficacy.

The score can be calculated for any patients with chronic HBV infection. Although this multiple regression model dealt with appropriate logarithmic transformation for non-normal distribution parameters, the regression analysis was based on a linear regression model. Very slight fibrosis can be calculated as less than 1.00, which is commonly found to a slight degree in chronic hepatitis with tiny fibrotic change as F0. Very severe fibrosis might be calculated as more than 4.00, which is an imaginary and nonsense number in the scoring system of fibrosis. The FSB is, however, very useful and valuable in a real clinical setting: estimation of severity of liver fibrosis in an outpatient clinic, evaluation of the natural progression of a patient's fibrosis over 10 years and assessment of a long-term administration of interferon in patients with chronic hepatitis B from the viewpoint of fibrotic change. Recent development of new nucleoside/nucleotide analogs requires evaluation for long-term histological advantage, for aggravation of hepatitis stage during viral and biochemical breakthrough caused by HBV mutation, and even for