

38. Tong X, Arasappan A, Bennett F, et al. Preclinical characterization of the antiviral activity of SCH 900518 (Narlaprevir), a novel mechanism-based inhibitor of hepatitis C virus NS3 protease. *Antimicrob Agents Chemother* 2010;54:2365–2370
39. Chase R, Skelton A, Xia E, et al. A novel HCV NS3 protease mutation selected by combination treatment of the protease inhibitor boceprevir and NS5B polymerase inhibitors. *Antiviral Res* 2009;84:178–184

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,¹ Naoya Sakamoto,² Manabu Iwasaki,³ Minoru Sakamoto,⁴ Yoshiyuki Suzuki,⁵ Naoki Hiramatsu,⁶ Fuminaka Sugauchi,⁷ Akihiro Tamori,⁸ Mina Nakagawa,² and Namiki Izumi, MD^{1*}

¹Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

²Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan

³Department of Computer and Information Science, Seikei University, Tokyo, Japan

⁴First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan

⁵Department of Hepatology, Toranomon Hospital, Tokyo, Japan

⁶Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan

⁷Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan

⁸Department 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) (≥ 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 with a high probability of sustained virological response. *J. Med. Virol.* 83:445–452, 2011.

© 2011 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 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 26 October 2010

DOI 10.1002/jmv.22005

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

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 kIU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic Systems, Pleasanton, CA), (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 µg/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 ²)	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 ⁹ /L)	164 (56)	172 (55)	0.68
HCV RNA (10 ³ 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 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

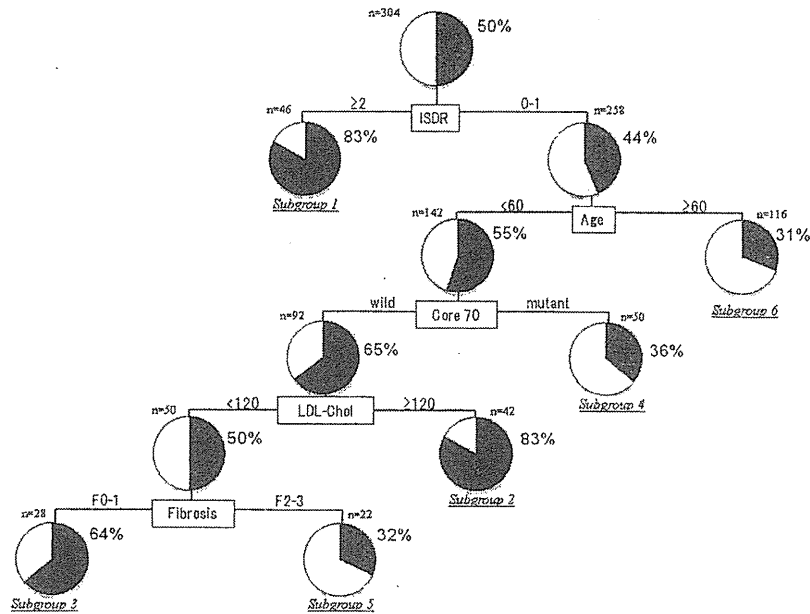


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to 7. The rate of sustained virological response was >80% in subgroups 1 and 2, 64% in subgroup 3, and 31–36% in subgroups 4, 5, and 6. LDL-C represents low-density lipoprotein cholesterol and Core70 represents amino acid substitution at position 70 of the core region.

120 mg/dl. Patients with higher LDL-C level had the higher probability of sustained virological response (83% vs. 50%). The stage of fibrosis was selected as the final variable of split, with significant fibrosis (F2–4) being the predictor of lower sustained virological response probability (64% vs. 32%).

Among the six subgroups derived by this decision tree, the subgroup of patients with two or more substitutions in ISDR (subgroup 1) or with a single or no substitution in ISDR but younger than 60 years of age, having the wild-type Core70 and high serum level of LDL-C (≥ 120 mg/dl) (subgroup 2) showed the highest probability of sustained virological response (83%).

Validation of the Decision-Tree Model

The decision-tree model was validated using a validation dataset of 201 cases that were not included in the model-building dataset. Each patient in the validation set was allocated to subgroups 1–6 using the flowchart form of the decision tree. The rates of sustained virological response were 75% for subgroup 1, 73% for subgroup 2, 65% for subgroup 3, 41% for subgroup 4, 46% for subgroup 5, and 33% for subgroup 6. The rates of sustained virological response for each subgroup of patients were correlated closely between the model building dataset and the validation dataset ($r^2 = 0.94$) (Fig. 2).

J. Med. Virol. DOI 10.1002/jmv

The six subgroups were reconstructed into three groups according to their rate of sustained virological response: the high-probability group consisted of subgroups 1 and 2, the intermediate-probability group consisted of subgroup 3, and the low-probability group consisted of subgroups 4, 5, and 6. The rate of sustained virological response in the high-probability group was high on a consistent basis: 83% for model-building patients and 74% for validation patients. The rate of sustained virological response in the intermediate-probability group was 64% for model building patients and 65% for internal validation patients. The rate of sustained virological response in the low-probability group was low on a consistent basis: 32% for model-building patients and 36% for internal validation patients (Fig. 3). Thirty percent of the patients were classified into the high-probability group and 10% of the patients were classified into intermediate-probability group, which means that about 40% of patients with higher than average probability of achieving sustained virological response were identified.

Effect of Dose Reductions of PEG-IFN and RBV

The possible effect of drug reductions was analyzed in the three groups of patients divided by decision tree (low-, intermediate-, and high-probability groups)

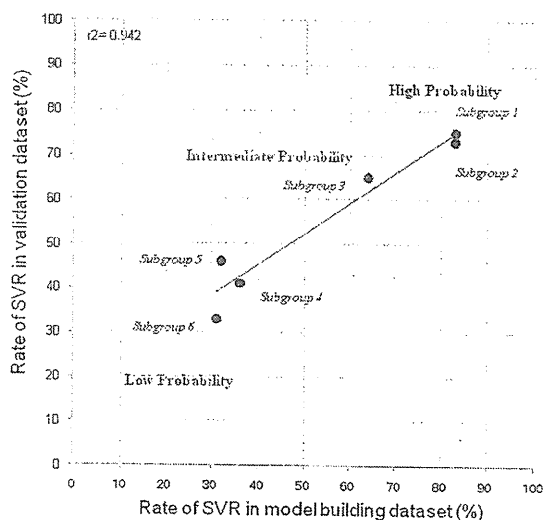


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were then calculated and plotted for each subgroup. The x-axis represents the rate of sustained virological response in the model-building datasets and the y-axis represents the rate of sustained virological response in the validation datasets. The rates of achieving sustained virological response in each subgroup of patients correlated closely between the model-building dataset and the validation dataset (correlation coefficient: $r^2 = 0.94$).

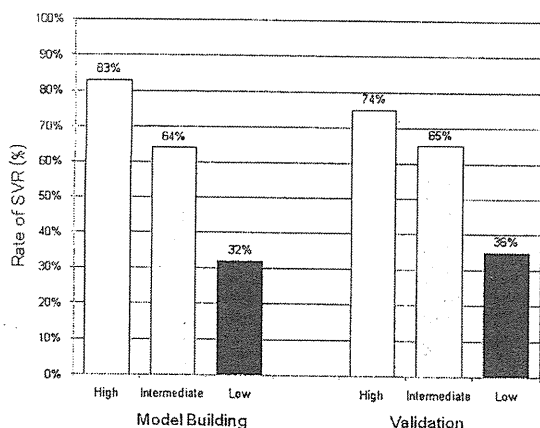


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup 4, 5, and 6), intermediate-probability group (subgroup 3), and high-probability group (subgroup 1 and 2), respectively. The rate of sustained virological response showed significant difference between the three groups.

(Fig. 4). Patients were stratified according to the cumulative drug exposure with PEG-IFN and RBV: the good adherence group consisted of patients who took $\geq 80\%$ planned doses of both PEG-IFN and RBV; the poor adherence group consisted of patients who took $< 80\%$ of planned doses of both PEG-IFN and RBV. Even after adjustment for drug adherence, the three groups of patients divided by decision-tree analysis still had low, intermediate, and high probability of achieving sustained virological response, respectively, indicating that this model predicts sustained virological response independent of drug exposure.

Multivariable Logistic Regression Analysis

Age, sex, serum levels of creatinine, ALT, GGT, LDL-C, hemoglobin, platelet count, HCV RNA titer, ISDR substitution, substitution at Core70, substitution at Core91, histological stage of fibrosis, and grade of activity were found to be associated with sustained virological response by standard univariable analysis. Multivariable analysis including these factors showed that age, sex, LDL-C levels, GGT levels, platelet count, ISDR substitution, and substitution at Core70 showed independent associations with sustained virological response (Table II). Substitution in ISDR had the highest odds ratio, at 9.92. Fibrosis, which was selected as a significant predictor of response in the decision-tree analysis, was not found to be an independent predictor of response in standard multivariable analysis, indicating that the decision-tree analysis could identify significant predictors that would apply specifically to selected patients.

DISCUSSION

The present study revealed that viral factors such as substitutions in ISDR and Core70 are significant and independent predictors of sustained virological response to PEG-IFN plus RBV in chronic hepatitis C. In a decision-tree model for the pretreatment prediction of sustained virological response, the number of substitutions in ISDR was the best predictor of sustained virological response, followed by younger age, wild-type sequence at Core70, higher level of LDL-C, and absent fibrosis. This decision-tree model could identify patients with high probability of sustained virological response (83%) among difficult-to-treat genotype 1b chronic hepatitis C patients. Using this model, rapid estimates of the response before treatment can be made by allocating patients to specific subgroups with a defined rate of response simply by following the flowchart form. Because more potent therapy, such as a combination of protease inhibitor, PEG-IFN, and RBV, is under clinical trial and may become available in the near future [Hezode et al., 2009; McHutchison et al., 2009], pretreatment prediction of the likelihood of sustained virological response may be useful for both patients and physicians to support clinical decisions whether to start current standard therapy or to wait for emerging new therapies.

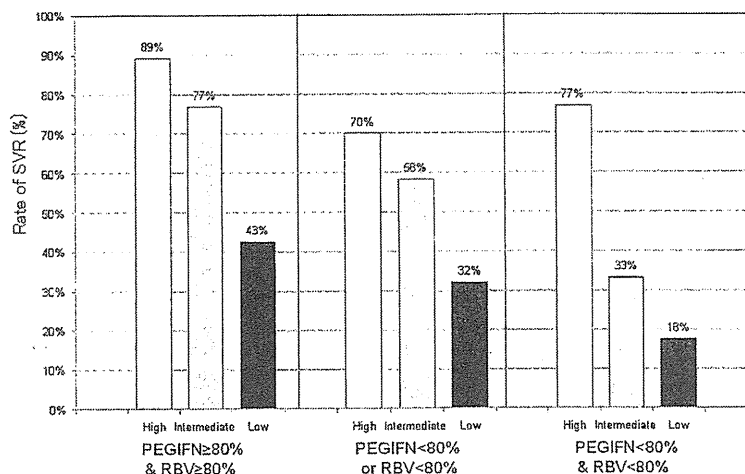


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further stratified according to cumulative drug exposure to PEG-IFN and RBV.

Two or more substitutions in ISDR had a strong impact on sustained virological response, because this factor was selected as a top variable in decision-tree analysis and had the highest odds ratio in multivariable analysis. Moreover, even among patients with unfavorable ISDR (0 or 1 mutation), younger patients (<60 years) with the wild-type sequence at Core70 and high level of LDL-C (≥ 120 mg/dl) had a high rate of sustained virological response. The sustained virological response rate of these two subgroups of patients was 83% in the model-building patients and 75% in the validation patients. Thus, patients with high possibility of sustained virological response could be extracted by the combined analysis of ISDR and Core70. These patients may be the best-suited candidates for treatment with the current combination therapy. Conversely, the following patients with 0–1 mutation in ISDR had a low probability of sustained virological response (32–35%): (1) older (>60 years); or (2) younger (<60 years) patients but having mutant-type sequence at Core70; or (3) younger (<60 years) patients having a wild-type sequence at Core70, but having a low level of LDL-C (<120 mg/dl) and advanced fibrosis. These patients may

be advised to wait for a more effective therapy. Decision may be made on a case-by-case basis, taking into account the potential risk of disease progression while waiting.

In a previous decision-tree model using simple and noninvasive standard tests that are available readily worldwide [Kurosaki et al., 2010b], the rate of sustained virological response was at most 65–76% among those in the high-probability group. That model focused on use by general physicians in routine general practice, especially where specialized resources, such as liver biopsy or determination of viral sequences, are not available. In that model, younger age, male sex, higher platelet counts, lower alpha-fetoprotein (AFP) levels, and lower GGT levels were identified as favorable predictive parameters. Higher AFP levels and lower platelet counts that are hallmarks of advanced fibrosis [Shiratori and Omata, 2000; Akuta et al., 2007b] were associated with low probability of sustained virological response in that model. On the other hand, the present analysis aimed to clarify the significance of viral factors for pretreatment prediction of sustained virological response, and to build an advanced model that may be used by specialist physicians engaged in the

TABLE II. Multivariable Logistic Regression Analysis for Factors Associated With SVR

Parameter		Odds	95% CI	P-value
Age (years)	<60 vs. ≥ 60	2.28	1.31–3.94	0.003
Sex	Male vs. female	3.36	1.87–5.99	<0.0001
GGT (IU/L)	<40 vs. ≥ 40	2.65	1.45–4.85	0.002
LDL-C (mg/dl)	≥ 120 vs. <120	1.79	0.91–3.53	0.094
Platelets (10 ⁹ /L)	≥ 120 vs. <120	2.69	1.22–5.90	0.014
ISDR mutations	≥ 2 vs. 0–1	9.92	3.71–26.54	<0.0001
Core70	Wild vs. mutant	1.92	1.07–3.47	0.030

GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region.

treatment of hepatitis. In the present model, stage of fibrosis was selected as a predictive factor, but at lower level of significance than HCV mutations. The predicted rate of sustained virological response in the high-probability group of the present model is higher than that in the previous model (75–83% vs. 65–76%). These results indicate that substitutions in ISDR and Core70 were important pretreatment predictors of sustained virological response. Determination of these viral factors is not available readily in clinical practice, but is of value for improving the accuracy of pretreatment prediction of sustained virological response.

Substitutions in ISDR and Core70 have been reported previously to be associated with efficacy of IFN therapy. The association between the number of substitutions in ISDR and response to therapy was demonstrated originally in patients treated with IFN mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997], but recent studies have reported a positive correlation with PEG-IFN and RBV combination therapy as well [Munoz de Rueda et al., 2008; Shirakawa et al., 2008; Ikeda et al., 2009]. Another important viral factor relevant to treatment response is amino acid substitution in Core70. The sequence of this amino acid was reported originally to be associated with nonresponse to therapy [Akuta et al., 2005], but subsequent studies confirmed the positive correlation of a wild-type Core70 with sustained virological response [Akuta et al., 2009]. The multiple logistic regression analysis showed that ISDR and Core70 were independent factors associated with sustained virological response along with host factors. How these important viral factors and other host factors can be combined to predict response to PEG-IFN plus RBV is an important clinical question. Decision-tree modeling can make the response probability apparent by combining all these factors. Some factors that may be associated with treatment outcome, such as levels of ferritin or homocysteine, were not included. This may be a potential limitation of the present study.

It is of interest that a recent study by Li et al. [2010] has shown that a high serum level of LDL-C is linked to the *IL28B* major allele (CC in rs12979860). In that study, a high serum level of LDL-C was associated with sustained virological response, but it was no longer significant when analyzed together with the *IL28B* genotype in multivariate analysis. Thus, the association between treatment response and LDL cholesterol levels in the present study may reflect the underlining link of LDL cholesterol levels to the *IL28B* genotype. Recent reports indicate that the *IL28B* genotype and HCV substitutions are correlated closely [Akuta et al., 2010; Kurosaki et al., 2010c]. Still, Core70 [Akuta et al., 2010] or ISDR [Kurosaki et al., 2010c] were predictors of response to therapy independent of *IL28B* genotype. Future study is needed to elucidate the possible mechanisms underlying the association between HCV sequences and host genetic factors, and also the role of host and viral factors for the prediction of treatment response.

In conclusion, a data mining analysis emphasized the impact of substitutions in ISDR and Core70 on pretreatment prediction of sustained virological response to PEG-IFN plus RBV therapy. A decision-tree model that includes substitutions in ISDR and Core70 of HCV could identify patients with high probability of sustained virological response, and could thereby improve the predictive accuracy over predictions that are based on standard tests.

REFERENCES

- Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. 2005. 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* 48:372–380.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007a. 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* 46:403–410.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007b. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. *J Med Virol* 79:1686–1695.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. 2009. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: Amino acid substitutions in HCV core region as predictor of sustained virological response. *J Med Virol* 81:452–458.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Chayama K, Nakamura Y, Kumada H. 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. *Hepatology* 52:421–429.
- Averbook BJ, Fu P, Rao JS, Mansour EG. 2002. 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* 132:589–602.
- Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttill RW. 2003. Phosphorus as an early predictive factor in patients with acute liver failure. *Transplantation* 75:2007–2014.
- Bedossa P, Poynard T. 1996. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 24:289–293.
- Breiman L, Friedman RA, Olshen CJ, Stone CM. 1980. *Classification and regression trees*. CA: Wadsworth.
- Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. 2003. Early virologic response to treatment with peginterferon alpha-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 38:645–652.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. 1995. 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* 96:224–230.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334:77–81.
- Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves FL, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. 2002. Peginterferon alpha-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 347:975–982.

- Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, Klein T, Mori M. 2005. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 23:4322-4329.
- Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461:399-401.
- Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, Bronowicki JP, Bourliere M, Gharakhanian S, Bengtsson L, McNair L, George S, Kieffer T, Kwong A, Kauffman RS, Alam J, Pawlowsky JM, Zeuzem S. 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 360:1839-1850.
- Ikeda H, Suzuki M, Okuse C, Yamada N, Okamoto M, Kobayashi M, Nagase Y, Takahashi H, Matsunaga K, Matsumoto N, Itoh F, Yotsuyanagi H, Koitabashi Y, Yasuda K, Iino S. 2009. Short-term prolongation of pegylated interferon and ribavirin therapy for genotype 1b chronic hepatitis C patients with early viral response. *Hepatology* 49:753-759.
- Izumi N, Nishiguchi S, Hino K, Suzuki F, Kumada Y, Itoh Y, Asahina Y, Tamori A, Hiramatsu N, Hayashi N, Kudo M. 2010. Management of hepatitis C; Report of the consensus meeting at the 45th annual meeting of the Japan society of hepatology (2009). *Hepatology* 50:347-368.
- Jensen DM, Morgan TR, Marcellin P, Pockros PJ, Reddy KR, Hadziyannis SJ, Ferenci P, Ackrill AM, Willems B. 2006. Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. *Hepatology* 43:954-960.
- Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, Ikeda T, Tozuka S, Izumi N, Marumo F, Sato C, Ogura Y. 1997. 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* 25:750-753.
- Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, Tamaki N, Hosokawa T, Ueda K, Tsuchiya K, Nakanishi H, Ikeda H, Itakura J, Takahashi Y, Asahina Y, Higak M, Enomoto N, Izumi N. 2010a. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatology* 50:251-260.
- Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, Sugauchi F, Yatsuhashi H, Izumi N. 2010b. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. *J Gastroenterol* DOI: 10.1007/s00535-010-0322-5.
- Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, Sugiyama M, Matsuura K, Sugauchi F, Asahina Y, Nakagawa M, Watanabe M, Sakamoto M, Maekawa S, Sakai A, Kaneko S, Ito K, Masaki N, Tokunaga K, Izumi N, Mizokami M. 2010c. Pretreatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. *J Hepatol* DOI: 10.1016/j.jhep.2010.07.037.
- LeBlanc M, Crowley J. 1995. A review of tree-based prognostic models. *Cancer Treat Res* 75:113-124.
- Lee SS, Ferenci P. 2008. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther* 13:9-16.
- Leiter U, Buettner PG, Eigentler TK, Garbe C. 2004. Prognostic factors of thin cutaneous melanoma: An analysis of the central malignant melanoma registry of the German Dermatological Society. *J Clin Oncol* 22:3660-3667.
- Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, Suchindran S, Muir AJ, Guyton JR, Gardner SD, McHutchison JG, McCarthy JJ. 2010. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. *Hepatology* 51:1904-1911.
- Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomized trial. *Lancet* 358:958-965.
- McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, McNair L, Alam J, Muir AJ. 2009. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* 360:1827-1838.
- Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. 2002. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 12:243-248.
- Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, Gila A, Quiles R, Leon J, Ruiz-Extremera A, Salmeron J. 2008. Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82:6644-6653.
- Segal MR, Bloch DA. 1989. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 8:539-550.
- Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, Ichijo T, Yoshizawa K, Kiyosawa K, Tanaka E. 2008. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology* 48:1753-1760.
- Shiratori Y, Omata M. 2000. 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* 15:E141-E151.
- Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, Bassendine M, Spengler U, Dore GJ, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Muller T, Bahlo M, Stewart GJ, Booth DR, George J. 2009. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 41:1100-1104.
- Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. 2009. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 41:1105-1109.
- Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, Hatakeyama K. 2007. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol* 14:34-40.
- Yu ML, Dai CY, Huang JF, Chiu CF, Yang YH, Hou NJ, Lee LP, Hsieh MY, Lin ZY, Chen SC, Wang LY, Chang WY, Chuang WL. 2008. Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: A randomized trial. *Hepatology* 47:1884-1893.
- Zlobec I, Steele R, Nigam N, Compton CC. 2005. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 11:5440-5443.

Review Article

Predictors of Virological Response to a Combination Therapy with Pegylated Interferon Plus Ribavirin Including Virus and Host Factors

Namiki Izumi, Yasuhiro Asahina, and Masayuki Kurosaki

Department of Gastroenterology and Hepatology, Musashino Red-Cross Hospital, Kyonancho 1-26-1, Musashinoshi, 180-8610 Tokyo, Japan

Correspondence should be addressed to Namiki Izumi, nizumi@musashino.jrc.or.jp

Received 25 April 2010; Revised 29 June 2010; Accepted 19 July 2010

Academic Editor: Ming-Lung Yu

Copyright © 2010 Namiki Izumi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A combination therapy with pegylated interferon (PEG-IFN) plus ribavirin (RBV) has made it possible to achieve a sustained virological response (SVR) of 50% in refractory cases with genotype 1b and high levels of plasma HCV RNA. Several factors including virus mutation and host factors such as age, gender, fibrosis of the liver, lipid metabolism, innate immunity, and single nucleotide polymorphism (SNPs) are reported to be correlated to therapeutic effects. However, it is difficult to determine which factor is the most important predictor for an individual patient. Data mining analysis is useful for combining all these together to predict the therapeutic effects. It is important to analyze blood tests and to predict therapeutic effects prior to initiating treatment. Since new anti-HCV agents are under development, it will be necessary in the future to select the patients who have a high possibility of achieving SVR if treatment is performed with standard regimen.

1. Progress in Virological Response in the Difficult-to-Treat Patients with Genotype 1 Hepatitis C Virus (HCV) Infection and Factors Correlated to the Efficacy

Recently, the average age of the patients with chronic hepatitis C has been increasing in Japan. Incidence of hepatocellular carcinoma (HCC) in the elderly patients with chronic hepatitis C (65 years or older) has demonstrated to be higher than younger ones when adjusted by the stage of hepatic fibrosis [1]. In Japan, refractory cases with genotype 1b and high HCV RNA levels are seen in as high as 70 percent of chronic hepatitis C patients. The outcome of conventional IFN monotherapy has been an SVR response of 3%–5% after 6 months of treatment in genotype 1b and high HCV RNA patients [2, 3], and virus mutation such as interferon sensitivity-determining region (ISDR) is shown to be correlated with the virological response [2]. The association of ISDR mutations and virological response to IFN monotherapy was denied in an Italian study [4];

however, it was confirmed by a Chinese study [5] and an international meta-analysis [6].

However, pegylated IFN (PEG-IFN) extends the duration of therapy and reduces adverse effects, and for this reason, PEG-IFN has become the cornerstone of therapy. Furthermore, by the combination therapy with PEG-IFN and ribavirin (RBV), the rate of SVR has dramatically improved. Even in the patients with genotype 1b and high HCV RNA level, SVR rate reaches as high as 40%–50%, thereby improving the therapeutic effects both in Western countries [7, 8] and in Japan [9, 10].

It is important to predict the rate of achieving SVR in the individual patient, before initiating treatment. Both virus- and host-related elements have been reported as factors correlated to therapeutic effects of combination therapy [11–13]. A particular focus has been placed on virus mutations, age, gender, fibrosis of the liver, lipid metabolism, and degree of fatty metamorphosis of the liver.

Among these factors related to PEG-IFN and RBV, innate immunity has been shown to be correlated in virological

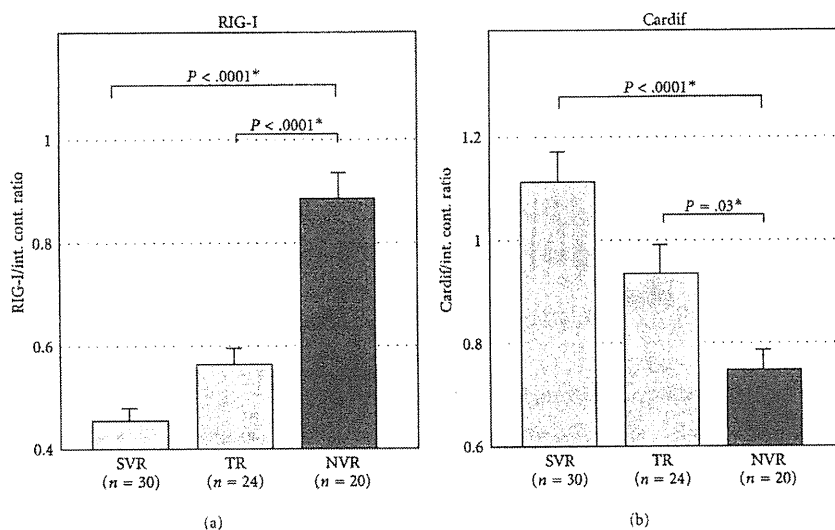


FIGURE 1: Expression of genes correlated to the intrahepatic innate immunity and virological response to PEG-IFN alpha-2b and RBV combination therapy. Open column indicates SVR ($n = 30$), gray column indicates TR ($n = 24$), and closed column indicates NVR ($n = 20$). Error bars indicate the standard error. The P values were analyzed by the Kruskal-Wallis test. Expression of RIG-I was significantly higher in NVR than in SVR patients, and Cardif expression was higher in SVR than in NVR. The figure was cited from [8].

response. Asahina et al. reported that liver biopsies were performed before the PEG-IFN and RBV combination therapy to examine the correlation between the gene expression involved in innate immunity and the therapeutic effects, and in the patients in whom RIG-I expression is high and the expression of Cardif, an adaptor gene, is low, it was found that there are many nonresponders (NVRs) in which HCVRNA does not become negative during the course of treatment [13]. It was therefore revealed that there are many NVRs among the patients in whom the ratio of RIG-I to Cardif in liver tissue is high and that this ratio is low in the SVR patients. Based on these findings, it has become clear that innate immunity is correlated to therapeutic effects (Figure 1).

Furthermore, it was recently discovered that a single nucleotide polymorphism (SNP) of the host gene IL28B is significantly involved in the therapeutic response to the PEG-IFN and RBV combination therapy [14, 15]. The possibility of becoming an NVR is high in cases of the minor allele carriers of IL28B. However, it is not possible to routinely measure an SNP of IL28B in the clinical setting. In this paper, factors which can actually be used in real clinical practice are discussed for the prediction of the efficacy of PEG-IFN and RBV combination therapy.

2. Amount of HCVRNA

In the patients with chronic hepatitis C, it is not possible to directly measure the amount of virus, and the

amount of HCVRNA is measured instead. Currently, a real-time PCR method which has an advantage of wide range and high sensitivity is utilized, and measurements can be taken from a single blood sample of a very small amount, that is, 1.2 log copies/ml, to a very large amount, that is, 8 log copies/ml. This method has a higher level of sensitivity than the conventional AmpliCor monitor test and can therefore detect HCVRNA even if only a very small amount exists in the plasma. If the amount of HCVRNA in plasma is less than the range of sensitivity of the real-time PCR method, it is recorded as undetectable level. If the indication is "less than 1.2 log copies/ml of HCVRNA", it means that a very small amount of HCVRNA exists in the plasma. Since the indication of the real-time PCR method is based on log counts, a decrease of 1.0 in the numerical value means that the amount of HCVRNA has decreased to 1/10. With the application of this real-time PCR method, it has become possible to measure amounts of HCVRNA up to 8 log copies/ml, and it has also become possible to predict the efficacy before treatment and to monitor appropriately the reactivity during the course of treatment. However, in the patients in whom a PEG-IFN and RBV combination therapy is performed, SVR can be acquired even when the amount of virus prior to the treatment is quite large. It is therefore difficult to predict the virological response solely from the amount of HCVRNA before starting the treatment. Once treatment has commenced, at what week HCVRNA becomes negative is important for the prediction of therapeutic effects, and this serves as a parameter for deciding the duration of treatment [16].

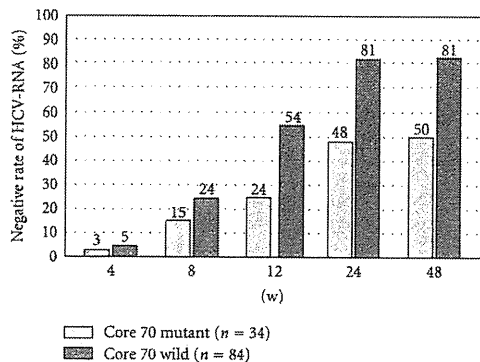


FIGURE 2: Comparison of aa70 mutations in the HCV core region and the rate of HCV-RNA becoming negative during the course of treatment. Compared with the wild type, among the patients of aa70 mutations, there were fewer patients in whom HCV-RNA had become negative during the course of treatment.

Measuring the rate of viral clearance from serum is helpful for predicting the likelihood of a response to PEGIFN and RBV and useful for determining the optimal duration of therapy if the patients start to receive the treatment [17]. In the AASLD practice guideline, response-guided therapy is highly recommended [18]. In two nationwide registration trials conducted in Japan [9, 10], the SVR rate was high, from 76% to 100%, in patients whose HCV-RNA was cleared rapidly from serum by week 4 (rapid virological response; RVR), and 71% to 73% in patients who achieved undetectable HCV-RNA from week 5 to week 12 (early virological response; EVR). In contrast, the SVR rate in patients with late clearance of HCV-RNA from week 13 to week 24 was low at 29% to 36%. No patients without clearance of HCV-RNA by week 24 achieved SVR.

The strategy of extending therapy in patients with delayed virological responses, defined as clearance of HCV-RNA between weeks 12 and 24, was evaluated in five studies [19–23]. These results cannot be compared directly with each other because of the heterogeneous study populations, differences in the baseline characteristics, and the different regimens utilized amongst them. Nevertheless, the results showed a trend toward a higher SVR rate by extending therapy from 48 to 72 weeks in patients with delayed virological response.

3. Viral Mutations in Core and NS5A Region

In the patients with genotype 1b HCV infection, the mutations in aa70 and aa91 in the core region have been shown to correlate with early virological response (EVR) during PEG-IFN and RBV combination treatment [11]. If aa70 in the core region is mutated to anything other than arginine and aa91 to anything other than leucine, it is difficult to achieve EVR, and it is thus highly possible that such cases

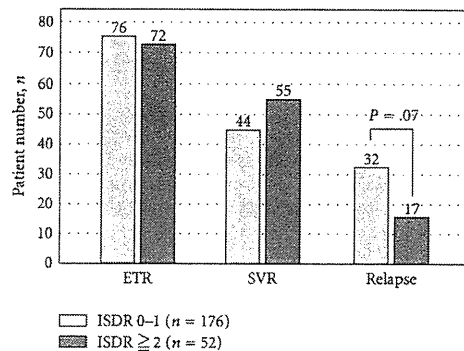


FIGURE 3: Number of ISDR substitutions and the comparison of virological response, SVR, and relapse at the end of the treatment. Compared with the patients with 2 or more sites of substitutions, the rate of SVR was lower and the rate of relapse was higher in the patients in whom there were fewer substitutions in ISDR, that is, 0 or 1 sites.

will become nonresponders. The examination results at our institution including 292 patients with genotype 1b infection demonstrated that, in the cases with mutations in aa70 in the core region, the rate of HCV-RNA becoming negative during the course of combination treatment was low compared to the wild type of aa70 (Figure 2). However, core aa70 mutation is shown to have quasispecies detectable by cloning, and 70Q clone was positively selected during combination treatment with PEGIFN and RBV [24].

Furthermore, Enomoto et al. reported that the patients with 4 or more amino acid mutations were observed in interferon sensitivity-determining region (ISDR) within NS5A region [2]; SVR rate is higher than 90% by IFN monotherapy, and SVR is less than 10% in the patients with no mutation in ISDR. It has also been reported that, in PEG-IFN and RBV combination therapy, the number of ISDR mutations is involved in the SVR [12].

We analyzed the relationship between virological response and ISDR mutations in the patients with genotype 1b infection treated by PEG-IFN alpha-2b and RBV combination therapy. In the patients with 0 or 1 ISDR mutation, even if the rate of HCV-RNA becoming negative at the end of treatment was the same, the rate of SVR would be lower compared with the patients having 2 or more mutations (Figure 3). This demonstrates that there is a higher incidence of relapse after the end of treatment in the patients with 0 or 1 ISDR mutation.

Enomoto and Maekawa reported that mutations both in NS5A-ISDR (interferon sensitivity-determining region) and core 70Q substitution are associated with no early viral response during PEGIFN and RBV combination therapy [25]. Association of core aa70 substitution and mutations in NS5A region is confirmed to be associated with viral response by PEGIFN and RBV combination therapy in a Japanese multicenter cooperative study [26]. The number of

mutations in the interferon sensitivity-determining region was shown to be associated with the viral response to PEGIFN and RBV combination treatment not only in Japan [27], but also in Tunisia [28].

Recently, a consensus has been established that mutations in aa70 in the core region are important for the prediction of HCVRNA becoming negative during the early course of treatment, and the number of ISDR mutations is important for the prediction of relapse after the end of treatment.

4. Adherence

It has been confirmed that it is important to ensure 80% or more of the scheduled dose of both PEG-IFN and RBV in order to improve the rate of SVR, and together with the duration of treatment, the 80 · 80 · 80 rule has been established. However, Schiffman et al. recently reported that the dose of PEG-IFN in the initial stage of administration is important and that, if a sufficient dose of PEG-IFN is administered, then 60% or more of the RBV dose would be enough [29]. It is therefore of primary importance to ensure the dose of PEG-IFN.

In Japan, the average age of patients with chronic hepatitis C is increasing, and achieving good adherence is difficult in many patients. Consequently, the rate of SVR is low in elderly patients. How to improve the rate of SVR in elderly patients is an important issue. With regard to the dose of RBV, reducing the RBV dose based on the calculation of the total body clearance (CL/F) has been proposed to be useful for decreasing the discontinuation and improving the rate of SVR. Although there is no consensus on an appropriate dose of PEG-IFN in elderly patients, if the initial dose is set lower than the usual dose, discontinuation would be reduced. Thus, it is necessary to investigate whether such an initial dose would improve the rate of SVR.

Recently, the risk of hemolytic anemia was clearly demonstrated to correlate with ITAP gene SNP [30]. The predictive implication should be analyzed prospectively in clinical practice.

5. Host Factors

Zeuzem et al. described the factors related to the less response to interferon-based therapy, and he showed that several host factors such as older age, race, and obesity are responsible factors for the poor response to IFN [31]. Recently, insulin resistance which was examined by homeostasis model assessment index (HOMA-IR) was shown to be associated with a lower rate of SVR, and body mass index (BMI) was not identified as a significant factor for the poor response to PEGIFN and ribavirin combination therapy [32]. Insulin resistance was confirmed as a related factor to the nonresponse to interferon-based treatment [33]. However, Charlton et al. reported that obesity itself is an associated factor for decreased efficacy of interferon-based therapies, and they discussed the possible mechanism [34], and obesity was shown to be associated with the increased enhancement

of suppressor of cytokine signaling (SOCS) family in the hepatocytes [35].

6. Data Mining Analysis

Both virus- and host-related factors are correlated to therapeutic effects of PEG-IFN and RBV. One important question is which of these factors should be focused on in order to predict the therapeutic effects in an individual patient. In addition, in each individual patient, the host and virological factors are different. It is therefore difficult to predict the virological response in each case before treatment. Furthermore, although it is important to predict the relapse rate when HCVRNA becomes within an undetectable level in an individual patient, prediction of the rate of SVR including virological and host factors and adherence to the treatment has never been carried out in an individual patient.

A data mining analysis is the process of analyzing a large amount of data by a computer in order to develop an algorithm. Conventional statistics have been used to examine certain hypothesis. Data mining is superior in that it can set an algorithm, using a computer, based on a large amount of data without a hypothesis.

We therefore conducted at our institute a data mining analysis of the patients with genotype 1b infection having high levels of HCVRNA to whom a PEG-IFN alpha-2b and RBV combination therapy was administered to investigate whether by the 12th week after the commencement of treatment HCVRNA became negative (EVR) (Figure 4) [36]. The most important factor for the prediction of EVR was the steatosis of the liver: when steatosis was observed in 30% or more of hepatocytes, EVR was found to be difficult to achieve. In the patients in whom steatosis was not severe, the second most important factor was the serum LDL cholesterol value. While the rate of EVR was 57% in the patients in whom this value was 100 mg/ml or above, the rate of EVR was 32% in the patients in whom the LDL cholesterol was less than 100 mg/dl.

The higher the LDL cholesterol value, the earlier the HCVRNA became negative. Among the patients with low LDL cholesterol values, while the rate of EVR was 15% in patients 60 years of age or older, the rate was high in the patients under the age of 60 years old, that is, 49%. Among patients under the age of 60, the rate of EVR was low, that is, 31%, in patients with a blood glucose level of 120 mg/dl or above whereas EVR was achieved in 71% of the patients with a blood glucose level of less than 120 mg/dl (Figure 4).

On the other hand, in the patients with high LDL cholesterol values, the next most important factor was age. While the rate of EVR was 50% in patients 50 years of age or older, EVR was achieved in 77% of the patients under the age of 50. Among patients of 50 years of age or older, the next most important factor was the gamma GTP value. While the rate of EVR was 35% in the patients in whom gamma GTP was 40 IU/L or above, EVR was achieved in 60% of the patients where the value was less than 40 IU/L.

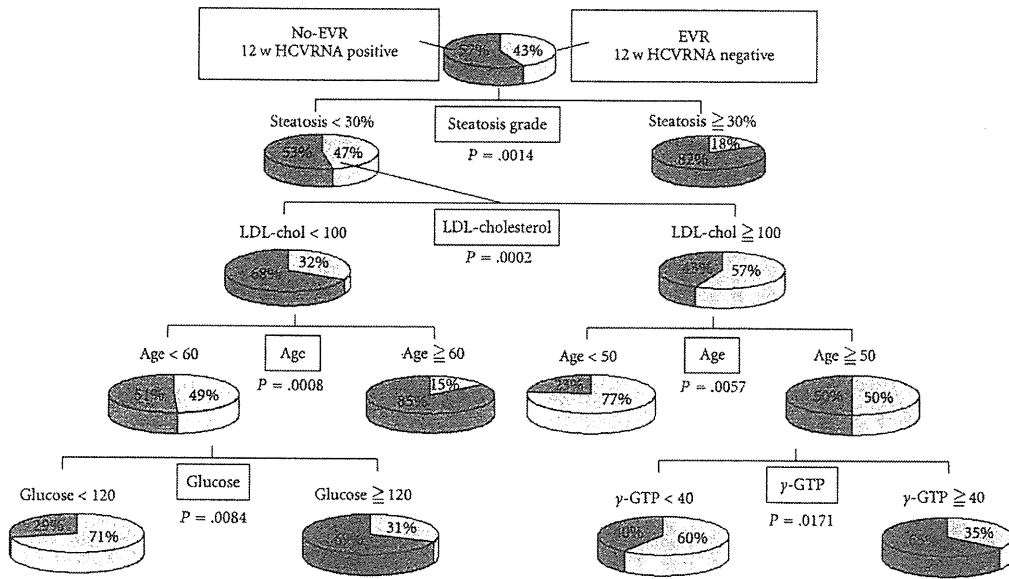


FIGURE 4: HCV RNA negative (EVR) algorithm at 12th week from data mining analysis of PEG-IFN alpha-2b plus RBV combination in the genotype 1b and high levels of HCV RNA. Both virological and host factors were evaluated by data mining analysis software from SPSS. This figure was cited from [36].

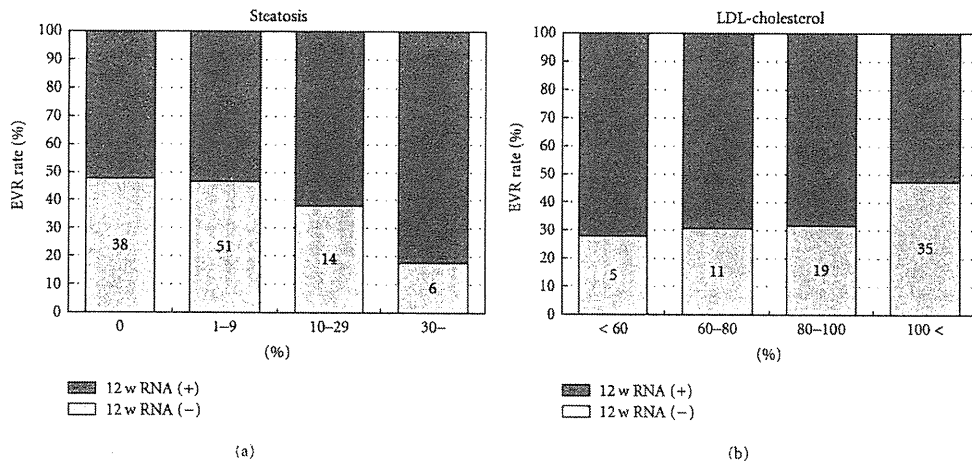


FIGURE 5: The rate of EVR in the patients with genotype 1b and high levels of HCV RNA, based on fatty deposition in the liver (a), and the LDL cholesterol value (b), respectively. EVR was highly achieved in the patients with less steatosis in the liver, and in those with high serum LDL-cholesterol levels. This is univariate analysis, and cited from [36].

We therefore compared these factors based on the original data. A univariate comparison of the fatty infiltration of the liver and the rate of EVR demonstrated that the rate of EVR was higher when the steatosis of the liver was less severe (Figure 5(a)). In addition, a comparison of the LDL cholesterol value and the rate of EVR demonstrated a significant correlation, confirming that the higher the LDL cholesterol value, the higher the rate of EVR (Figure 5(b)). Therefore, it was also proposed by the results of univariate analysis of each factor extracted from the data mining analysis that these factors were correlated to the rate of EVR.

From these observations, it is likely to improve the viral response to PEGIFN and ribavirin by reducing steatosis of the liver through daily walking or abstaining alcohol intake or by refraining from high-fat diet.

By utilizing data mining, it is therefore possible to assess virus- and host-related factors together and to predict the virological response in each patient, and thereby clinically useful information can be obtained. The algorithm should be validated including a large number of the patients.

Acknowledgment

This paper was supported by a grant from the Japanese Ministry of Health, Welfare and Labor.

References

- [1] Y. Asahina, K. Tsuchiya, I. Hirayama, et al., "Effect of aging on risk for hepatocellular carcinoma in chronic hepatitis C virus infection," *Hepatology*, vol. 52, no. 2, pp. 518–527, 2010.
- [2] N. Enomoto, I. Sakuma, Y. Asahina et al., "Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection," *New England Journal of Medicine*, vol. 334, no. 2, pp. 77–81, 1996.
- [3] S. Iino, F. Ichida, A. Sakuma et al., "A randomized clinical trial with natural interferon- α monotherapy for 24 or 48 weeks on patients with chronic hepatitis C having genotype 1b infection in high viral titers," *Hepatology Research*, vol. 24, no. 4, pp. 338–345, 2002.
- [4] G. Squadrito, M. E. Orlando, I. Cacciola et al., "Long-term response to interferon alpha is unrelated to "interferon sensitivity determining region" variability in patients with chronic hepatitis C virus-1b infection," *Journal of Hepatology*, vol. 30, no. 6, pp. 1023–1027, 1999.
- [5] C. Shen, T. Hu, L. Shen, L. Gao, W. Xie, and J. Zhang, "Mutations in ISDR of NS5A gene influence interferon efficacy in Chinese patients with chronic hepatitis C virus genotype 1b infection," *Journal of Gastroenterology and Hepatology*, vol. 22, no. 11, pp. 1898–1903, 2007.
- [6] M. Pascu, P. Martus, M. Höhne et al., "Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences," *Gut*, vol. 53, no. 9, pp. 1345–1351, 2004.
- [7] S. J. Hadziyannis, H. Sette Jr., T. R. Morgan et al., "Peginterferon- α 2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose," *Annals of Internal Medicine*, vol. 140, no. 5, pp. 346–355, 2004.
- [8] M. P. Manns, J. G. McHutchison, S. C. Gordon et al., "Peginterferon alfa-2b plus ribavirin compared with interferon- α 2b plus ribavirin for initial treatment of chronic hepatitis C: a randomized trial," *Lancet*, vol. 358, no. 9286, pp. 958–965, 2001.
- [9] S. Iino, E. Tomita, H. Kumada et al., "Prediction of treatment outcome with daily high-dose IFN α -2b plus ribavirin in patients with chronic hepatitis C with genotype 1b and high HCV RNA levels: relationship of baseline viral levels and viral dynamics during and after therapy," *Hepatology Research*, vol. 30, no. 2, pp. 63–70, 2004.
- [10] G. Yamada, S. Iino, T. Okuno et al., "Virological response in patients with hepatitis C virus genotype 1b and a high viral load: impact of peginterferon- α -2a plus ribavirin dose reductions and host-related factors," *Clinical Drug Investigation*, vol. 28, no. 1, pp. 9–16, 2008.
- [11] N. Akuta, F. Suzuki, Y. Kawamura 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," *Journal of Hepatology*, vol. 46, no. 3, pp. 403–410, 2007.
- [12] H. Shirakawa, A. Matsumoto, S. Joshita et al., "Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors," *Hepatology*, vol. 48, no. 6, pp. 1753–1760, 2008.
- [13] Y. Asahina, N. Izumi, I. Hirayama et al., "Potential relevance of cytoplasmic viral sensors and related regulators involving innate immunity in antiviral response," *Gastroenterology*, vol. 134, no. 5, pp. 1396–1405, 2008.
- [14] D. Ge, J. Fellay, A. J. Thompson et al., "Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance," *Nature*, vol. 461, no. 7262, pp. 399–401, 2009.
- [15] Y. Tanaka, N. Nishida, M. Sugiyama et al., "Genome-wide association of IL28B with response to pegylated interferon- α and ribavirin therapy for chronic hepatitis C," *Nature Genetics*, vol. 41, no. 10, pp. 1105–1109, 2009.
- [16] P. Ferenci, H. Laferl, T. Scherzer et al., "Peginterferon alfa-2a and ribavirin for 24 weeks in hepatitis C type 1 and 4 patients with rapid virological response," *Gastroenterology*, vol. 135, no. 2, pp. 451–458, 2008.
- [17] N. Izumi, S. Nishiguchi, K. Hino et al., "Management of hepatitis C; Report of the Consensus Meeting at the 45th Annual Meeting of the Japan Society of Hepatology (2009)," *Hepatology Research*, vol. 40, no. 4, pp. 347–368, 2010.
- [18] M. G. Ghany, D. B. Strader, D. L. Thomas, and L. B. Seeff, "Diagnosis, management, and treatment of hepatitis C: an update," *Hepatology*, vol. 49, no. 4, pp. 1335–1374, 2009.
- [19] T. Berg, M. von Wagner, S. Nasser et al., "Extended treatment duration for hepatitis C virus type 1: comparing 48 versus 72 weeks of peginterferon- α -2a plus ribavirin," *Gastroenterology*, vol. 130, no. 4, pp. 1086–1097, 2006.
- [20] J. M. Sánchez-Tapias, M. Diago, P. Escartin et al., "Peginterferon- α 2a plus ribavirin for 48 versus 72 weeks in patients with detectable hepatitis C virus RNA at week 4 of treatment," *Gastroenterology*, vol. 131, no. 2, pp. 451–460, 2006.
- [21] P. Ferenci, H. Laferl, T. M. Scherzer, et al., "Customizing treatment with peginterferon alfa-2a (40kD)(PEGASYS[®]) plus ribavirin (COPEGUS[®]) in patient with HCV genotype 1 or 4 infection: interim results of a prospective randomized trial," *Hepatology*, vol. 44, no. 336a, 2006.
- [22] B. L. Pearlman, C. Ehleben, and S. Saifee, "Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis C

- genotype 1-infected slow responders," *Hepatology*, vol. 46, no. 6, pp. 1688–1694, 2007.
- [23] M. Buti, Y. Lurie, N. G. Zakharova, et al., "Extended treatment duration in chronic hepatitis C genotype 1-infected slow responders: final results of the SUCCEED study," *Journal of Hepatology*, vol. 50, supplement 1, p. S58, abstract 141, 2009.
- [24] E. Kurbanov, Y. Tanaka, K. Matsuura et al., "Positive selection of core 70Q variant genotype 1b hepatitis C virus strains induced by pegylated interferon and ribavirin," *Journal of Infectious Diseases*, vol. 201, no. 11, pp. 1663–1671, 2010.
- [25] N. Enomoto and S. Maekawa, "HCV genetic elements determining the early response to peginterferon and ribavirin therapy," *Intervirology*, vol. 53, no. 1, pp. 66–69, 2010.
- [26] T. Okanou, Y. Itoh, H. Hashimoto 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," *Journal of Gastroenterology*, vol. 44, no. 9, pp. 952–963, 2009.
- [27] M. Nakagawa, N. Sakamoto, M. Ueyama et al., "Mutations in the interferon sensitivity determining region and virological response to combination therapy with pegylated-interferon alpha 2b plus ribavirin in patients with chronic hepatitis C-1b infection," *Journal of Gastroenterology*, vol. 45, no. 6, pp. 656–665, 2010.
- [28] N. Bouzgarrou, E. Hassen, W. Mahfoudh et al., "NS5AISDR-V3 region genetic variability of Tunisian HCV-1b strains: correlation with the response to the combined interferon/ribavirin therapy," *Journal of Medical Virology*, vol. 81, no. 12, pp. 2021–2028, 2009.
- [29] M. L. Shiffman, M. G. Ghany, T. R. Morgan et al., "Impact of reducing peginterferon alfa-2a and ribavirin dose during retreatment in patients with chronic hepatitis C," *Gastroenterology*, vol. 132, no. 1, pp. 103–112, 2007.
- [30] J. Fellay, A. J. Thompson, D. Ge et al., "ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C," *Nature*, vol. 464, no. 7287, pp. 405–408, 2010.
- [31] S. Zeuzem, "Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well?" *Annals of Internal Medicine*, vol. 140, no. 5, pp. 370–381, 2004.
- [32] H. S. Conjeevaram, D. E. Kleiner, J. E. Everhart et al., "Race, insulin resistance and hepatic steatosis in chronic hepatitis C," *Hepatology*, vol. 45, no. 1, pp. 80–87, 2007.
- [33] G. Tarantino, P. Conca, P. Sorrentino, and M. Ariello, "Metabolic factors involved in the therapeutic response of patients with hepatitis C virus-related chronic hepatitis," *Journal of Gastroenterology and Hepatology*, vol. 21, no. 8, pp. 1266–1268, 2006.
- [34] M. R. Charlton, P. J. Pockros, and S. A. Harrison, "Impact of obesity on treatment of chronic hepatitis C," *Hepatology*, vol. 43, no. 6, pp. 1177–1186, 2006.
- [35] M. J. Walsh, J. R. Jonsson, M. M. Richardson et al., "Non-response to antiviral therapy is associated with obesity and increased hepatic expression of suppressor of cytokine signaling 3 (SOCS-3) in patients with chronic hepatitis C, viral genotype 1," *Gut*, vol. 55, no. 4, pp. 529–535, 2006.
- [36] M. Kurosaki, K. Matsunaga, I. Hirayama, et al., "A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis," *Hepatology Research*, vol. 40, no. 3, pp. 251–260, 2010.

Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis

Masayuki Kurosaki · Naoya Sakamoto · Manabu Iwasaki · Minoru Sakamoto · Yoshiyuki Suzuki · Naoki Hiramatsu · Fuminaka Sugauchi · Hiroshi Yatsuhashi · Namiki Izumi

Received: 22 June 2010 / Accepted: 21 August 2010 / Published online: 10 September 2010
© Springer 2010

Abstract

Background This study aimed to develop a model for the pre-treatment prediction of sustained virological response (SVR) to peg-interferon plus ribavirin therapy in chronic hepatitis C.

Methods Data from 800 genotype 1b chronic hepatitis C patients with high viral load ($>100,000$ IU/ml) treated by peg-interferon plus ribavirin at 6 hospitals in Japan were randomly assigned to a model building ($n = 506$) or an internal validation ($n = 294$). Data from 524 patients treated at 29 hospitals in Japan were used for an external validation. Factors predictive of SVR were explored using data mining analysis.

Results Age (<50 years), alpha-fetoprotein (AFP) (<8 ng/mL), platelet count ($\geq 120 \times 10^9/l$), gamma-glutamyl-transferase (GGT) (<40 IU/l), and male gender were used to build the decision tree model, which divided patients into 7 subgroups with variable rates of SVR ranging from 22 to 77%. The reproducibility of the model was confirmed by the internal and external validation ($r^2 = 0.92$ and 0.93 , respectively). When reconstructed into 3 groups, the rate of SVR was 75% for the high probability group, 44% for the intermediate probability group and 23% for the low probability group. Poor adherence to drugs lowered the rate of SVR in the low probability group, but not in the high probability group.

M. Kurosaki · N. Izumi (✉)
Division of Gastroenterology and Hepatology,
Musashino Red Cross Hospital, 1-26-1 Kyonan-cho,
Musashino, Tokyo 180-8610, Japan
e-mail: nizumi@musashino.jrc.or.jp

M. Kurosaki
e-mail: kurosaki@musashino.jrc.or.jp

N. Sakamoto
Department of Gastroenterology and Hepatology,
Tokyo Medical and Dental University, Tokyo, Japan
e-mail: nsakamoto.gast@tmd.ac.jp

M. Iwasaki
Department of Computer and Information Science,
Seikei University, Tokyo, Japan
e-mail: iwasaki@st.seikei.ac.jp

M. Sakamoto
First Department of Internal Medicine, University of Yamanashi,
Yamanashi, Japan
e-mail: msakamoto@yamanashi.ac.jp

Y. Suzuki
Department of Hepatology, Toranomon Hospital, Tokyo, Japan
e-mail: suzunari@interlink.or.jp

N. Hiramatsu
Department of Gastroenterology and Hepatology,
Osaka University Graduate School of Medicine,
Osaka, Japan
e-mail: hiramatsu@gh.med.osaka-u.ac.jp

F. Sugauchi
Department of Gastroenterology and Metabolism,
Nagoya City University Graduate School of Medical Sciences,
Nagoya, Japan
e-mail: fsugauch@med.nagoya-cu.ac.jp

H. Yatsuhashi
Clinical Research Center, National Nagasaki Medical Center,
Nagasaki, Japan
e-mail: yatsuhashi@nmc.hosp.go.jp

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, Yamashiro 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 ^a	261/506 (52%)	160/295 (54%)
Body mass index (kg/m ²)	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 ⁹ /l)	164 (52–380)	158 (43–312)
AFP (ng/ml)	14.7 (0.9–680)	13 (0.8–323)
HCV RNA (10 ³ 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

^a 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

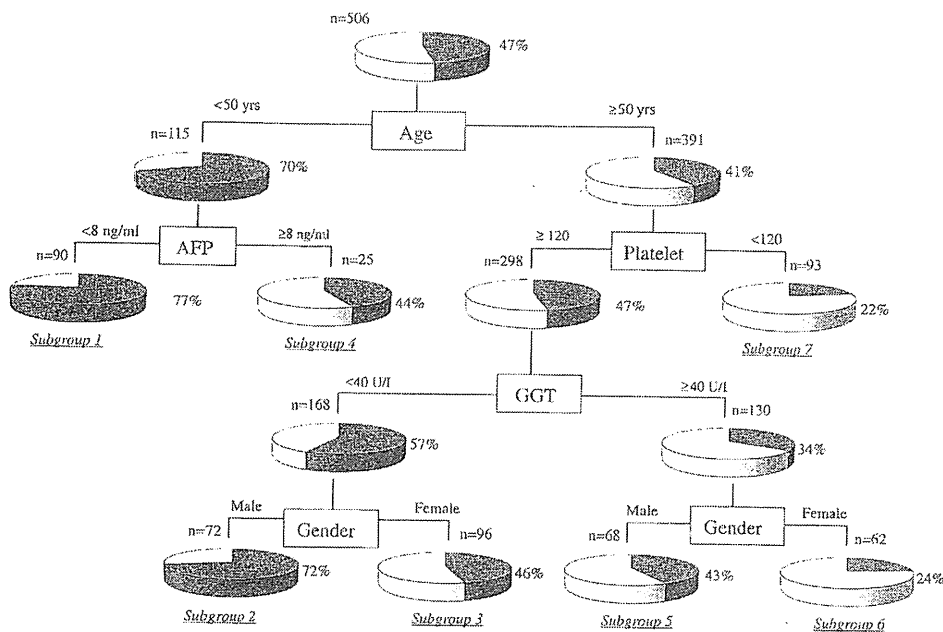


Fig. 1 Decision tree analysis. Boxes indicate the factors for splitting and the cutoff value for the split. Pie charts indicate the rate of SVR for each group. Terminal subgroups of patients discriminated by the

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 ≥ 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 (≥ 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 (≥ 50 years) patients with low platelet counts (< $120 \times 10^9/l$) (subgroup 7) and older (≥ 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