

**Conclusions** A decision tree model that includes age, gender, AFP, platelet counts, and GGT is useful for predicting the probability of response to therapy with peg-interferon plus ribavirin and has the potential to support clinical decisions regarding the selection of patients for therapy.

**Keywords** Data mining · Decision tree · Alpha-fetoprotein · HCV · Peg-interferon

## Introduction

The current standard therapy for genotype 1 chronic hepatitis C is 48 weeks of pegylated interferon (PEG-IFN) plus ribavirin (RBV) [1]. Sustained virological response (SVR), defined as undetectable HCVRNA post-treatment is regarded as a cure of chronic hepatitis C. However, the rate of SVR to this regimen is only 50% in patients with HCV genotype 1b and a high HCVRNA titer [2, 3]. Since PEG-IFN and RBV combination therapy is costly and accompanied by potential adverse effects, the ability to predict the possibility of SVR before therapy may significantly influence the selection of patients for therapy. A recent report revealed that single nucleotide polymorphisms located in the *IL28B* are strongly associated with a response to PEG-IFN plus RBV therapy [4–6]. Besides, the amino acid substitutions in the NS5A [7–9] or core region of HCV were also associated with response to therapy [10, 11]. Unfortunately, these host genetic and viral factors are not yet readily available for general application in actual clinical practice. Fibrosis of the liver is also an important predictor of response, but resources may be limited in some countries. Clinical and non-invasive parameters may be better suited for general practice, but there is no established means by which the likelihood of a response can be predicted prior to therapy.

Data mining is a method of predictive analysis that explores data, without setting the hypothesis, to discover hidden patterns and relationships in highly complex datasets and enables the development of predictive models. Decision tree analysis is a core component of data mining and predictive modeling [12], and it is utilized by decision makers in various fields of business. Recent publications on decision tree analysis indicate its usefulness for defining prognostic factors in various diseases such as prostate cancer [13], diabetes [14], melanoma [15, 16], colorectal carcinoma [17, 18], and liver failure [19]. The results of the analysis are presented as a tree structure, which is intuitive and facilitates the allocation of patients into subgroups by following the flow chart form [20]. We have recently reported the usefulness of decision tree analysis for the prediction of early virological response (undetectable

HCVRNA within 12 weeks of therapy) to PEG-IFN and RBV combination therapy in chronic hepatitis C [21].

In the present study, we used decision tree analysis to explore baseline predictors of response to PEG-IFN/RBV therapy so that a pre-treatment algorithm could be created to discriminate chronic hepatitis C patients who are likely to respond to PEG-IFN/RBV therapy from those who are not. For the purpose of use in general practice, only clinical and non-invasive parameters were included in the analysis.

## Materials and methods

### Patients

This was a multicenter retrospective cohort study supported by the Japanese Ministry of Health, Labor and Welfare. Data were collected from a total of 800 chronic hepatitis C patients who received therapy for 48 weeks with PEG-IFN alpha-2b and RBV at Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Osaka University, Nagoya City University Graduate School of Medical Sciences, Yamanashi University, and their related hospitals. The inclusion criteria to be enrolled in this study were as follows (1) infection by genotype 1b, (2) HCVRNA higher than 100,000 IU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic systems, CA), which is typically used for the definition of high viral load in Japan, (3) lack of co-infection with hepatitis B virus or human immunodeficiency virus, (4) lack of other causes of liver disease such as autoimmune hepatitis and primary biliary cirrhosis and (5) completion of at least 12 weeks of therapy. Patients received PEG-IFN alpha-2b (1.5 µg/kg) subcutaneously every week and were administered a weight-adjusted dose of RBV (600 mg for <60 kg, 800 mg for 60–80 kg, and 1,000 mg for >80 kg), which is the recommended dosage in Japan. Patients who were treated for more than 49 weeks were not included in the study. For the analysis, patients were randomly assigned to either the model building ( $n = 506$ ) or the internal validation ( $n = 295$ ) group. Consent was obtained from each patient. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional review committee. The baseline characteristics and representative laboratory test results are listed in Table 1. The overall rate of SVR was 47% in the model building set and 49% in the validation set. There were no significant differences in the clinical backgrounds between these 2 groups.

For external validation of the model, we collaborated with another study group supported by the Japanese Ministry of Health, Labor and Welfare. This multicenter study group consisted of 29 medical centers and hospitals

**Table 1** Comparison of pre-treatment factors between model building and internal validation patients

	Model (n = 506)	Validation (n = 295)
Age (years)	56 (14–75)	55 (18–74)
Male gender <sup>a</sup>	261/506 (52%)	160/295 (54%)
Body mass index (kg/m <sup>2</sup> )	22.9 (14.3–34.0)	23.2 (16.1–33.8)
Albumin (g/dl)	4 (2.7–5.0)	4 (2.8–4.9)
Creatinine (mg/dl)	0.7 (0.4–1.5)	0.7 (0.4–1.1)
AST (IU/l)	60 (11–370)	62 (11–240)
ALT (IU/l)	73 (11–413)	73 (14–390)
GGT (IU/l)	56 (10–328)	55 (7–409)
Total cholesterol (mg/dl)	173 (73–297)	171 (29–273)
Triglyceride (mg/dl)	105 (33–474)	109 (32–372)
White blood cell count (/μl)	4,745 (1,800–10,900)	4,823 (1,200–9,700)
Neutrophil count (/μl)	2,563 (667–7,870)	2,484 (508–7,579)
Red blood cell count (/μl)	448 (313–577)	451 (313–574)
Hemoglobin (g/dl)	14.1 (9.4–18.3)	14.1 (10.0–18.0)
Hematocrit (%)	41.7 (13.3–53.7)	41.9 (15.5–52.7)
Platelets (10 <sup>9</sup> /l)	164 (52–380)	158 (43–312)
AFP (ng/ml)	14.7 (0.9–680)	13 (0.8–323)
HCV RNA (10 <sup>3</sup> IU/ml)	1,852 (100–5,100)	1,870 (100–5,100)
Fibrosis stage: F3–4	73/417 (18%)	48/247 (19%)

Data expressed as median (range) unless otherwise indicated

AST aspartate aminotransferase, ALT alanine aminotransferase, GGT gamma-glutamyltransferase, AFP alpha-fetoprotein

<sup>a</sup> Data expressed as number/available data (percentage)

belonging to the National Hospital Organization. A dataset collected from 524 patients who were treated with PEG-IFN alpha-2b/RBV was used as an external validation dataset, i.e., completely independent from the dataset that was used for model building.

#### Laboratory tests

Blood samples were obtained before therapy and at least once every month during therapy, and were used for hematologic tests, blood chemistry analysis and determination of HCV RNA. Pretreatment levels of HCV RNA were quantified by Cobas Amplicor (Roche Diagnostic Systems, Pleasanton, CA). SVR was defined as undetectable HCV RNA at week 24 after completion of therapy, as determined by qualitative PCR with a lower end detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems). Liver biopsy was available in 664 patients. Fibrosis and activity

were scored according to the METAVIR scoring system [22]. Fibrosis was staged on a scale of 0–4: F0 (no fibrosis), F1 (mild fibrosis: portal fibrosis without septa), F2 (moderate fibrosis: few septa), F3 (severe fibrosis: numerous septa without cirrhosis) and F4 (cirrhosis). Activity of necroinflammation was graded on a scale of 0–3: A0 (no activity), A1 (mild activity), A2 (moderate activity) and A3 (severe activity).

#### Statistical analysis

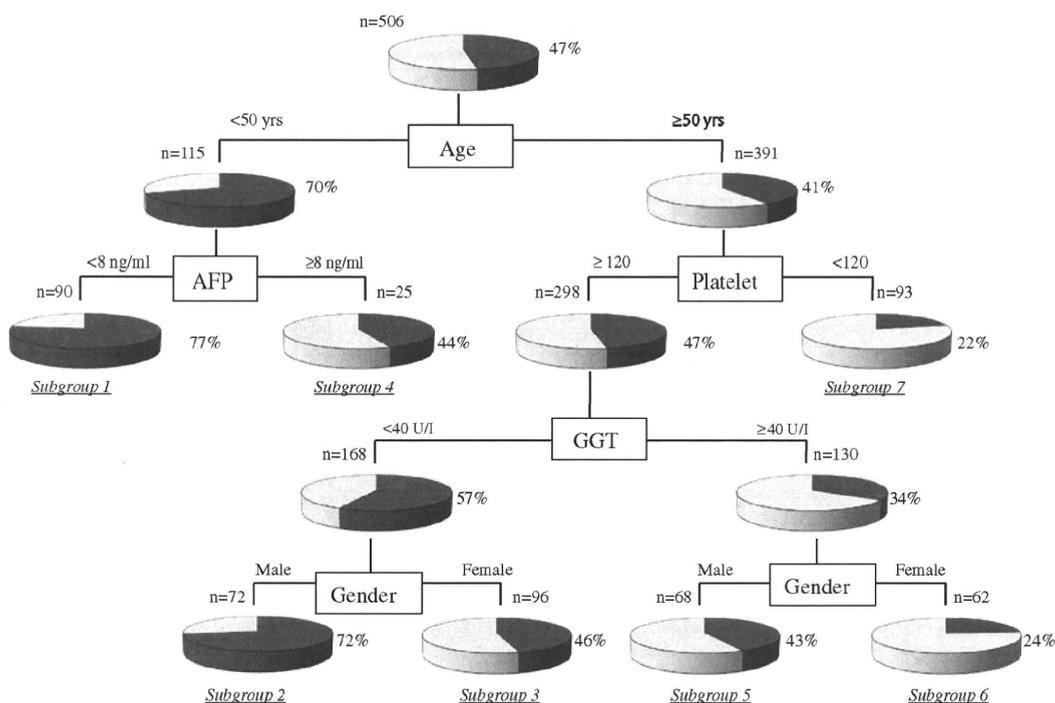
A database of pretreatment variables was created containing 6 variables from hematological tests (red blood cells, hemoglobin, hematocrit, white blood cells, neutrocytes and platelets), 8 variables from the blood chemistry test [creatinine, albumin, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyltransferase (GGT), total cholesterol, triglyceride and alpha-fetoprotein (AFP)], serum level of HCV RNA and 3 variables for patient characteristics (age, gender and body mass index). Based on this database, the recursive partitioning analysis algorithm referred to as decision tree analysis was implemented to define meaningful subgroups of patients with respect to the possibility of achieving SVR.

Decision tree analysis is a family of nonparametric regression methods. Software is used to automatically explore the data to search for optimal split variables and to build a decision tree structure [23]. For the analysis, the entire study population was evaluated to determine which variables and cutoff points yielded the most significant division into 2 prognostic subgroups that were as homogeneous as possible for the probability of SVR. Thereafter, the same analytic process was applied to all newly defined subgroups. A restriction was imposed on the tree construction such that the procedure stopped when either no additional significant variable was detected or when the sample size was below 20. For this analysis, the data mining software IBM SPSS Modeler 13 (IBM SPSS Inc., Chicago, IL) was utilized. SPSS software v.15.0 (SPSS Inc., Chicago, IL) was used for multivariate logistic regression analysis.

## Results

#### Decision tree analysis

Decision tree analysis was carried out on the model building dataset from 506 patients using 18 variables. Figure 1 shows the results. The analysis automatically selected 5 predictive variables to produce a total of 7 subgroups of patients. Age was selected as the variable of initial split with an optimal cutoff of 50 years. The possibility of achieving SVR was 41% for patients older than 50 compared to 70% for patients



**Fig. 1** Decision tree analysis. Boxes indicate the factors for splitting and the cutoff value for the split. Pie charts indicate the rate of SVR for each group. Terminal subgroups of patients discriminated by the

analysis are numbered from 1 to 7. AFP alpha-fetoprotein, GGT gamma-glutamyltransferase

younger than 50. Among patients younger than 50, the level of serum AFP, with an optimal cutoff of 8 ng/ml, was selected as the variable of second split. Patients with lower AFP levels had a higher probability of SVR (77 vs. 44%). Among older patients, platelet count was selected as the second variable of split, with an optimal cutoff of  $120 \times 10^9/l$ . Patients with higher platelet counts had a higher probability of SVR (47 vs. 22%). Among patients with platelet counts higher than  $120 \times 10^9/l$ , GGT was selected as the third variable of split with an optimal cutoff of 40 IU/l. Patients with a lower GGT level had a higher probability of SVR (57 vs. 34%). Gender was selected as the fourth variable of split, with male gender being a predictor of a higher SVR probability (72 vs. 46% in patients with GGT levels <40 IU/l and 43 vs. 24% in those with GGT  $\geq 40$  IU/l). HCVRNA load was included in the analysis but was not selected as a significant variable.

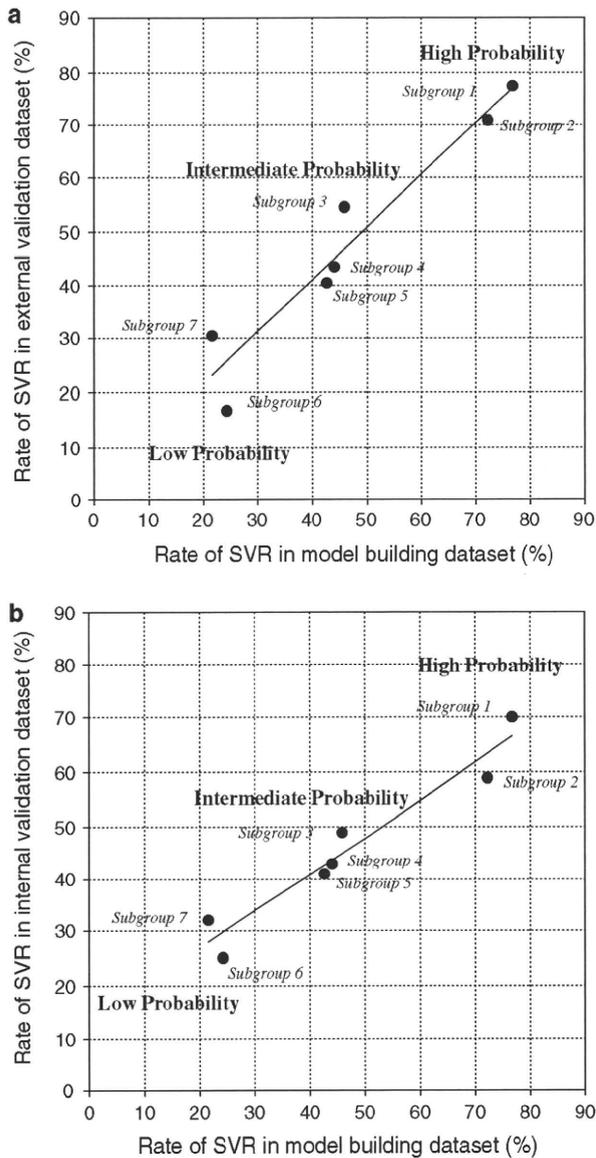
The probabilities of SVR for the 7 subgroups derived by this process were highly variable. The subgroup of young patients (<50 years) with low serum AFP (<8 ng/ml) (subgroup 1) or the subgroup of older ( $\geq 50$  years) male patients with high platelet counts ( $\geq 120 \times 10^9/l$ ) and low serum GGT (<40 IU/l) (subgroup 2) showed the highest

probability of SVR (72 and 77%), while the subgroup of older ( $\geq 50$  years) patients with low platelet counts (< $120 \times 10^9/l$ ) (subgroup 7) and older ( $\geq 50$  years) female patients with high serum GGT (subgroup 6) showed the lowest probability of SVR (22 and 24%).

#### Validation of the decision tree

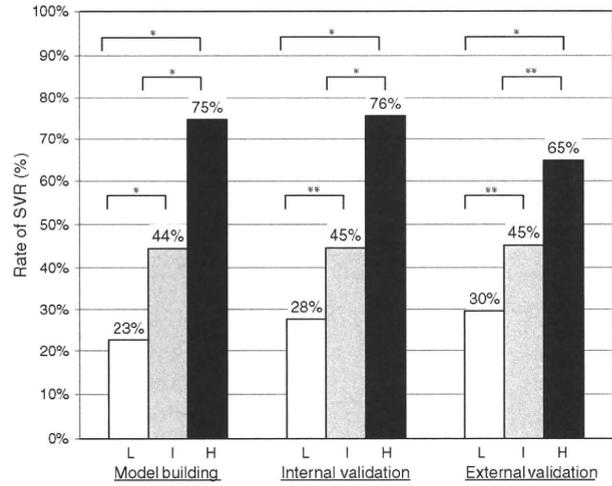
The results of the decision tree analysis were validated with an internal validation dataset of 295 cases, which was independent of the model building dataset. Each patient in the validation set was allocated to subgroups 1–7 using the flow-chart form of the decision tree. The rates of SVR were 77% for subgroup 1, 71% for subgroup 2, 55% for subgroup 3, 44% for subgroup 4, 41% for subgroup 5, 17% for subgroup 6, and 30% for subgroup 7. The rates of SVR for each subgroup of patients were closely correlated between the model building dataset and the internal validation dataset ( $r^2 = 0.925$ ) (Fig. 2a).

To further confirm the universality of the results, data collected from 524 patients by a collaborating study group were used for external validation. Thus, the dataset used for external validation was completely independent of the



**Fig. 2** Validation of the decision tree analysis by an internal and external validation dataset: subgroup-stratified comparison of the SVR rate. The rate of SVR in each subgroup was plotted. The X axis represents the model building, and the Y axis represents the validation datasets. **a** Internal validation and **b** external validation. There was a close correlation between the model building and the internal validation dataset (correlation coefficient  $r^2 = 0.925$ ) and between the model building and the external validation dataset (correlation coefficient  $r^2 = 0.936$ )

original dataset used for model building. Each patient in the external validation set was allocated to subgroups 1–7 using the flow-chart form of the tree. The rates of SVR were 70% for subgroup 1, 59% for subgroup 2, 49% for subgroup 3, 43% for subgroup 4, 41% for subgroup 5, 25% for subgroup 6, and 32% for subgroup 7. The rates of SVR for each subgroup of patients were closely correlated



**Fig. 3** Comparison of SVR rates between groups divided by the decision tree. The rate of SVR was compared among the 3 groups of patients divided by the decision tree analysis (white, gray and black boxes, indicating a low (L), intermediate (I) and high (H) probability group, respectively). The rate of SVR was significantly different among the 3 groups. \* $p < 0.0001$ , \*\* $p < 0.001$

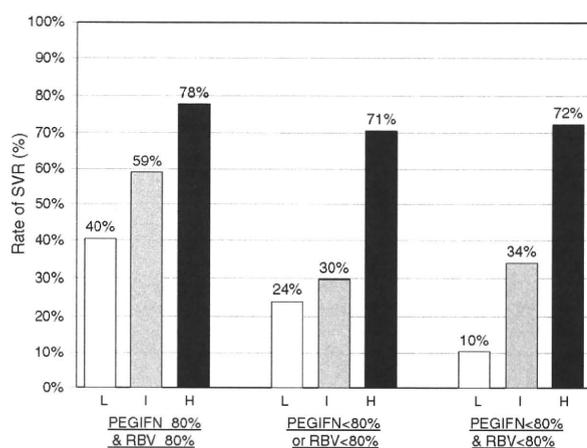
between the model-building dataset and the validation dataset ( $r^2 = 0.936$ ) (Fig. 2b).

Construction of 3 groups according to the probability of SVR

Seven subgroups were reconstructed into 3 groups according to their predicted rates of SVR: the high probability group consisted of subgroups 1 and 2, the intermediate probability group consisted of subgroups 3, 4 and 5, and the low probability group consisted of subgroups 6 and 7. The rate of SVR was significantly different among the 3 groups (Fig. 3). The rate of SVR in the high probability group was consistently high: 75% for model building patients, 76% for internal validation patients and 65% for external validation patients. Conversely, the rate of SVR in the low probability group was consistently low: 23% for model building patients, 28% for internal validation patients and 30% for external validation patients. The rate of SVR in the intermediate probability group was 44% for model building patients, 45% for internal validation patients and 45% for external validation patients. Since 28–32% of patients were classified as high probability and 30–32% were classified as low probability, roughly 60% of patients were classified as having either a high or low probability of achieving SVR.

Effect of dose reductions of PEG-IFN and RBV on SVR

The cumulative dose of PEG-IFN and RBV was not included as a variable of analysis since the present study



**Fig. 4** Comparison of SVR rates among groups stratified by drug adherence. The 3 groups of patients divided by the decision tree analysis (white, gray and black boxes indicating a low (L), intermediate (I) and high (H) probability group, respectively) were further stratified according to the cumulative drug exposure of PEG-IFN and RBV. The good adherence group ( $\geq 80\%$  planned dose of both PEG-IFN and RBV) had a higher rate of SVR compared with the poor adherence group ( $< 80\%$  planned dose of both PEG-IFN and RBV) in the low ( $p = 0.0003$ ) and intermediate ( $p = 0.007$ ) but not in the high probability group ( $p = 0.53$ )

aimed to develop a pre-treatment model for the prediction of response. To analyze the possible effect of drug reductions on the result of the decision tree analysis, 3 groups of patients divided by the decision tree analysis (low, intermediate and high probability group) were further stratified according to the cumulative drug exposure of PEG-IFN and RBV (Fig. 4). Even after adjustment for adherence, 3 groups of patients still had low, intermediate and high probability of achieving SVR, respectively. Of note, the good adherence group ( $\geq 80\%$  planned dose of both PEG-IFN and RBV) had higher rates of SVR compared with the poor adherence group ( $< 80\%$  planned dose of both PEG-IFN and RBV) in the low ( $p = 0.0003$ ) and intermediate ( $p = 0.007$ ) probability group, but not in the high probability group ( $p = 0.53$ ).

Factors associated with SVR by multivariate logistic regression analysis

We also explored the factors associated with SVR using a standard statistical analysis. By univariate analysis, age, gender, serum albumin, creatinine, alanine aminotransferase, GGT, red blood cell count, hemoglobin, hematocrit, platelet count and AFP were found to be associated with SVR (Table 2). HCV RNA load was not associated with SVR. By multivariate analysis, age, gender, GGT and platelet count were found to be independently associated with SVR (Table 3). Of note, AFP, which was selected as a

significant predictor of response in the decision tree analysis, was not found to be an independent response predictor in the standard multivariate analysis. This indicates a unique feature of the decision tree analysis; i.e., it could identify significant predictors that specifically apply to selected patients, in this case patients younger than 50 years old.

Relationships between decision tree model and stage of fibrosis or HCV RNA load

Liver biopsy was performed in 664 patients. The distribution of fibrosis in three probability groups differed significantly. Advanced fibrosis (F3 or F4) was higher in the low probability group (39%) compared to the intermediate probability group (13%) ( $p < 0.0001$ ) and to the high probability group (6%) ( $p < 0.0001$ ). Advanced fibrosis was also higher in the intermediate group compared to the high probability group ( $p = 0.01$ ). AFP was significantly associated with liver fibrosis stage: medians of AFP levels were 4.9, 5.9, 13.0 and 18.6 for F1, F2, F3 and F4, respectively ( $p < 0.0001$ , Spearman's rank correlations). Lower platelet counts correlated with advanced fibrosis stages (data not shown). The SVR rate was higher in the high probability group compared to the intermediate or low probability group after stratification by HCV RNA load. Among patients with low HCV RNA load ( $< 400,000$  IU/ml), the rate of SVR was 93, 59 and 50% for the high, intermediate and low probability group, respectively ( $p = 0.002$  for high vs. intermediate and  $p < 0.001$  for high vs. low probability groups). Among patients with a high HCV RNA load ( $\geq 400,000$  IU/ml), the rate of SVR was 73, 42 and 21% for the high, intermediate and low probability group, respectively ( $p < 0.001$  for high vs. low, high vs. intermediate and intermediate vs. low probability groups).

## Discussion

Currently, the combination of PEG-IFN and RBV is the recommended therapy for chronic HCV infection. The rate of SVR with 48 weeks of therapy is around 50% in patients with HCV genotype 1b and a high HCV RNA titer [2, 3]. To date, the virological response during therapy is the most reliable means for predicting the likelihood of SVR [2, 24, 25]. More potent therapy, such as a triple combination of protease inhibitor, PEG-IFN and RBV, is being evaluated in clinical trials but is not readily available [26, 27]. Under the circumstances, pre-treatment prediction of the likelihood of SVR may be useful for both patients and physicians to support clinical decisions as to whether to start PEG-IFN/RBV therapy or delay treatment until a new more effective therapy becomes available.

**Table 2** Comparison of pre-treatment factors between patients with and without sustained virological response (SVR) among the model building dataset (*n* = 506)

	SVR ( <i>n</i> = 240)	Non-SVR ( <i>n</i> = 266)	<i>p</i>
Age (years)	54 (25–75)	60 (36–73)	<0.0001
Male gender <sup>a</sup>	151/240 (63%)	171/266 (41%)	<0.0001
Body mass index (kg/m <sup>2</sup> )	22.5 (16.8–32.0)	22.6 (15.5–33.3)	0.244
Albumin (g/dl)	4.1 (3.2–5.0)	4 (2.7–4.9)	0.004
Creatinine (mg/dl)	0.7 (0.44–1.14)	0.69 (0.39–1.47)	<0.0001
AST (IU/l)	59 (11–370)	61 (17–261)	0.457
ALT (IU/l)	58 (11–413)	53 (11–316)	0.031
GGT (IU/l)	31 (10–322)	43 (12–328)	0.005
Total cholesterol (mg/dl)	175 (87–297)	171 (73–274)	0.184
Triglyceride (mg/dl)	105 (36–474)	105 (33–294)	0.992
White blood cell count (/μl)	4,600 (2,200–10,900)	4,425 (1,800–10,810)	0.479
Neutrophils (/μl)	2,507 (667–7,870)	2,423 (900–7,281)	0.321
Red blood cell count (/μl)	455 (336–577)	441 (313–564)	0.001
Hemoglobin (g/dl)	14.3 (10.2–17.6)	13.9 (9.4–17.9)	0.004
Hematocrit (%)	42.1 (13.3–53.7)	41.2 (30.7–52.0)	0.031
Platelets (10 <sup>9</sup> /l)	178 (81–380)	142 (60–320)	<0.0001
AFP (ng/ml)	4.3 (0.9–680)	6.4 (1.9–468)	0.041
HCVRNA (10 <sup>3</sup> IU/ml)	1,400 (100–5,100)	1,700 (100–5,100)	0.659
Fibrosis stage: F3–4 <sup>a</sup>	21/198 (11%)	52/219 (24%)	<0.0001

Data expressed as median (range) unless otherwise indicated

AST aspartate aminotransferase, ALT alanine aminotransferase, GGT gamma-glutamyltransferase, AFP alpha-fetoprotein

<sup>a</sup> Data expressed as number/available data (percentage)

**Table 3** Multivariate logistic regression analysis for factors associated with sustained virological response (SVR)

	Odds	95% CI	<i>p</i> value
Age (years)	0.96	0.94–0.98	0.001
Platelets (10 <sup>9</sup> /l)	1.09	1.04–1.14	<0.0001
ALT (IU/l)	1.01	1.00–1.01	0.001
GGT (IU/l)	0.99	0.98–0.99	<0.0001
Male gender	2.92	1.87–4.55	<0.0001

GGT gamma-glutamyltransferase

Using the data mining analysis, we constructed a simple decision tree model for the pre-treatment prediction of response to PEG-IFN/RBV. The analysis highlighted 5 variables relevant to response: age, gender, platelet count, AFP and GGT. Classification based on these variables identified subgroups of patients with high probabilities of achieving SVR among difficult to treat genotype 1b chronic hepatitis C patients. The reproducibility of the model was confirmed by the independent internal and external validation datasets. An advantage of the decision tree analysis over traditional regression models is that the decision tree model is user-intuitive and can be readily interpreted by medical professionals without any specific knowledge of statistics. Patients can be allocated to specific subgroups with a defined rate of response simply by following the flow-chart form. Using this model, an estimate of the response before treatment can be rapidly obtained, which may facilitate clinical decision making. Thus, this model could be readily applicable to clinical practice.

According to the results of the decision tree analysis, patients were categorized into 3 groups: the rate of SVR was 23–30% for the low probability group, 44–45% for the intermediate probability group and 65–76% for the high probability group. About 30% of patients were each categorized in the high and low probability group and the remaining 40% of patients in the intermediate probability group. These results support the evidence-based approach for selecting an optimum treatment strategy for individual patients. For example, patients in the high probability group may be the most suitable candidates for PEG-IFN/RBV therapy, while patients in the low probability group may be advised to wait for a future therapy, such as the combination of protease inhibitor, PEG-IFN and RBV. However, the estimation of low probability should not be used to preclude patients from therapy, and the final decision should be made on a case-by-case basis, taking into consideration the acceptance by the patient of a low likelihood of response and the potential risk of disease progression while waiting for a future therapy.

Another important finding was that poor adherence to drugs lowered the rate of SVR in the low and intermediate probability groups, which implies that effort should be made to maintain ≥80% of the planned dose of PEG-IFN and RBV in those patients. On the other hand, the rate of SVR was high irrespective of drug adherence in the high probability group. Whether shorter duration of therapy is sufficient in this group of patients should be confirmed in future study.

The variables used in the decision tree have been previously reported to associate with the efficacy of IFN therapy. Younger age and male gender are associated with a favorable response [28]. Lower platelet count is a hallmark of advanced fibrosis in chronic hepatitis C and is reported to be associated with poor response to IFN [29]. AFP is usually used for the screening or the diagnosis of hepatocellular carcinoma, but recent studies suggest an association between higher AFP levels and poor response to IFN therapy [30–33]. Previous report speculated that higher expression of AFP by hepatic progenitor cells may be associated with non-response to therapy [30]. Another report speculated that AFP levels predict poor response to therapy through the underlining link to advanced liver fibrosis [31]. Our data support the latter speculation since advanced fibrosis was associated with elevation of AFP levels. Fibrosis of the liver is an important predictor of response, but we did not include this factor in the decision tree analysis since liver biopsy may not always be available in general practice. As a result, two predictive factors that correlate with fibrosis stage (platelet counts and AFP) were selected in the model, and three probability groups reflected the different distribution of fibrosis stage. GGT is reported to be associated with insulin resistance and hepatic steatosis [34–37], a factor that confers resistance to IFN therapy [38–44]. What is unique to the present study is the visualization of response probability by combining these factors and its high reproducibility revealed by a high-quality validation of the model by internal and external validation datasets that were completely independent of the model building dataset. Since factors used in the model were clinical parameters that are readily available by the usual workup of patients, this model could be immediately applicable to clinical practice without imposing costs for additional examinations.

A potential limitation of this study is that data mining analysis has an intrinsic risk of showing relationships that fit to the original dataset but are not reproducible in different populations. Although internal and external validations showed that our model had high reproducibility, we recognize that further validation on a larger external validation cohort, especially in populations other than Japanese, may be necessary to further verify the reliability of our model.

In conclusion, we built a pre-treatment model for the prediction of virological response to PEG-IFN/RBV. Because this decision tree model was made up of simple variables, it can be easily applied to clinical practice. This model may have the potential to support decisions about patient selection for PEG-IFN/RBV based on a possibility of response weighed against the potential risk of adverse events or costs.

**Acknowledgments** This study was supported by a grant-in-aid from the Ministry of Health, Labor and Welfare, Japan.

## References

1. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology*. 2004;39:1147–71.
2. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL Jr, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med*. 2002;347:975–82.
3. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*. 2001;358:958–65.
4. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet*. 2009;41:1105–9.
5. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet*. 2009;41:1100–4.
6. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*. 2009;461:399–401.
7. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest*. 1995;96:224–30.
8. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med*. 1996;334:77–81.
9. Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, et al. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology*. 1997;25:750–3.
10. Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology*. 2005;48:372–80.
11. Okanoue T, Itoh Y, Hashimoto H, Yasui K, Minami M, Takehara T, et al. Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study. *J Gastroenterol*. 2009;44:952–63.
12. Breiman L, Friedman RA, Olshen CJ, Stone CM. *Classification and regression trees*. Calif: Wadsworth; 1980.
13. Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, et al. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol*. 2005;23:4322–9.
14. Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol*. 2002;12:243–8.
15. Averbuch BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center.

- implications on the future of cancer staging. *Surgery*. 2002;132:589–602.
16. Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. *J Clin Oncol*. 2004;22:3660–7.
  17. Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol*. 2007;14:34–40.
  18. Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res*. 2005;11:5440–3.
  19. Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, et al. Phosphorus ans an early predictive factor in patients with acute liver failure. *Transplantation*. 2003;75:2007–14.
  20. LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res*. 1995;75:113–24.
  21. Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatol Res*. 2010;40:251–60.
  22. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. *Hepatology*. 1996;24:289–93.
  23. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med*. 1989;8:539–50.
  24. Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology*. 2003;38:645–52.
  25. Lee SS, Ferenci P. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther*. 2008;13(Suppl 1):9–16.
  26. Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med*. 2009;360:1839–50.
  27. McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med*. 2009;360:1827–38.
  28. Sezaki H, Suzuki F, Kawamura Y, Yatsuji H, Hosaka T, Akuta N, et al. Poor response to pegylated interferon and ribavirin in older women infected with hepatitis C virus of genotype 1b in high viral loads. *Dig Dis Sci*. 2009;54:1317–24.
  29. Shiratori Y, Omata M. Predictors of the efficacy of interferon therapy for patients with chronic hepatitis C before and during therapy: how does this modify the treatment course? *J Gastroenterol Hepatol*. 2000;15(Suppl):E141–51.
  30. Gad RR, Males S, El Makhzangy H, Shouman S, Hasan A, Attala M, et al. Predictors of a sustained virological response in patients with genotype 4 chronic hepatitis C. *Liver Int*. 2008;28:1112–9.
  31. Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. *J Med Virol*. 2007;79:1686–95.
  32. Males S, Gad RR, Esmat G, Abobakr H, Anwar M, Rekaewicz C, et al. Serum alpha-fetoprotein level predicts treatment outcome in chronic hepatitis C. *Antivir Ther*. 2007;12:797–803.
  33. Bayati N, Silverman AL, Gordon SC. Serum alpha-fetoprotein levels and liver histology in patients with chronic hepatitis C. *Am J Gastroenterol*. 1998;93:2452–6.
  34. Fraser A, Ebrahim S, Smith GD, Lawlor DA. A comparison of associations of alanine aminotransferase and gamma-glutamyltransferase with fasting glucose, fasting insulin, and glycated hemoglobin in women with and without diabetes. *Hepatology*. 2007;46:158–65.
  35. Marchesini G, Avagnina S, Barantani EG, Ciccarone AM, Corica F, Dall'Aglio E, et al. Aminotransferase and gamma-glutamyltranspeptidase levels in obesity are associated with insulin resistance and the metabolic syndrome. *J Endocrinol Investig*. 2005;28:333–9.
  36. Soresi M, Tripi S, Franco V, Giannitrapani L, Alessandri A, Rappa F, et al. Impact of liver steatosis on the antiviral response in the hepatitis C virus-associated chronic hepatitis. *Liver Int*. 2006;26:1119–25.
  37. Yaginuma R, Ikejima K, Okumura K, Kon K, Suzuki S, Takei Y, et al. Hepatic steatosis is a predictor of poor response to interferon alpha-2b and ribavirin combination therapy in Japanese patients with chronic hepatitis C. *Hepatol Res*. 2006;35:19–25.
  38. Adinolfi LE, Gambardella M, Andrea A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology*. 2001;33:1358–64.
  39. Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol*. 2007;46:403–10.
  40. Berg T, Sarrazin C, Herrmann E, Hinrichsen H, Gerlach T, Zachoval R, et al. Prediction of treatment outcome in patients with chronic hepatitis C: significance of baseline parameters and viral dynamics during therapy. *Hepatology*. 2003;37:600–9.
  41. Mazzella G, Salzetta A, Casanova S, Morelli MC, Villanova N, Miniero R, et al. Treatment of chronic sporadic-type non-A, non-B hepatitis with lymphoblastoid interferon: gamma GT levels predictive for response. *Dig Dis Sci*. 1994;39:866–70.
  42. Thomopoulos KC, Theocharis GJ, Tsamantas AC, Siagris D, Dimitropoulou D, Gogos CA, et al. Liver steatosis is an independent risk factor for treatment failure in patients with chronic hepatitis C. *Eur J Gastroenterol Hepatol*. 2005;17:149–53.
  43. Villela-Nogueira CA, Perez RM, de Segadas Soares JA, Coelho HS. Gamma-glutamyl transferase (GGT) as an independent predictive factor of sustained virologic response in patients with hepatitis C treated with interferon-alpha and ribavirin. *J Clin Gastroenterol*. 2005;39:728–30.
  44. Camps J, Crisostomo S, Garcia-Granero M, Riezu-Boj JI, Civeira MP, Prieto J. Prediction of the response of chronic hepatitis C to interferon alfa: a statistical analysis of pretreatment variables. *Gut*. 1993;34:1714–7.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

Original Article

Hepatic steatosis in chronic hepatitis C is a significant risk factor for developing hepatocellular carcinoma independent of age, sex, obesity, fibrosis stage and response to interferon therapy

Masayuki Kurosaki,<sup>1</sup> Takanori Hosokawa,<sup>1</sup> Kotaro Matsunaga,<sup>2</sup> Itsuko Hirayama,<sup>1</sup> Tomohiro Tanaka,<sup>1</sup> Mitsuaki Sato,<sup>1</sup> Yutaka Yasui,<sup>1</sup> Nobuharu Tamaki,<sup>1</sup> Ken Ueda,<sup>1</sup> Kaoru Tsuchiya,<sup>1</sup> Teiji Kuzuya,<sup>1</sup> Hiroyuki Nakanishi,<sup>1</sup> June Itakura,<sup>1</sup> Yuka Takahashi,<sup>1</sup> Yasuhiro Asahina,<sup>1</sup> Nobuyuki Enomoto<sup>3</sup> and Namiki Izumi<sup>1</sup>

Divisions of <sup>1</sup>Gastroenterology and Hepatology, and <sup>2</sup>Pathology, Musashino Red Cross Hospital, Tokyo, and <sup>3</sup>First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan

**Aim:** Hepatic steatosis is linked to development of hepatocellular carcinoma (HCC) in non-viral liver disease such as non-alcoholic steatohepatitis. The present study aimed to assess whether hepatic steatosis is associated with the development of HCC in chronic hepatitis C.

**Methods:** We studied a retrospective cohort of 1279 patients with chronic hepatitis C who received interferon (IFN) therapy between 1994 and 2005 at a single regional hospital in Japan. Of these patients, 393 had a sustained virological response (SVR) and 886 had non-SVR to IFN therapy. After IFN therapy, these patients were screened for development of HCC every 6 months. The average period of observation was 4.5 years.

**Results:** HCC developed in 68 patients. The annual incidence of HCC was 2.73% for patients with a steatosis grade of 10% or greater and 0.69% for patients with a steatosis grade of 0–9%.

On multivariate analysis, higher grade of steatosis was a significant risk factor for HCC independent of older age, male sex, higher body mass index (BMI), advanced fibrosis stage and non-SVR to IFN therapy. The adjusted risk ratio of hepatic steatosis was 3.04 (confidence interval 1.82–5.06,  $P < 0.0001$ ), which was higher than that of older age (1.09), male sex (2.12), non-SVR to IFN (2.43) and higher BMI (1.69).

**Conclusion:** Hepatic steatosis is a significant risk factor for development of HCC in chronic hepatitis C independent of other known risk factors, which suggest the possibility that amelioration of hepatic steatosis may prevent hepatocarcinogenesis.

**Key words:** body mass index, fibrosis, sex, hepatocellular carcinoma, interferon, steatosis, virological response.

35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

INTRODUCTION

HEPATOCELLULAR CARCINOMA (HCC) is one of the most common cancers worldwide and its incidence has been increasing. This recent increase in HCC incidence may likely be attributed to the higher

prevalence of non-alcoholic fatty liver disease (NAFLD) and hepatitis C virus (HCV) infection.<sup>1</sup>

Non-alcoholic fatty liver disease is characterized by hepatic steatosis with or without inflammation in the absence of excessive alcohol consumption. Several studies have indicated the etiological association between NAFLD and development of HCC.<sup>2–4</sup> Other studies have shown that obesity or diabetes, a common etiology of non-alcoholic hepatic steatosis, is associated with development of HCC.<sup>5–7</sup> Although the mechanism of carcinogenesis in NAFLD has not been determined, an animal model showed that obesity-related hepatic steatosis leads to the development of hepatic

Correspondence: •• Namiki Izumi, Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. Email: nizumi@musashino.jrc.or.jp  
Received 23 January 2010; revision 10 May 2010; accepted 21 May 2010.

hyperplasia, suggesting the possibility that hepatic steatosis is a pre-malignant condition.<sup>8</sup>

Another important etiological agent for HCC is HCV infection. Because steatosis is a common pathological feature of HCV-infected patients,<sup>9</sup> the important question is whether steatosis influences the progression of liver disease in hepatitis C, by analogy with NAFLD. Several studies, including ours<sup>10</sup> indicated that hepatic steatosis promotes the progression of hepatic fibrosis.<sup>11-15</sup> The association between hepatic steatosis and the development of HCC in chronic hepatitis C has been proposed<sup>16</sup> and was confirmed in two studies<sup>17,18</sup> while another study failed to show such an association.<sup>19</sup> The present study was conducted to analyze the association between hepatic steatosis and development of HCC in a large cohort of chronic hepatitis C patients, which enabled to adjust for known risk factors for HCC.

## METHODS

### Patients

A TOTAL OF 1437 chronic hepatitis C patients were treated with interferon (IFN) at Musashino Red Cross Hospital between October 1994 and October 2005. Among them, 1279 patients who fulfilled the following inclusion criteria were enrolled in this study: (i) positive for HCV RNA by reverse-transcription polymerase chain reaction before IFN therapy; (ii) absence of other causes of liver disease, such as co-infection with hepatitis B virus, autoimmune hepatitis or primary biliary cirrhosis; (iii) had undergone liver biopsy within the 12 months prior to IFN treatment; (iv) were followed for more than 1 year after the completion of IFN therapy; and (v) absence of HCC during and within 1 year after the completion of therapy. A total of 158 patients were excluded: two patients who were positive for hepatitis B surface antigen, 97 patients lacking liver biopsy, 53 patients with less than 1 year's duration of follow up, and six patients who developed HCC within 1 year of the completion of IFN therapy. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional ethics review committee.

Patients were followed up by regular visits to our hospital every 1-3 months. Six patients died of liver-unrelated disease (two patients with gastric cancer and one patient each with lung cancer, colon cancer, pancreatic cancer and leukemia). There were 122 patients who were lost to follow up because of relocation. We included their data in the analysis, censored at the time

of their last visit. The start of follow up was defined as the date of completion of first IFN therapy and the end of follow up was defined as the date of diagnosis of HCC or the date of the last visit. The average period of follow up was 4.5 years.

Clinical characteristics and laboratory data were collected at the most recent time point before liver biopsy. Diabetes mellitus was diagnosed based on a fasting plasma glucose concentration that exceeded 126 mg/dL, a casual plasma glucose concentration that exceeded 200 mg/dL, or the need for insulin or oral anti-hyperglycemic drugs. Information regarding alcohol consumption was obtained through an interview. Body mass index (BMI) was calculated using the following formula: weight in kilograms/height in meters squared. The baseline clinical features of patients at enrollment are summarized in Table 1.

### Histological examination

Liver biopsy specimens were obtained from all patients before therapy. The median length of liver biopsy specimens was 13 mm (range 10-42 mm) and median number of portal tracts was 11 (range 4-30). Histological findings were re-evaluated recently by three independent pathologists who were blinded to the clinical details to ensure consistency over time. Fibrosis and activity were scored according to the METAVIR scoring system.<sup>20</sup> Fibrosis was staged on a scale of 0-4: F0 (no fibrosis); F1 (mild fibrosis: portal fibrosis without septa); F2 (moderate fibrosis: few septa); F3 (severe fibrosis: numerous septa without cirrhosis); and F4 (cirrhosis). Activity of necroinflammation was graded on a scale of 0-3: A0 (no activity); A1 (mild activity); A2 (moderate activity); and A3 (severe activity). Percentage of steatosis was quantified by determining the average proportion of hepatocytes affected by steatosis and graded on a scale of 0%, 1-9%, 10-29% and 30% or greater as reported previously.<sup>10</sup> All three pathologists assigned the same scale in 85% of cases for fibrosis staging, 87% for inflammation grading and 95% for steatosis grading. If there was discordance, the scores assigned by two pathologists were used for the analysis.

### Screening for HCC

At enrollment, no patient had HCC or any suspicious lesion on abdominal ultrasonography or computed tomography. Patients were examined for HCC by abdominal ultrasonography or computed tomography at least every 6 months. Suspicious lesions were examined further by a triphasic contrast-enhanced computerized tomography or magnetic resonance imaging

**Table 1** Clinical characteristics of patients

Male, n (%)	643 (50%)
Age (years)	54.2 ± 11.9
BMI (kg/m <sup>2</sup> )	23.4 ± 3.1
Alcohol consumption ≥20 g/day, n (%)	44 (3%)
Diabetes Mellitus, n (%)	197 (15%)
AST level (IU/L)	68.9 ± 45.3
ALT level (IU/L)	92.9 ± 75.9
GGT level (IU/L)	41.2 ± 38.2
Platelet count (×10 <sup>10</sup> /L)	16.4 ± 5.2
HCV genotype, n (%)	
1b	873 (68.2%)
2a	236 (18.4%)
2b	139 (10.9%)
3	2 (0.2%)
Not determined	29 (2.3%)
Histological findings	
Grade of activity, n (%)	
A0	154 (12%)
A1	574 (45%)
A2	441 (34%)
A3	110 (9%)
Stage of fibrosis, n (%)	
F0	24 (2%)
F1	591 (46%)
F2	378 (30%)
F3	242 (19%)
F4	44 (3%)
Grade of steatosis, n (%)	
0%	384 (30%)
1–9%	543 (42%)
10–29%	215 (17%)
≥30%	137 (11%)
SVR to interferon therapy, n (%)	393 (31%)
Development of HCC, n (%)	68 (5%)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; GGT,  $\gamma$ -glutamyltransferase; HCC, hepatocellular carcinoma; SVR, sustained virological response.

angiography or tumor biopsy to confirm the diagnosis. Diagnostic criteria of HCC on radiological findings were hyper-vascularity at angiography or hyper-attenuation at triphasic contrast-enhanced computerized tomography or magnetic resonance imaging during the hepatic arterial phase.

#### Statistical analysis

The SPSS software package ver. 15.0 was used for statistical analysis. Categorical data were analyzed using Fisher's exact test. Continuous variables were compared with Student's *t*-test. The time for the development of HCC was defined as the time from the completion of IFN

therapy to the time of diagnosis. Annual incidence of HCC was calculated using the person-years method. Effect of hepatic steatosis on time to development of HCC was analyzed by the Kaplan–Meier method and log–rank test, after stratification by age, sex, BMI, degree of fibrosis and response to IFN therapy, as well as multivariate analysis using Cox proportional hazards regression analysis. A *P*-value of less than 0.05 was considered statistically significant.

## RESULTS

### Background factors for steatosis

PATIENTS WITH A steatosis grade of 10% or greater were older (53.6 ± 12.6 vs 56.0 ± 9.8, *P* = 0.001), had a higher BMI (23.0 ± 3.0 vs 24.6 ± 3.3, *P* < 0.0001), higher frequency of diabetes (12% vs 24%, *P* < 0.0001), higher serum levels of aspartate aminotransferase (AST) (66 ± 46 vs 75 ± 43, *P* = 0.002),  $\gamma$ -glutamyltransferase (GGT) (37 ± 52 vs 52 ± 33, *P* < 0.0001), total cholesterol (173 ± 32 vs 179 ± 33, *P* = 0.005), triglycerides (123 ± 56 vs 145 ± 68, *P* < 0.0001), and a lower serum level of albumin (4.2 ± 0.3 vs 4.1 ± 0.3, *P* = 0.005) and lower platelet counts (16.6 ± 5.2 vs 15.7 ± 5.1, *P* = 0.007). Histological grade of activity (A2–3: 39% vs 54%, *P* < 0.0001), and stage of fibrosis (F3–4: 18% vs 34%, *P* < 0.0001) were higher. The proportion of non-sustained virological response (SVR) to IFN also was higher (35% vs 19%, *P* < 0.0001). These results indicate that hepatic steatosis in hepatitis C is related to metabolic factors and associated with other risk factors for the development of HCC such as older age, advanced stage of fibrosis, and non-SVR to IFN therapy.

### Factors associated with the development of HCC

Hepatocellular carcinoma developed in 68 patients during follow up. An overall annual incidence of HCC development was 1.19% by person-years. The annual incidence of HCC development by person-years was higher in patients with higher grade of steatosis: 0.45% for patients without steatosis, 0.78% for patients with 1–9% of steatosis, 2.30% for patients with 10–29% of steatosis, and 3.56% for patients with 30% of steatosis. The relative risk of hepatic steatosis (grade of ≥10%) for HCC development was 4.39 (95% confidence interval 2.66–7.26, *P* < 0.0001). The difference remained significant, even after stratification for other risk factors such as IFN therapy, stage of fibrosis, age, sex and BMI (Fig. 1). When analyzed by the multivariate Cox proportional

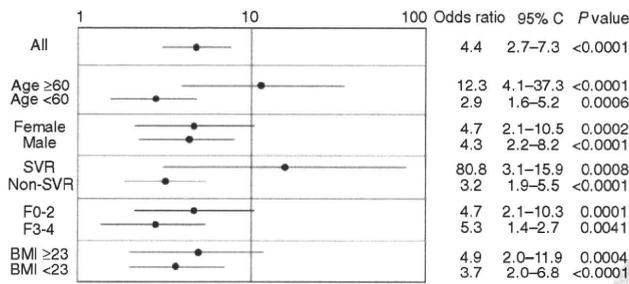


Figure 1 Relative risk differences of hepatocellular carcinoma (HCC) among patients with and without steatosis. The relative risk of hepatic steatosis (grade ≥10%) for HCC development was analyzed, after stratification for other risk factors such as interferon (IFN) therapy, stage of fibrosis, age, sex and body mass index (BMI). SVR, sustained virological response.

hazards regression method, a higher grade of steatosis, older age, male sex, higher BMI, an advanced stage of fibrosis and non-SVR to IFN therapy were independent risk factors associated with the development of HCC (Table 2). The adjusted risk ratio of hepatic steatosis was 3.04 (95% confidence interval 1.82-5.06,  $P < 0.0001$ ). The presence of diabetes and consumption of ethanol were not significant. Figure 2(a) shows the Kaplan-Meier curve of the time to development of HCC in the entire cohort. The cumulative incidence of HCC was significantly higher with hepatic steatosis of 10% or greater. To adjust for other risk factors, patients were stratified according to response to IFN therapy, stage of fibrosis, age, sex and BMI. The difference remained sig-

nificant, even after stratification for these confounding factors (Fig. 2b-f). Three patients died after the development of HCC. All were over 60 years old, and had significant steatosis. The impact of hepatic steatosis on the survival rate could not be analyzed due to the small number of death.

### DISCUSSION

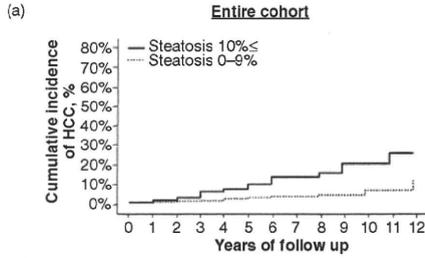
IN THIS STUDY, we have shown that the presence of significant steatosis is an independent risk factor for the development of HCC in chronic hepatitis C. Our study involved the largest number of patients, compared to previous reports, and this enabled us to adjust for

Table 2 Multivariate analysis of risk factors for hepatocellular carcinoma

Predictor		Odds ratio (95% CI)	P-value
Age	By every 10 years	1.09 (1.05-1.13)	<0.0001
Sex	Male vs female	2.12 (1.28-3.51)	0.004
Stage of fibrosis	F3-4 vs F0-2	4.30 (2.59-7.14)	<0.0001
Grade of steatosis	≥10% vs <10%	3.04 (1.82-5.06)	<0.0001
Response to IFN	Non-SVR vs SVR	2.43 (1.13-5.23)	0.023
Diabetes	Present vs absent	0.75 (0.42-1.33)	0.319
Ethanol consumption (g/day)	≥20 vs <20	0.50 (0.07-3.60)	0.478
BMI (kg/m <sup>2</sup> )	≥23 vs <23	1.69 (1.02-2.86)	0.043

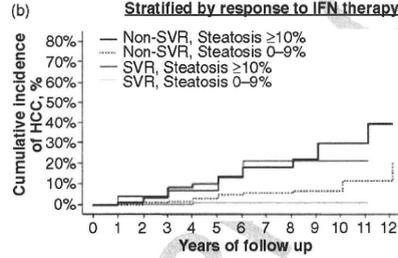
BMI, body mass index; CI, confidence interval; IFN, interferon; SVR, sustained virological response.

Figure 2 Cumulative incidence of hepatocellular carcinoma (HCC) among patients with steatosis (solid line) and without steatosis (dotted line), stratified by other risk factors. The cumulative incidence of HCC was (a) significantly higher in patients with a steatosis grade of 10% or greater ( $P < 0.0001$  by the log-rank test), even after (b) stratification by the response to interferon therapy ( $P < 0.0001$  for sustained virological response [SVR] and non-SVR by the log-rank test), (c) stratification by the stage of fibrosis ( $P < 0.0001$  for F0-2 and  $P = 0.0036$  for F3-4 by the log-rank test), (d) stratification by age ( $P = 0.0001$  for age ≥60 and  $P < 0.0001$  for age <60 by the log-rank test), (e) stratification by sex ( $P < 0.0001$  for men and women by the log-rank test), and (f) stratification by body mass index (BMI) ( $P < 0.0001$  for BMI ≥23 kg/m<sup>2</sup> and <23 kg/m<sup>2</sup> by the log-rank test). The number of patients at risk is shown below each graph.



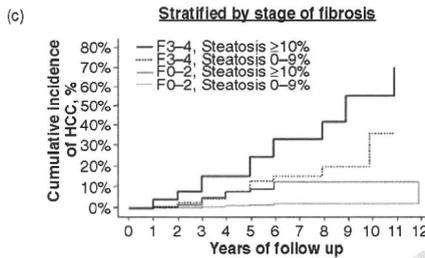
**Number of patients at risk**

Steatosis 0-9%	927	824	620	503	320	227	161	117	77	49	27	10
Steatosis ≥10%	352	271	207	157	113	83	54	48	32	17	9	1



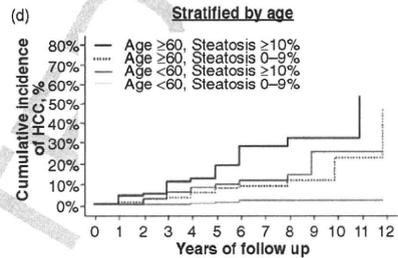
**Number of patients at risk**

SVR												
Steatosis 0-9%	326	254	204	153	81	55	33	21	15	10	5	0
Steatosis ≥10%	67	50	34	22	14	10	4	4	4	2	2	0
Non-SVR												
Steatosis 0-9%	601	507	416	350	239	172	128	96	62	39	22	10
Steatosis ≥10%	285	221	173	135	99	73	50	44	28	15	7	1



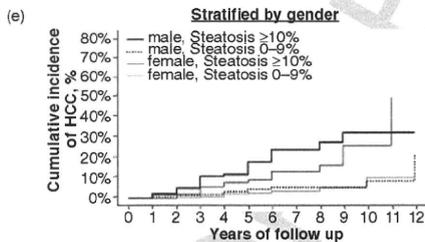
**Number of patients at risk**

F0-2												
Steatosis 0-9%	759	623	509	415	266	188	137	99	64	39	25	10
Steatosis ≥10%	234	190	146	107	77	55	37	32	19	11	6	1
F3-4												
Steatosis 0-9%	118	81	61	50	36	28	17	16	13	6	3	0
Steatosis ≥10%	168	138	111	88	54	39	23	18	13	10	2	0



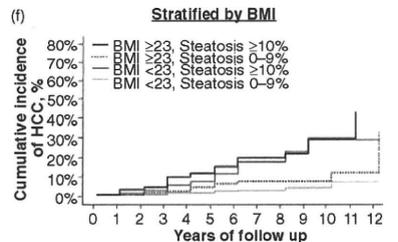
**Number of patients at risk**

Age <60												
Steatosis 0-9%	549	457	367	298	188	148	111	83	53	33	19	7
Steatosis ≥10%	193	154	111	83	61	48	34	31	23	12	6	1
Age ≥60												
Steatosis 0-9%	378	304	253	205	132	79	50	34	24	16	8	3
Steatosis ≥10%	159	117	96	74	52	35	20	17	9	5	3	0



**Number of patients at risk**

Male												
Steatosis 0-9%	470	389	319	265	169	126	90	65	46	30	17	7
Steatosis ≥10%	173	134	98	73	54	40	21	21	15	8	6	1
Female												
Steatosis 0-9%	457	372	301	238	151	101	71	52	31	19	10	3
Steatosis ≥10%	179	137	109	84	59	43	33	27	17	9	3	0



**Number of patients at risk**

BMI ≥23												
Steatosis 0-9%	417	346	269	213	129	94	66	49	31	19	8	4
Steatosis ≥10%	226	176	137	101	71	55	34	33	20	10	5	0
BMI <23												
Steatosis 0-9%	510	415	351	290	191	133	95	68	46	30	19	6
Steatosis ≥10%	126	95	70	56	42	28	20	15	12	7	4	1

1 other known risk factors for HCC. The impact of steato-  
2 sis on HCC development remained significant even after  
3 adjusting for other risk factors such as older age, male  
4 sex, higher BMI, advanced fibrosis and non-SVR to IFN  
5 therapy. These findings indicate the need of intensive  
6 surveillance for HCC in patients with significant steato-  
7 sis and provide an argument for therapeutic interven-  
8 tions aimed at reducing steatosis, in order to reduce the  
9 risk of HCC.

10 The association between hepatic steatosis and the  
11 development of HCC in chronic hepatitis C has been  
12 proposed and the possible mechanism has been dis-  
13 cussed.<sup>16</sup> There are several cohort studies on this topic  
14 but their results are conflicting. The first report  
15 included 20 patients with SVR to IFN, 51 patients with  
16 non-SVR to IFN and 90 patients who did not receive  
17 IFN therapy.<sup>17</sup> In this cohort of 161 patients, older age,  
18 absence of IFN therapy, cirrhosis and steatosis were  
19 associated with HCC development. Another study  
20 involved 25 patients with HCC and an equal number  
21 of patients who did not develop HCC, matched for  
22 age, sex, HCV genotype and stage of fibrosis.<sup>19</sup> In this  
23 study, only ALT and albumin were identified as predic-  
24 tors of HCC and steatosis was not. The authors  
25 acknowledged the small size of the cohort as a limita-  
26 tion and emphasized the need for larger cohort  
27 studies. The third study analyzed explanted liver from  
28 cirrhotic patients who underwent liver transplantation  
29 and included 32 patients with HCC and 62 patients  
30 without HCC.<sup>18</sup> The authors found that older age,  
31 higher  $\alpha$ -fetoprotein levels and steatosis were signifi-  
32 cantly associated with HCC. The major advantage of  
33 this study was the standardization of fibrosis stage to  
34 cirrhosis. On the other hand, a limitation was the ret-  
35 rospective nature of the study; steatosis was evaluated  
36 after the diagnosis of HCC, when cirrhosis already was  
37 present (fibrosis stage F4). Because steatosis has been  
38 reported to decrease once cirrhosis has developed, this  
39 study may have underestimated the grade of steatosis  
40 present prior to the development of HCC. Thus, we  
41 cannot simply apply their findings to a clinical setting  
42 where biopsies are usually obtained before the devel-  
43 opment of cirrhosis and years before the development  
44 of HCC. Based on that background, the principal aim  
45 of this study was to analyze the association between  
46 hepatic steatosis and the development of HCC in  
47 chronic hepatitis C patients, adjusting for known risk  
48 factors. We found that steatosis was an independent  
49 risk factor by the multivariate Cox proportional  
50 hazards regression analysis and by the Kaplan-Meier  
51 method and log-rank test after stratification by other

52 risk factors. To our surprise, the adjusted risk ratio of  
53 hepatic steatosis was higher than that of older age,  
54 male sex, non-SVR to IFN and higher BMI.

55 How steatosis contributes to the development of HCC  
56 remains unclear. Several studies including ours,<sup>10</sup> indi-  
57 cated that hepatic steatosis promotes the progression of  
58 hepatic fibrosis,<sup>11-15</sup> which potentiates the risk of HCC  
59 indirectly. On the other hand, the ob/ob mouse model  
60 of NAFLD showed that hepatic neoplasia developed in  
61 the absence of advanced fibrosis, supporting the concept  
62 that metabolic abnormalities related to obesity initiate  
63 the neoplastic process.<sup>8</sup> Leptin, an adipocytokine related  
64 to steatosis in chronic hepatitis C,<sup>21</sup> was shown recently  
65 to be mitogenic in human liver<sup>22</sup> and thus may be a link  
66 between steatosis and HCC development. Otherwise,  
67 steatosis may be responsible for increased lipid peroxi-  
68 dation and reactive oxygen species which induce genetic  
69 damage.<sup>23-25</sup> Another study showed that mice transgenic  
70 for the HCV core gene developed hepatic steatosis early  
71 in life and thereafter HCC which indicates that the HCV  
72 core protein has a chief role in the development of both  
73 steatosis and HCC development.<sup>26</sup> The precise mecha-  
74 nism of the association between steatosis and carcino-  
75 genesis needs further investigation.

76 The higher incidence of HCC in patients with signifi-  
77 cant steatosis has important clinical implications. The  
78 most important question is whether therapeutic inter-  
79 ventions aimed at reducing steatosis could reduce the  
80 risk of HCC in chronic hepatitis C. Because the adjusted  
81 risk ratio of hepatic steatosis was higher than that of  
82 older age, male sex, non-SVR to IFN and higher BMI,  
83 we hypothesize that modification of lifestyle and the  
84 amelioration of hepatic steatosis may efficiently prevent  
85 hepatocarcinogenesis in patients having concomitant  
86 risk factors. Apparently, further prospective studies  
87 focusing on this point are necessary. Weight reduction  
88 may provide an important treatment strategy because  
89 one study indicated that weight reduction in chronic  
90 hepatitis C leads to a reduction in steatosis and an  
91 improvement in fibrosis despite the persistence of HCV  
92 infection.<sup>27</sup> Alternatively, insulin resistance may be  
93 another target of therapy because a study showed that  
94 the administration of pioglitazone led to metabolic  
95 and histological improvement in subjects with non-  
96 alcoholic steatohepatitis.<sup>28</sup> A limitation of the present  
97 study was that data for the plasma insulin concentration  
98 was not available and thus insulin resistance could not  
99 be assessed. Whether insulin resistance plays a role in  
100 hepatocarcinogenesis or its amelioration could improve  
101 steatosis and ultimately prevent development of HCC in  
102 chronic hepatitis C awaits future investigation.

Another important finding of the present study was that steatosis was a significant risk factor for the development of HCC in patients with SVR to IFN therapy. Thus, steatosis may play a role in carcinogenesis in patients who have cleared HCV. Several studies have shown that the incidence of HCC is reduced but not eliminated in those with SVR to IFN.<sup>29-31</sup> Because the predictors of HCC development in SVR patients have not been established to date, steatosis may be used to identify patients who need intensive surveillance and long-term follow up, even after the clearance of HCV. In conclusion, we showed that hepatic steatosis is significantly associated with the development of HCC in chronic hepatitis C independent of age, sex, BMI, degree of fibrosis and response to previous IFN therapy. Steatosis may be a useful marker for identifying patients at higher risk for HCC. Further studies are needed to evaluate the hypothesis that therapeutic interventions aimed at reducing steatosis may prevent hepatocarcinogenesis.

#### ACKNOWLEDGMENTS

THIS STUDY WAS supported by a Grant-in-Aid from the Ministry of Health, Labor and Welfare, Japan.

#### REFERENCES

- 1 El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 2007; 132: 2557-76.
- 2 Shimada M, Hashimoto E, Taniai M *et al.* Hepatocellular carcinoma in patients with non-alcoholic steatohepatitis. *J Hepatol* 2002; 37: 154-60.
- 3 Bugianesi E, Leone N, Vanni E *et al.* Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology* 2002; 123: 134-40.
- 4 Marrero JA, Fontana RJ, Su GL, Conjeevaram HS, Emick DM, Lok AS. NAFLD may be a common underlying liver disease in patients with hepatocellular carcinoma in the United States. *Hepatology* 2002; 36: 1349-54.
- 5 Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* 2003; 348: 1625-38.
- 6 El-Serag HB, Tran T, Everhart JE. Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. *Gastroenterology* 2004; 126: 460-8.
- 7 Davila JA, Morgan RO, Shaib Y, McGlynn KA, El-Serag HB. Diabetes increases the risk of hepatocellular carcinoma in the United States: a population based case control study. *Gut* 2005; 54: 533-9.
- 8 Yang S, Lin HZ, Hwang J, Chacko VP, Diehl AM. Hepatic hyperplasia in noncirrhotic fatty livers: is obesity-related hepatic steatosis a premalignant condition? *Cancer Res* 2001; 61: 5016-23.
- 9 Lefkowitz JH, Schiff ER, Davis GL *et al.* Pathological diagnosis of chronic hepatitis C: a multicenter comparative study with chronic hepatitis B. The Hepatitis Interventional Therapy Group. *Gastroenterology* 1993; 104: 595-603.
- 10 Kurosaki M, Matsunaga K, Hirayama I *et al.* The presence of steatosis and elevation of alanine aminotransferase levels are associated with fibrosis progression in chronic hepatitis C with non-response to interferon therapy. *J Hepatol* 2008; 48: 736-42.
- 11 Hourigan LF, Macdonald GA, Purdie D *et al.* Fibrosis in chronic hepatitis C correlates significantly with body mass index and steatosis. *Hepatology* 1999; 29: 1215-19.
- 12 Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* 2001; 33: 1358-64.
- 13 Westin J, Nordlinder H, Lagging M, Norkrans G, Wejstal R. Steatosis accelerates fibrosis development over time in hepatitis C virus genotype 3 infected patients. *J Hepatol* 2002; 37: 837-42.
- 14 Fartoux L, Chazouilleres O, Wendum D, Poupon R, Serfaty L. Impact of steatosis on progression of fibrosis in patients with mild hepatitis C. *Hepatology* 2005; 41: 82-7.
- 15 Leandro G, Mangia A, Hui J *et al.* Relationship between steatosis, inflammation, and fibrosis in chronic hepatitis C: a meta-analysis of individual patient data. *Gastroenterology* 2006; 130: 1636-42.
- 16 Koike K. Hepatitis C virus contributes to hepatocarcinogenesis by modulating metabolic and intracellular signaling pathways. *J Gastroenterol Hepatol* 2007; 22 (Suppl 1): S108-11.
- 17 Ohata K, Hamasaki K, Toriyama K *et al.* Hepatic steatosis is a risk factor for hepatocellular carcinoma in patients with chronic hepatitis C virus infection. *Cancer* 2003; 97: 3036-43.
- 18 Pekow JR, Bhan AK, Zheng H, Chung RT. Hepatic steatosis is associated with increased frequency of hepatocellular carcinoma in patients with hepatitis C-related cirrhosis. *Cancer* 2007; 109: 2490-6.
- 19 Kumar D, Farrell GC, Kench J, George J. Hepatic steatosis and the risk of hepatocellular carcinoma in chronic hepatitis C. *J Gastroenterol Hepatol* 2005; 20: 1395-400.
- 20 Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; 24: 289-93.
- 21 Romero-Gomez M, Castellano-Megias VM, Grande L *et al.* Serum leptin levels correlate with hepatic steatosis in chronic hepatitis C. *Am J Gastroenterol* 2003; 98: 1135-41.
- 22 Ramani K, Yang H, Xia M, Ara AI, Mato JM, Lu SC. Leptin's mitogenic effect in human liver cancer cells requires induc-

- 1 tion of both methionine adenosyltransferase 2A and 2beta. 19  
2 *Hepatology* 2008; 47: 521–31. 20  
3 23 Okuda M, Li K, Beard MR *et al.* Mitochondrial injury, oxi- 21  
4 dative stress, and antioxidant gene expression are induced 22  
5 by hepatitis C virus core protein. *Gastroenterology* 2002; 23  
6 122: 366–75. 24  
7 24 Cai D, Yuan M, Frantz DF *et al.* Local and systemic insulin 25  
8 resistance resulting from hepatic activation of IKK-beta and 26  
9 NF-kappaB. *Nat Med* 2005; 11: 183–90. 27  
10 25 Arkan MC, Hevener AL, Greten FR *et al.* IKK-beta links 28  
11 inflammation to obesity-induced insulin resistance. *Nat 29*  
12 *Med* 2005; 11: 191–8. 30  
13 26 Moriya K, Fujie H, Shintani Y *et al.* The core protein of 31  
14 hepatitis C virus induces hepatocellular carcinoma in 32  
15 transgenic mice. *Nat Med* 1998; 4: 1065–7. 33  
16 27 Hickman IJ, Clouston AD, Macdonald GA *et al.* Effect of 34  
17 weight reduction on liver histology and biochemistry in 35  
18 patients with chronic hepatitis C. *Gut* 2002; 51: 89–94.
- 28 Belfort R, Harrison SA, Brown K *et al.* A placebo-controlled 19  
trial of pioglitazone in subjects with nonalcoholic steato- 20  
hepatitis. *N Engl J Med* 2006; 355: 2297–307. 21  
29 Yoshida H, Shiratori Y, Moriyama M *et al.* Interferon 22  
therapy reduces the risk for hepatocellular carcinoma: 23  
national surveillance program of cirrhotic and noncirrhotic 24  
patients with chronic hepatitis C in Japan. IHIT Study 25  
Group. Inhibition of Hepatocarcinogenesis by Interferon 26  
Therapy. *Ann Intern Med* 1999; 131: 174–81. 27  
30 Nishiguchi S, Shiomi S, Nakatani S *et al.* Prevention of 28  
hepatocellular carcinoma in patients with chronic active 29  
hepatitis C and cirrhosis. *Lancet* 2001; 357: 196–7. 30  
31 Shiratori Y, Ito Y, Yokosuka O *et al.* Antiviral therapy for 31  
cirrhotic hepatitis C: association with reduced hepatocel- 32  
lular carcinoma development and improved survival. *Ann 33*  
*Intern Med* 2005; 142: 105–14. 34

Toppan Best-set Premedia Limited	
Journal Code: HEPR	Proofreader: Momy
Article No: 692	Delivery date: 19 July 2010
Page Extent: 8	

## AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remark
q1	AUTHOR: Please provide the title for corresponding author, e.g. Dr, Professor.	
q2	AUTHOR: As per journal style, 3-6 keywords should be kept in the article. Please remove one or more keywords accordingly.	

## Sequences in the Interferon Sensitivity-Determining Region and Core Region of Hepatitis C Virus Impact Pretreatment Prediction of Response to PEG-Interferon Plus Ribavirin: Data Mining Analysis

Masayuki Kurosaki,<sup>1</sup> Naoya Sakamoto,<sup>2</sup> Manabu Iwasaki,<sup>3</sup> Minoru Sakamoto,<sup>4</sup> Yoshiyuki Suzuki,<sup>5</sup> Naoki Hiramatsu,<sup>6</sup> Fuminaka Sugauchi,<sup>7</sup> Akihiro Tamori,<sup>8</sup> Mina Nakagawa,<sup>2</sup> and Namiki Izumi, MD<sup>1\*</sup>

<sup>1</sup>Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

<sup>2</sup>Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan

<sup>3</sup>Department of Computer and Information Science, Seikei University, Tokyo, Japan

<sup>4</sup>First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan

<sup>5</sup>Department of Hepatology, Toranomon Hospital, Tokyo, Japan

<sup>6</sup>Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan

<sup>7</sup>Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

<sup>8</sup>Department of Hepatology, Osaka City University Medical School, Osaka, Japan

The aim of the present study was to clarify the significance of viral factors for pretreatment prediction of sustained virological response to pegylated-interferon (PEG-IFN) plus ribavirin (RBV) therapy for chronic hepatitis C using data mining analysis. Substitutions in the IFN sensitivity-determining region (ISDR) and at position 70 of the HCV core region (Core70) were determined in 505 patients with genotype 1b chronic hepatitis C treated with PEG-IFN plus RBV. Data mining analysis was used to build a predictive model of sustained virological response in patients selected randomly ( $n = 304$ ). The reproducibility of the model was validated in the remaining 201 patients. Substitutions in ISDR (odds ratio = 9.92,  $P < 0.0001$ ) and Core70 (odds ratio = 1.92,  $P = 0.01$ ) predicted sustained virological response independent of other covariates. The decision-tree model revealed that the rate of sustained virological response was highest (83%) in patients with two or more substitutions in ISDR. The overall rate of sustained virological response was 44% in patients with a low number of substitutions in ISDR (0–1) but was 83% in selected subgroups of younger patients (<60 years), wild-type sequence at Core70, and higher level of low-density lipoprotein cholesterol (LDL-C) ( $\geq 120$  mg/dl). Reproducibility of the model was validated ( $r^2 = 0.94$ ,  $P < 0.001$ ). In conclusion, substitutions in ISDR and Core70 of

HCV are significant predictors of response to PEG-IFN plus RBV therapy. A decision-tree model that includes these viral factors as predictors could identify patients efficiently with a high probability of sustained virological response.

**J. Med. Virol.** 9999:1–8, 2010.

© 2010 Wiley-Liss, Inc.

**KEY WORDS:** data mining; decision-tree model; ISDR; core region; PEG-interferon

### INTRODUCTION

The combination of pegylated-interferon (PEG-IFN) plus ribavirin (RBV) is currently the most effective therapy for chronic hepatitis C, but the rate of sustained virological response after 48 weeks of therapy is about 50% in patients with HCV genotype 1b and a high HCV

Grant sponsor: Ministry of Health, Labor and Welfare, Japan.

The authors report no conflicts of interest.

\*Correspondence to: Namiki Izumi, MD, Division<sup>Q1</sup> of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. E-mail: nizumi@musashino.jrc.or.jp

Accepted XX XX XXXX

DOI 10.1002/jmv.00000

Published online in Wiley Online Library  
(wileyonlinelibrary.com).

© 2010 WILEY-LISS, INC.

RNA titer [Manns et al., 2001; Fried et al., 2002]. The most reliable means to predict sustained virological response is to monitor the viral response during the early weeks of treatment. The early virological response, defined as undetectable HCV RNA at week 12, is associated with a high rate of sustained virological response [Davis et al., 2003; Lee and Ferenci, 2008]. The rapid virological response, defined as undetectable HCV RNA at week 4 of therapy, is even more predictive of sustained virological response than the early virological response [Jensen et al., 2006; Yu et al., 2008; Izumi et al., 2010]. However, there is no established means that predicts the virological response before commencing treatment. Recent reports have revealed that single nucleotide polymorphisms located near the *IL28B* gene show a strong association with the response to PEG-IFN plus RBV therapy [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Kurosaki et al., 2010c]. These findings indicate that the host factor is an important determinant of the treatment response. On the other hand, the present study's authors have reported that a stretch of 40 amino acids in the NS5A region of HCV, designated as the interferon sensitivity-determining region (ISDR), has a close association with the virological response to interferon mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997]. More recently, amino acid substitutions at positions 70 and 91 of the core region have been reported to be associated with response to PEG-IFN plus RBV combination therapy [Akuta et al., 2005, 2007a]. The impact of these HCV substitutions on treatment response is yet to be validated.

Decision-tree analysis is a core component of data mining analysis that can be used to build predictive models [Breiman et al., 1980]. This method has been used to define prognostic factors in various diseases such as prostate cancer [Garzotto et al., 2005], diabetes [Miyaki et al., 2002], melanoma [Averbook et al., 2002; Leiter et al., 2004], colorectal carcinoma [Zlobec et al., 2005; Valera et al., 2007], and liver failure [Baquerizo et al., 2003]. The major advantage of decision-tree analysis over logistic regression analysis is that the results of analysis are easy to understand. The simple allocation of patients into subgroups by following the flowchart form could define the predicted possibility of outcome [LeBlanc and Crowley, 1995].

Decision-tree analysis was used for the prediction of early virological response (undetectable HCV RNA within 12 weeks of therapy) to PEG-IFN and RBV combination therapy in chronic hepatitis C [Kurosaki et al., 2010a], and more recently for the pretreatment prediction of sustained virological response [Kurosaki et al., 2010b]. In the latter model, simple and noninvasive standard tests were used as parameters; specialized tests such as viral mutations and host genetics, or invasive tests such as liver histology, were not included because the aim of that model was for use in general medical practice, especially in some countries or areas where resources are limited. Thus, the impact of viral mutations or liver histology was not considered in that model.

The present study examined whether including viral substitutions in ISDR and the core region of HCV in the decision-tree model could improve its predictive accuracy over the previous model to identify chronic hepatitis C patients who are likely to respond to PEG-IFN plus RBV therapy.

## MATERIALS AND METHODS

### Patients

This multicenter retrospective cohort study included 505 chronic hepatitis C patients who were treated with PEG-IFN alpha-2b and RBV at Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Osaka University, Nagoya City University Graduate School of Medical Sciences, Yamanashi University, Osaka City University, and their related hospitals. The inclusion criteria were: (1) genotype 1b, (2) HCV RNA titer higher than 100  $\text{kIU}^{Q2}/\text{ml}$  by quantitative PCR (Cobas Amplifier HCV Monitor v 2.0, Roche Diagnostic Systems, CA<sup>Q3</sup>), (3) no co-infection with hepatitis B virus or human immunodeficiency virus, (4) no other causes of liver disease, (5) patients having undergone liver biopsy prior to IFN treatment, (6) number of substitutions in ISDR having been determined, (7) substitutions in the amino acid positions 70 and 91 of the core region having been determined, and (8) completion of at least 12 weeks of therapy. Patients were treated with PEG-IFN alpha-2b (1.5  $\mu\text{g}/\text{kg}$ ) weekly plus RBV. The daily dose of RBV was adjusted by weight: 600 mg for <60 kg, 800 mg for 60–80 kg, and 1,000 mg for >80 kg. For the analysis, patients were assigned randomly to either the model building (304 patients) or validation (201 patients) groups. There were no significant differences in the clinical backgrounds between these two groups (Table I). Informed consent was obtained from each patient. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional review committees of all concerned hospitals.

### Laboratory Tests

Hematological tests, blood chemistry, and HCV RNA titer were analyzed before therapy and at least once every month during therapy. Sequences of ISDR and the core region of HCV were determined by direct sequencing after amplification by reverse transcription and polymerase chain reaction as reported previously. At position 70 of the core region (Core70), arginine was defined as the wild type, and glutamine or histidine was defined as the mutant type. At position 91 of the core region, leucine was defined as the wild type and methionine was defined as the mutant type, as described previously [Akuta et al., 2005]. Fibrosis and activity were scored according to the METAVIR scoring system [Bedossa and Poynard, 1996]. Fibrosis was staged on a scale of 0–4: F0 (no fibrosis), F1 (mild fibrosis), F2 (moderate fibrosis), F3 (severe fibrosis), and F4 (cirrhosis). Activity of necroinflammation was graded on a scale of

TABLE I. Comparison of Pretreatment Factors Between Model Building and Validation Patients

	Model (n = 304)	Validation (n = 201)	P-value
Age (years)	55.6 (9.4)	56.0 (12.2)	0.80
Male (%)	53 (%)	55 (%)	0.13
Body mass index (kg/m <sup>2</sup> )	23.1 (3.1)	23.1 (4.0)	0.99
Albumin (g/dl)	4.0 (0.3)	4.0 (0.3)	0.47
Creatinine (mg/dl)	0.72 (0.15)	0.72 (0.14)	0.62
AST (IU/L)	63.3 (45.6)	58.9 (46.4)	0.91
ALT (IU/L)	78.7 (58.6)	74.5 (67.5)	0.68
GGT (IU/L)	53.2 (49.1)	57.4 (63.5)	0.43
Total cholesterol (mg/dl)	170.9 (32.6)	169.4 (34.1)	0.33
Triglyceride (mg/dl)	107.0 (44.7)	105.7 (48.0)	0.90
LDL-C (mg/dl)	95.5 (28.0)	96.4 (28.8)	0.34
White blood cell count (/μl)	4,902 (1,489)	4,906 (1,319)	0.86
Hemoglobin (g/dl)	14.1 (1.3)	14.3 (1.4)	0.09
Platelets (10 <sup>9</sup> /L)	164 (56)	172 (55)	0.68
HCV RNA (10 <sup>3</sup> IU/ml)	1,859 (1,468)	2,021 (1,393)	0.09
ISDR mutations: ≥2 (%)	15 (%)	20 (%)	0.11
Core70: mutant (%)	36 (%)	29 (%)	0.22
Core91: mutant (%)	40 (%)	36 (%)	0.20
Fibrosis: F2-4 (%)	49 (%)	48 (%)	0.36
Activity: A2-3 (%)	42 (%)	34 (%)	0.10

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region. Data expressed as mean (SD).

0-3: A0 (no activity), A1 (mild activity), A2 (moderate activity), and A3 (severe activity). Sustained virological response was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems) at week 24 after the completion of therapy.

### Statistical Analysis

A database of pretreatment variables included hematological tests (hemoglobin level, white blood cell count, and platelet count), blood chemistry tests (serum levels of creatinine, albumin, aspartate aminotransferase, alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), total cholesterol, triglyceride, and low-density lipoprotein cholesterol (LDL-C)), viral factors (HCV RNA titer, number of substitutions in ISDR, substitutions in the amino acid positions 70 and 91 of the core region), histological findings (stage of fibrosis and grade of activity) and patient characteristics (age, sex, and body mass index). Based on this database, decision-tree analysis was used to define a predictive model for sustained virological response.

Student's *t*-test was used for the univariable comparison of quantitative variables and Fisher's exact test was used for the comparison of qualitative variables. For the multivariable analysis for factors associated with sustained virological response, logistic regression models with backward selection were used to identify independent predictors of sustained virological response. Variables that showed significant association with sustained virological response by univariable analysis were included in the multivariable analysis. IBM-SPSS software v.15.0 (SPSS, Inc., Chicago, IL) was used for these analyses. For the decision-tree analysis

[Segal<sup>Q4</sup> and Bloch, 1989], the data mining software IBM SPSS Modeler 13 (IBM SPSS, Inc.) was used, as reported previously [Kurosaki et al., 2010a,b]. In brief, the software searched for the optimal split variables to build a decision-tree structure. The entire study population was first evaluated to determine the variables and cut-off points for the most significant division into two subgroups having different probabilities of sustained virological response. Thereafter, analysis was repeated on all subgroups in the same way until either no additional significant variable was detected or the sample size was below 20.

## RESULTS

### Generation of the Decision-Tree Model

The decision-tree analysis selected five predictive variables to produce six subgroups of patients (Fig. 1). The number of substitutions in ISDR was selected as the best predictor of sustained virological response. The possibility of achieving sustained virological response was 83% for patients with two or more substitutions in ISDR compared with 44% for patients with a single or no substitution. Among patients with a single or no substitution in ISDR, age, with an optimal cut-off of 60 years, was selected as the variable of second split. Patients younger than 60 had the higher probability of sustained virological response (55%) compared with those older than 60 years (31%). Among younger patients, amino acid substitution at Core70 was selected as the third variable of split—wild-type sequence being the predictor of favorable response compared with the mutant type (65% vs. 36%). Among patients with wild-type Core70, the level of serum LDL-C was selected as the fourth variable of split, with an optimal cutoff of