

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 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)

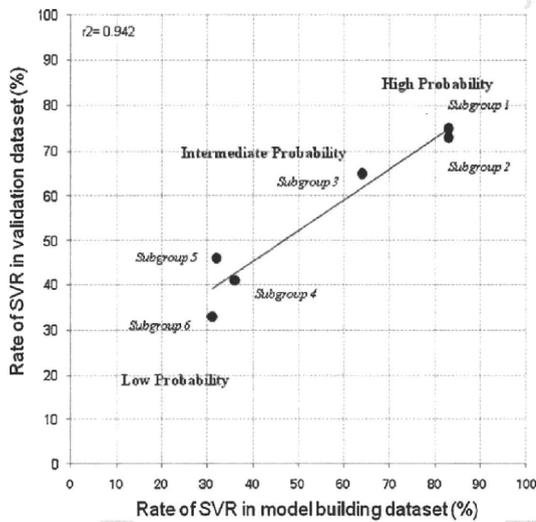


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1-6 by following the flowchart form of the decision tree, and the rates of sustained virological response were 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: $r^2 = 0.94$).

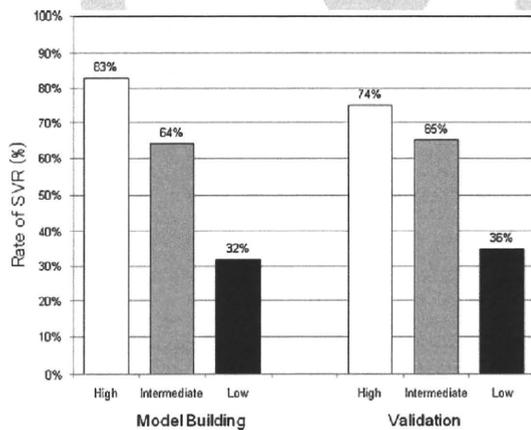


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. White, gray, and black boxes indicate the low-probability group (subgroup 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.

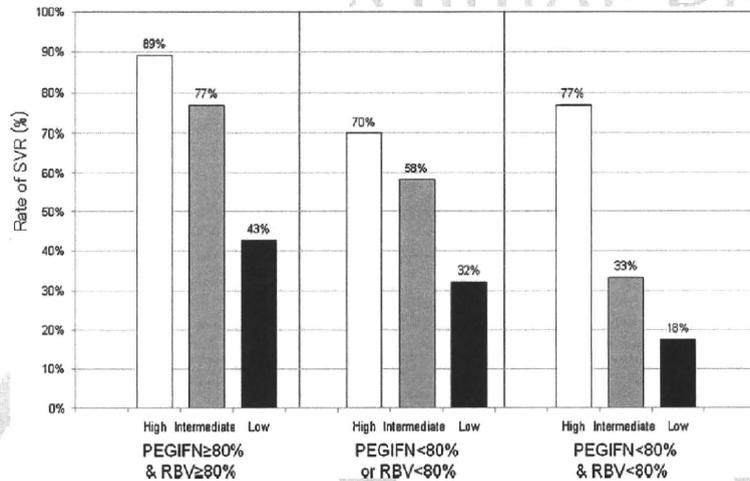


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further 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
Gender	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 alfa-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 alfa-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. *Hepatol Res* 39:753–759.
- 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). *Hepatol Res* 40:347–368.
- 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. *Hepatol Res* 40:251–260.
- 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* (in press⁴⁵).
- 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* (in press⁴⁶).
- 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 randomised 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.
- 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.

Q1: Author: Prod. Editor: Please provide the accepted date.

Q2: Author: Please check the change made.

Q3: Author: Please provide the city name.

Q4: Author: Reference (Segal and Bloch, 1989) is not given in the list. Please provide it in the reference list.

Q5: Author: Please update.

Q6: Author: Please update.



111 RIVER STREET, HOBOKEN, NJ 07030

*****IMMEDIATE RESPONSE REQUIRED*****

Please follow these instructions to avoid delay of publication.

READ PROOFS CAREFULLY

- This will be your only chance to review these proofs.
- Please note that the volume and page numbers shown on the proofs are for position only.

ANSWER ALL QUERIES ON PROOFS (Queries for you to answer are attached as the last page of your proof.)

- Mark all corrections directly on the proofs. Note that excessive author alterations may ultimately result in delay of publication and extra costs may be charged to you.

CHECK FIGURES AND TABLES CAREFULLY (Color figures will be sent under separate cover.)

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

COMPLETE REPRINT ORDER FORM

- Fill out the attached reprint order form. It is important to return the form even if you are not ordering reprints. You may, if you wish, pay for the reprints with a credit card. Reprints will be mailed only after your article appears in print. This is the most opportune time to order reprints. If you wait until after your article comes off press, the reprints will be considerably more expensive.

RETURN

- PROOFS**
- REPRINT ORDER FORM**
- CTA (if you have not already signed one)**

RETURN WITHIN 48 HOURS OF RECEIPT TO Lori-Marie Hamilton at jmvprod@wiley.com; lorimarieh@woh.rr.com

QUESTIONS?

Lori-Marie Hamilton, Production Editor
E-mail: jmvprod@wiley.com; lorimarieh@woh.rr.com
Refer to journal acronym and article production number
(i.e., JMV 00-001 for Journal of Medical Virology ms 00-001).

COPYRIGHT TRANSFER AGREEMENT

Date: _____ Contributor name: _____

Contributor address: _____

Manuscript number (Editorial office only): _____

Re: Manuscript entitled _____

_____ (the "Contribution")

for publication in _____ (the "Journal")

published by _____ ("Wiley-Blackwell").

Dear Contributor(s):

Thank you for submitting your Contribution for publication. In order to expedite the editing and publishing process and enable Wiley-Blackwell to disseminate your Contribution to the fullest extent, we need to have this Copyright Transfer Agreement signed and returned as directed in the Journal's instructions for authors as soon as possible. If the Contribution is not accepted for publication, or if the Contribution is subsequently rejected, this Agreement shall be null and void. **Publication cannot proceed without a signed copy of this Agreement.**

A. COPYRIGHT

1. The Contributor assigns to Wiley-Blackwell, during the full term of copyright and any extensions or renewals, all copyright in and to the Contribution, and all rights therein, including but not limited to the right to publish, republish, transmit, sell, distribute and otherwise use the Contribution in whole or in part in electronic and print editions of the Journal and in derivative works throughout the world, in all languages and in all media of expression now known or later developed, and to license or permit others to do so.

2. Reproduction, posting, transmission or other distribution or use of the final Contribution in whole or in part in any medium by the Contributor as permitted by this Agreement requires a citation to the Journal and an appropriate credit to Wiley-Blackwell as Publisher, and/or the Society if applicable, suitable in form and content as follows: (Title of Article, Author, Journal Title and Volume/Issue, Copyright © [year], copyright owner as specified in the Journal). Links to the final article on Wiley-Blackwell's website are encouraged where appropriate.

B. RETAINED RIGHTS

Notwithstanding the above, the Contributor or, if applicable, the Contributor's Employer, retains all proprietary rights other than copyright, such as patent rights, in any process, procedure or article of manufacture described in the Contribution.

C. PERMITTED USES BY CONTRIBUTOR

1. **Submitted Version.** Wiley-Blackwell licenses back the following rights to the Contributor in the version of the Contribution as originally submitted for publication:

a. After publication of the final article, the right to self-archive on the Contributor's personal website or in the Contributor's institution's/employer's institutional repository or archive. This right extends to both intranets and the Internet. The Contributor may not update the submission version or replace it with the published Contribution. The version posted must contain a legend as follows: This is the pre-peer reviewed version of the following article: FULL CITE, which has been published in final form at [Link to final article].

b. The right to transmit, print and share copies with colleagues.

2. **Accepted Version.** Re-use of the accepted and peer-reviewed (but not final) version of the Contribution shall be by separate agreement with Wiley-Blackwell. Wiley-Blackwell has agreements with certain funding agencies governing reuse of this version. The details of those relationships, and other offerings allowing open web use, are set forth at the following website: <http://www.wiley.com/go/funderstatement>. NIH grantees should check the box at the bottom of this document.

3. **Final Published Version.** Wiley-Blackwell hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution:

a. Copies for colleagues. The personal right of the Contributor only to send or transmit individual copies of the final published version in any format to colleagues upon their specific request provided no fee is charged, and further-provided that there is no systematic distribution of the Contribution, e.g. posting on a listserve, website or automated delivery.

b. Re-use in other publications. The right to re-use the final Contribution or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications should be accurately noted.

c. Teaching duties. The right to include the Contribution in teaching or training duties at the Contributor's institution/place of employment including in course packs, e-reserves, presentation at professional conferences, in-house training, or distance learning. The Contribution may not be used in seminars outside of normal teaching obligations (e.g. commercial seminars). Electronic posting of the final published version in connection with teaching/training at the Contributor's institution/place of employment is permitted subject to the implementation of reasonable access control mechanisms, such as user name and password. Posting the final published version on the open Internet is not permitted.

d. Oral presentations. The right to make oral presentations based on the Contribution.

4. **Article Abstracts, Figures, Tables, Data Sets, Artwork and Selected Text (up to 250 words).**

a. Contributors may re-use unmodified abstracts for any non-commercial purpose. For on-line uses of the abstracts, Wiley-Blackwell encourages but does not require linking back to the final published versions.

b. Contributors may re-use figures, tables, data sets, artwork, and selected text up to 250 words from their Contributions, provided the following conditions are met:

- (i) Full and accurate credit must be given to the Contribution.
- (ii) Modifications to the figures, tables and data must be noted. Otherwise, no changes may be made.
- (iii) The reuse may not be made for direct commercial purposes, or for financial consideration to the Contributor.
- (iv) Nothing herein shall permit dual publication in violation of journal ethical practices.

D. CONTRIBUTIONS OWNED BY EMPLOYER

1. If the Contribution was written by the Contributor in the course of the Contributor's employment (as a "work-made-for-hire" in the course of employment), the Contribution is owned by the company/employer which must sign this Agreement (in addition to the Contributor's signature) in the space provided below. In such case, the company/employer hereby assigns to Wiley-Blackwell, during the full term of copyright, all copyright in and to the Contribution for the full term of copyright throughout the world as specified in paragraph A above.

2. In addition to the rights specified as retained in paragraph B above and the rights granted back to the Contributor pursuant to paragraph C above, Wiley-Blackwell hereby grants back, without charge, to such company/employer, its subsidiaries and divisions, the right to make copies of and distribute the final published Contribution internally in print format or electronically on the Company's internal network. Copies so used may not be resold or distributed externally. However the company/employer may include information and text from the Contribution as part of an information package included with software or other products offered for sale or license or included in patent applications. Posting of the final published Contribution by the institution on a public access website may only be done with Wiley-Blackwell's written permission, and payment of any applicable fee(s). Also, upon payment of Wiley-Blackwell's reprint fee, the institution may distribute print copies of the published Contribution externally.

E. GOVERNMENT CONTRACTS

In the case of a Contribution prepared under U.S. Government contract or grant, the U.S. Government may reproduce, without charge, all or portions of the Contribution and may authorize others to do so, for official U.S. Govern-

ment purposes only, if the U.S. Government contract or grant so requires. (U.S. Government, U.K. Government, and other government employees: see notes at end)

F. COPYRIGHT NOTICE

The Contributor and the company/employer agree that any and all copies of the final published version of the Contribution or any part thereof distributed or posted by them in print or electronic format as permitted herein will include the notice of copyright as stipulated in the Journal and a full citation to the Journal as published by Wiley-Blackwell.

G. CONTRIBUTOR'S REPRESENTATIONS

The Contributor represents that the Contribution is the Contributor's original work, all individuals identified as Contributors actually contributed to the Contribution, and all individuals who contributed are included. If the Contribution was prepared jointly, the Contributor agrees to inform the co-Contributors of the terms of this Agreement and to obtain their signature to this Agreement or their written permission to sign on their behalf. The Contribution is submitted only to this Journal and has not been published before. (If excerpts from copyrighted works owned by third parties are included, the Contributor will obtain written permission from the copyright owners for all uses as set forth in Wiley-Blackwell's permissions form or in the Journal's Instructions for Contributors, and show credit to the sources in the Contribution.) The Contributor also warrants that the Contribution contains no libelous or unlawful statements, does not infringe upon the rights (including without limitation the copyright, patent or trademark rights) or the privacy of others, or contain material or instructions that might cause harm or injury.

CHECK ONE BOX:

Contributor-owned work

ATTACH ADDITIONAL SIGNATURE
PAGES AS NECESSARY

Contributor's signature _____ Date _____

Type or print name and title _____

Co-contributor's signature _____ Date _____

Type or print name and title _____

Company/Institution-owned work

(made-for-hire in the
course of employment)

Company or Institution (Employer-for-Hire) _____ Date _____

Authorized signature of Employer _____ Date _____

U.S. Government work

Note to U.S. Government Employees

A contribution prepared by a U.S. federal government employee as part of the employee's official duties, or which is an official U.S. Government publication, is called a "U.S. Government work," and is in the public domain in the United States. In such case, the employee may cross out Paragraph A.1 but must sign (in the Contributor's signature line) and return this Agreement. If the Contribution was not prepared as part of the employee's duties or is not an official U.S. Government publication, it is not a U.S. Government work.

U.K. Government work
(Crown Copyright)

Note to U.K. Government Employees

The rights in a Contribution prepared by an employee of a U.K. government department, agency or other Crown body as part of his/her official duties, or which is an official government publication, belong to the Crown. U.K. government authors should submit a signed declaration form together with this Agreement. The form can be obtained via <http://www.opsi.gov.uk/advice/crown-copyright/copyright-guidance/publication-of-articles-written-by-ministers-and-civil-servants.htm>

Other Government work

Note to Non-U.S., Non-U.K. Government Employees

If your status as a government employee legally prevents you from signing this Agreement, please contact the editorial office.

NIH Grantees

Note to NIH Grantees

Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of Contributions authored by NIH grant-holders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

CIA-A

COLOR REPRODUCTION IN YOUR ARTICLE

These proofs have been typeset using figure files transmitted to production when this article was accepted for publication. Please review all figures and note your approval with your submitted proof corrections. In the event that color figure were included with the final manuscript files that we received for your article, this form must be completed and returned with your corrected proofs.

Because of the high cost of color printing, we can only print figures in color if authors cover the expense. Please indicate if you agree to pay charges for color figures to be printed in color (see table below for prices) or prefer black and white reproduction.

You will be invoiced for color charges once the article has been published in print

Failure to return this form with your article proofs may delay the publication of your article.

JOURNAL OF MEDICAL VIROLOGY

JOURNAL _____ MS. NO. _____ NO. COLOR PAGES _____

MANUSCRIPT TITLE _____

AUTHOR(S) _____

No. Color Pages	Color Charge	No. Color Pages	Color Charge	No. Color Pages	Color Charge
<input type="checkbox"/> 1	\$950	<input type="checkbox"/> 5	\$3400	<input type="checkbox"/> 9	\$5850
<input type="checkbox"/> 2	\$1450	<input type="checkbox"/> 6	\$3900	<input type="checkbox"/> 10	\$6350
<input type="checkbox"/> 3	\$1950	<input type="checkbox"/> 7	\$4400	<input type="checkbox"/> 11	\$6850
<input type="checkbox"/> 4	\$2450	<input type="checkbox"/> 8	\$4900	<input type="checkbox"/> 12	\$7350

*****Contact lorimarieh@woh.rr.com for a quote if you have more than 12 pages of color*****

Please print my figures color Please print my figures in black and white

Please print the following figures in color _____
and convert these figures to black and white _____

Approved by _____

Billing Address _____ E-mail _____

_____ Telephone _____

_____ Fax _____



WILEY

Publishers Since 1807

THE JOURNAL OF MEDICAL VIROLOGY

Telephone Number:

• Facsimile Number:

To: Ms. Lori-Marie Hamilton

Email: jmvprod@wiley.com; lorimarieh@woh.rr.com

From: Dr.

Date: _____

Re: *Journal of Medical Virology, ms #*

Dear Ms. Hamilton,

Attached please find corrections to ms# _____. Please contact me should you have any difficulty reading this fax at the numbers listed below.

Office phone:

Email:

Fax:

Lab phone:

I will return color figure proofs (if applicable) once I have checked them for accuracy.

Thank you,

Dr.

E-proofing feedback comments:



Additional reprint and journal issue purchases

Should you wish to purchase additional copies of your article, please click on the link and follow the instructions provided:
<https://caesar.sheridan.com/reprints/redir.php?pub=10089&acro=JMV>

Corresponding authors are invited to inform their co-authors of the reprint options available.

Please note that regardless of the form in which they are acquired, reprints should not be resold, nor further disseminated in electronic form, nor deployed in part or in whole in any marketing, promotional or educational contexts without authorization from Wiley. Permissions requests should be directed to mailto: permissionsus@wiley.com

For information about 'Pay-Per-View and Article Select' click on the following link: <http://wileyonlinelibrary.com/ppv>

Softproofing for advanced Adobe Acrobat Users - NOTES tool

NOTE: ADOBE READER FROM THE INTERNET DOES NOT CONTAIN THE NOTES TOOL USED IN THIS PROCEDURE.

Acrobat annotation tools can be very useful for indicating changes to the PDF proof of your article. By using Acrobat annotation tools, a full digital pathway can be maintained for your page proofs.

The NOTES annotation tool can be used with either Adobe Acrobat 4.0, 5.0 or 6.0. Other annotation tools are also available in Acrobat 4.0, but this instruction sheet will concentrate on how to use the NOTES tool. Acrobat Reader, the free Internet download software from Adobe, DOES NOT contain the NOTES tool. In order to softproof using the NOTES tool you must have the full software suite Adobe Acrobat 4.0, 5.0 or 6.0 installed on your computer.

Steps for Softproofing using Adobe Acrobat NOTES tool:

1. Open the PDF page proof of your article using either Adobe Acrobat 4.0, 5.0 or 6.0. Proof your article on-screen or print a copy for markup of changes.
2. Go to File/Preferences/Annotations (in Acrobat 4.0) or Document/Add a Comment (in Acrobat 6.0) and enter your name into the "default user" or "author" field. Also, set the font size at 9 or 10 point.
3. When you have decided on the corrections to your article, select the NOTES tool from the Acrobat toolbox and click in the margin next to the text to be changed.
4. Enter your corrections into the NOTES text box window. Be sure to clearly indicate where the correction is to be placed and what text it will effect. If necessary to avoid confusion, you can use your TEXT SELECTION tool to copy the text to be corrected and paste it into the NOTES text box window. At this point, you can type the corrections directly into the NOTES text box window. **DO NOT correct the text by typing directly on the PDF page.**
5. Go through your entire article using the NOTES tool as described in Step 4.
6. When you have completed the corrections to your article, go to File/Export/Annotations (in Acrobat 4.0) or Document/Add a Comment (in Acrobat 6.0).
7. **When closing your article PDF be sure NOT to save changes to original file.**
8. To make changes to a NOTES file you have exported, simply re-open the original PDF proof file, go to File/Import/Notes and import the NOTES file you saved. Make changes and re-export NOTES file keeping the same file name.
9. When complete, attach your NOTES file to a reply e-mail message. Be sure to include your name, the date, and the title of the journal your article will be printed in.

Original Article

A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis

Masayuki Kurosaki,¹ Kotaro Matsunaga,² Itsuko Hirayama,¹ Tomohiro Tanaka,¹ Mitsuaki Sato,¹ Yutaka Yasui,¹ Nobuharu Tamaki,¹ Takanori Hosokawa,¹ Ken Ueda,¹ Kaoru Tsuchiya,¹ Hiroyuki Nakanishi,¹ Hiroki Ikeda,¹ Jun Itakura,¹ Yuka Takahashi,¹ Yasuhiro Asahina,¹ Megumu Higaki,⁴ Nobuyuki Enomoto³ and Namiki Izumi¹

¹Division of Gastroenterology and Hepatology and ²Division of Pathology, Musashino Red Cross Hospital, Tokyo, ³First Department of Internal Medicine, University of Yamanashi, Yamanashi, and ⁴Department of Medical Science, Jikei Medical University, Tokyo, Japan

Aim: Early disappearance of serum hepatitis C virus (HCV) RNA is the prerequisite for achieving sustained virological response (SVR) in peg-interferon (PEG-IFN) plus ribavirin (RBV) therapy for chronic hepatitis C. This study aimed to develop a decision tree model for the pre-treatment prediction of response.

Methods: Genotype 1b chronic hepatitis C treated with PEG-IFN alpha-2b and RBV were studied. Predictive factors of rapid or complete early virological response (RVR/cEVR) were explored in 400 consecutive patients using a recursive partitioning analysis, referred to as classification and regression tree (CART) and validated.

Results: CART analysis identified hepatic steatosis (<30%) as the first predictor of response followed by low-density-lipoprotein cholesterol (LDL-C) (≥ 100 mg/dL), age (<50 and <60 years), blood sugar (<120 mg/dL), and gamma-glutamyltransferase (GGT) (<40 IU/L) and built decision tree

model. The model consisted of seven groups with variable response rates from low (15%) to high (77%). The reproducibility of the model was confirmed by the independent validation group ($r^2 = 0.987$). When reconstructed into three groups, the rate of RVR/cEVR was 16% for low probability group, 46% for intermediate probability group and 75% for high probability group.

Conclusions: A decision tree model that includes hepatic steatosis, LDL-C, age, blood sugar, and GGT may be useful for the prediction of response before PEG-IFN plus RBV therapy, and has the potential to support clinical decisions in selecting patients for therapy and may provide a rationale for treating metabolic factors to improve the efficacy of antiviral therapy.

Key words: data mining, decision tree, HCV, low-density-lipoprotein-cholesterol, steatosis

INTRODUCTION

COMBINATION THERAPY WITH pegylated interferon (PEG-IFN) and ribavirin (RBV) is now recognized as a standard treatment for patients with chronic hepatitis C.¹ However, the rate of sustained virological response (SVR) to 48 weeks of PEG-IFN RBV combina-

tion therapy is only 50% in patients with hepatitis C virus (HCV) genotype 1b and high HCV RNA titer, so called difficult to treat chronic hepatitis C patients.^{2,3} Within this difficult to treat group, the response to treatment sometimes can be highly heterogeneous for cases which are apparently equivalent in HCV RNA titer, making the prediction of response before treatment a difficult task. It has been suggested that early virological response (EVR), defined as either undetectable HCV RNA or a 2 log drop in HCV RNA at week 12, is a reliable means to predict SVR.^{2,4} More recently, it has been suggested that patients with a rapid virological response (RVR: undetectable HCV RNA at week 4) and a complete EVR (cEVR: undetectable HCV RNA at week 12)

Correspondence: Dr 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 26 May 2009; revision 25 August 2009; accepted 26 August 2009.

achieve high SVR rates, while patients with a partial EVR (pEVR: 2 log drop in HCV RNA but still detectable at week 12) have lower rates of SVR.⁵ Since PEG-IFN RBV combination therapy is costly and accompanied by potential adverse effects, the ability to predict the possibility of RVR or cEVR before therapy and identifying curable patients may significantly influence the selection of patients for therapy. Moreover, identification of baseline predictors of poor response is particularly important to establish a rationale for identifying therapeutic targets to improve the efficacy of antiviral therapy.

Data mining is a method of predictive analysis which explores tremendous volumes of data to discover hidden patterns and relationships in highly complex datasets and enables the development of predictive models. The classification and regression tree (CART) analysis is a core component of the decision tree tool for data mining and predictive modeling,⁶ is deployed to decision makers in various fields of business, and currently is being used in the area of biomedicine.⁷⁻¹³ The results of CART analysis are presented as a decision tree, which is intuitive and facilitates the allocation of patients into subgroups by following the flow-chart form.¹⁴ CART has been shown to be competitive with other traditional statistical techniques such as logistic regression analysis.¹⁵

In the present study, we used the CART analysis to explore baseline predictors of response to PEG-IFN plus RBV therapy among clinical, biochemical, virological and histological pretreatment variables and to define a pre-treatment algorithm to discriminate chronic hepatitis C patients who are likely to respond to PEG-IFN plus RBV therapy.

MATERIALS AND METHODS

Patients

A TOTAL OF 419 chronic hepatitis C patients were treated with PEG-IFN alpha-2b and RBV at Musashino Red Cross Hospital between December 2001 and December 2007. Among them, 400 patients who fulfilled the following inclusion criteria were enrolled in the present study. (i) infection by genotype 1b (ii) HCV RNA higher than 100 KIU/mL by quantitative PCR (Cobas Amplicor HCV Monitor, Roche Diagnostic systems, CA) which is usually used for the definition of high viral load in Japan (iii) lack of co-infection with hepatitis B virus or human immunodeficiency virus (iv) lack of other causes of liver disease such as autoimmune hepatitis, primary biliary cirrhosis, or alcohol intake of more than 20 g per day, and (v) having completed at

least 12 weeks of therapy with an early virological response that could be evaluated. Patients received PEG-IFN alpha-2b (1.5 microgram/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 1000 mg for >80 kg) which is the recommended dosage in Japan. Data from two third of patients (269 patients) were used for the model building set and the remaining one third of patients (131 patients) were used as a validation set. Consent in writing was obtained from each patient and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the institutional review committee.

Laboratory tests

Blood samples were obtained before therapy, and at least once every month during therapy and analyzed for hematologic tests, blood chemistries, and HCV RNA. In the present study, RVR and cEVR was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/mL (Amplicor, Roche Diagnostic systems, CA) at week 4 and 12, respectively. SVR was defined as undetectable HCV RNA at week 24 after the completion of therapy.

Histological examination

For all patients, liver biopsy specimens were obtained before therapy and were evaluated independently by three pathologists who were blinded to the clinical details. If there was a disagreement, the scores assigned by the majority of pathologists were used for the analysis. Fibrosis and activity were scored according to the METAVIR scoring system.¹⁶ 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 0–3: grade 0 (no steatosis), grade 1 (0–9%), grade 2 (10–29%), and grade 3 (over 30%) as we reported previously.¹⁷

Database for analysis

A pretreatment database of 72 variables was created containing histological findings (grade of fibrosis, activity, and steatosis), laboratory tests including the quantity of HCV RNA by Cobas Amplicor, and clinical information (age, gender, body weight, and body mass index).

The baseline characteristics and test results are listed in Table 1. The overall rate of RVR/cEVR was 43% in the model building set and 48% in the validation set. There were no significant differences in the clinical backgrounds between these two groups. Hepatitis C viral mutations, such as mutations in interferon-sensitivity determining region or core amino acid residues 70 and 91, were not included in the present analysis. The dataset of laboratory tests was based on the digitized records in this hospital. Continuous data was split into categorized data by increment of 10; For example, age was categorized into <30, 30–39, 40–49, 50–59, 60–69, and ≥70.

Statistical analysis

Based on this database, the recursive partitioning analysis algorithm referred to as CART was implemented to define meaningful subgroups of patients with respect to the possibility of achieving RVR/cEVR. The CART belongs to a family of nonparametric regression methods based on binary recursive partitioning of data. The software automatically explore the data to search for optimal split variables, builds a decision tree structure and finally classifies all subjects into particular subgroups that are homogeneous with respect to the outcome of interest.¹⁸ During the CART analysis, first, the entire study population, and thereafter, all newly defined subgroups, were investigated at every step of the analysis to determine which variable at what cut-off point yielded the most significant division into two prognostic subgroups that were as homogeneous as possible with respect to estimates of RVR/cEVR possibilities. This algorithm uses the impurity function (Gini criterion function) for splitting.¹⁹ A restriction was imposed on the tree construction such that terminal subgroups resulting from any given split must have at least 20 patients. The CART procedure stopped when either no additional significant variable was detected or when the sample size was below 20. The resulting final subgroups were most homogeneous with respect to the probability of achieving RVR/cEVR. For this analysis, data mining software Clementine version 12.0 (SPSS Inc, Chicago, IL) was utilized. SPSS 15.0 (SPSS Inc, Chicago, IL) was used for logistic regression analysis.

RESULTS

Factors associated with RVR/cEVR by standard statistical analysis

WE FIRST ANALYZED 72 variables by univariate and multivariate logistic regression analysis to find factors associated with RVR/cEVR (Table 2).

Patients with RVR/cEVR were significantly younger than those without. Among histological findings, grade of steatosis and stage of fibrosis was significantly lower in RVR/cEVR. Among hematologic tests, hemoglobin and hematocrit was significantly higher in RVR/cEVR. Among blood chemistry tests, creatinine and low-density lipoprotein cholesterol (LDL-C) was significantly higher and gamma-glutamyltransferase (GGT), low-density-lipoprotein cholesterol (LDL-C), and blood sugar were significantly lower in RVR/cEVR. The level of HCV RNA was significantly lower in RVR/cEVR. There were no significant differences in other tests.

Multivariate logistic regression analysis was performed on age, fibrosis stage, steatosis, HCV RNA, creatinine, hemoglobin, GGT, LDL-C, and blood sugar: hematocrit was not included since it is closely associated with hemoglobin. On multivariate analysis, age, grade of steatosis, level of HCV RNA, creatinine, hemoglobin, GGT, and LDL-cholesterol remained significant whereas stage of fibrosis, hemoglobin and blood sugar were not.

The CART analysis

The CART analysis was carried out on the model building set of 269 patients using the same variables as logistic regression analysis. Figure 1 shows the resulting decision tree. The CART analysis automatically selected five predictive variables to produce a total of seven subgroups of patients. The grade of steatosis was selected as the variable of initial split with an optimal cut-off of 30%. The possibility of achieving RVR/cEVR was only 18% for patients with hepatic steatosis of 30% or more compared to 47% for patients with hepatic steatosis of less than 30%. Among patients with hepatic steatosis of less than 30%, the level of serum LDL-C, with an optimal cut-off of 100 mg/dL, was selected as the variable of second split. Patients with higher LDL-C level had the higher probability of RVR/cEVR (57% vs. 32%). Among patients with LDL-C of less than 100 mg/dL, age, with an optimal cut-off of 60, was selected as the third variable of split. Younger patients had the higher probability of RVR/cEVR (49% vs. 15%). Among patients younger than 60, the blood sugar, with an optimal cut-off of 120 mg/dL, was selected as the fourth variable of split. Patients with lower blood sugar level had the higher probability of RVR/cEVR (71% vs. 31%). Among patients with hepatic steatosis of less than 30% and LDL-C of 100 mg/dL or more, age, with an optimal cut-off of 50, was selected as the third variable of split, younger being the predictor of higher RVR/cEVR probability (77% vs. 50%). Among patients older than 50,

Table 1 Clinical characteristics of patients

	Model set n = 269	Validation set n = 131	P-value
Sex (M/F)	127/142	55/76	0.325
Age (years)	57.7 ± 10.1	57.6 ± 10.0	0.932
Body weight (kg)	59.6 ± 11.0	57.5 ± 9.5	0.094
Body mass index (kg/m ²)	23.2 ± 3.1	23.3 ± 3.8	0.934
Total protein (g/dL)	7.6 ± 0.5	7.7 ± 0.6	0.558
Albumin (g/dL)	4.2 ± 0.3	4.2 ± 0.3	0.349
Globulin (g/dL)	3.4 ± 0.5	3.4 ± 0.6	0.989
Aspartate aminotransferase (IU/L)	58.1 ± 43.1	55.8 ± 37.5	0.601
Alanine aminotransferase (IU/L)	70.9 ± 49.2	66.4 ± 52.6	0.462
Gamma-glutamyltransferase (IU/L)	49.6 ± 44.0	45.2 ± 34.4	0.33
Lactate dehydrogenase (IU/L)	289.3 ± 112.3	301.5 ± 109.3	0.417
Total bilirubin (mg/dL)	0.71 ± 0.28	0.69 ± 0.23	0.317
Direct bilirubin (mg/dL)	0.23 ± 0.12	0.25 ± 0.10	0.147
Indirect bilirubin (mg/dL)	0.48 ± 0.21	0.44 ± 0.16	0.064
Alkaline phosphatase (IU/L)	290.9 ± 107.6	292.5 ± 107.6	0.917
Leucine aminopeptidase (IU/L)	64.3 ± 14.3	65.5 ± 12.3	0.543
Thymol turbidity test (KU)	7.1 ± 3.4	8.0 ± 3.7	0.062
Zinc sulfate turbidity test (KU)	15.4 ± 4.9	16.3 ± 5.4	0.188
Choline esterase (IU/L)	318.1 ± 81.7	321.1 ± 78.1	0.798
Ammonia (microg/dL)	39.7 ± 20.2	45.0 ± 15.6	0.668
Blood sugar (mg/dL)	125.9 ± 41.1	117.4 ± 47.9	0.081
Glycohemoglobin (%)	5.6 ± 1.6	5.4 ± 1.2	0.797
Total cholesterol (mg/dL)	170.8 ± 33.9	175.6 ± 36.8	0.170
Low-density-lipoprotein-cholesterol (mg/dL)	96.5 ± 25.2	100.9 ± 28.5	0.153
High-density-lipoprotein-cholesterol (mg/dL)	54.2 ± 15.9	55.2 ± 17.4	0.612
Triglyceride (mg/dL)	108.5 ± 47.8	102.8 ± 46.4	0.306
Creatinine (mg/dL)	0.72 ± 0.15	0.74 ± 0.17	0.236
Urea nitrogen (mg/dL)	14.1 ± 3.4	14.9 ± 3.9	0.123
Uric acid (mg/dL)	5.3 ± 1.2	5.2 ± 1.2	0.715
Sodium (mEq/L)	142.2 ± 2.0	142.4 ± 2.0	0.471
Potassium (mEq/L)	4.3 ± 0.3	4.3 ± 0.4	0.578
Chloride (mEq/L)	104.0 ± 2.2	104.0 ± 2.6	0.905
Calcium (mg/dL)	9.1 ± 0.4	9.2 ± 0.4	0.479
Phosphorus (mg/dL)	3.5 ± 0.5	3.5 ± 0.6	0.814
Magnesium (mg/dL)	2.2 ± 0.2	2.3 ± 0.3	0.390
Amylase (IU/L)	178.7 ± 125.8	175.1 ± 133.1	0.118
Creatine kinase (IU/L)	114.9 ± 147.6	119.3 ± 73.7	0.849
Iron (microg/dL)	104.7 ± 53.2	109 ± 37	0.726
Ferritin (ng/mL)	111.3 ± 103.3	59.7 ± 118.5	0.405
C-reactive peptide (mg/dL)	0.2 ± 1.1	0.1 ± 0.1	0.586
Immunoglobulin G (mg/dL)	1849 ± 426	1988 ± 525	0.129
Immunoglobulin M (mg/dL)	141 ± 69	205 ± 106	0.200
Immunoglobulin A (mg/dL)	323 ± 675	291 ± 81	0.784
Triiodothyronine (pg/mL)	2.3 ± 0.3	2.2 ± 0.3	0.358
Thyroxin (ng/dL)	0.9 ± 0.1	0.9 ± 0.1	0.872
Thyroid stimulating hormone (micro IU/mL)	1.8 ± 1.4	1.7 ± 0.7	0.939
White blood cell count (/microl)	5243 ± 1591	5286 ± 1101	0.843
Segmented neutrophils (%)	55.4 ± 10.8	57.0 ± 10.0	0.297
Band neutrophils (%)	1.5 ± 1.6	0.5 ± 0.6	0.250
Eosinophils (%)	2.9 ± 2.3	2.4 ± 1.4	0.127

Table 1 Continued

	Model set n = 269	Validation set n = 131	P-value
Basophiles (%)	0.6 ± 0.4	0.6 ± 0.3	0.727
Lymphocytes (%)	34.6 ± 9.6	34.0 ± 9.3	0.682
Monocytes (%)	6.6 ± 2.2	6.2 ± 2.6	0.149
Red blood cell count (10 ⁴ /microl)	458 ± 43	455 ± 47	0.643
Hemoglobin (g/dL)	14.4 ± 1.5	14.5 ± 1.5	0.618
Hematcrit (%)	42.7 ± 4.0	42.9 ± 4.4	0.717
Reticulocytes (%)	1.4 ± 0.4	1.4 ± 0.4	0.762
Mean corpuscular volume (fL)	93.3 ± 4.5	93.8 ± 5.41	0.466
Mean corpuscular hemoglobin concentration (pg)	31.5 ± 1.9	31.7 ± 2.3	0.583
Mean corpuscular hemoglobin concentration (g/dL)	33.8 ± 0.9	33.7 ± 1.3	0.910
Platelets (10 ⁴ /microl)	16.8 ± 5.4	16.3 ± 4.5	0.480
Prothrombin time (s)	11.7 ± 1.2	11.7 ± 0.9	0.762
Prothrombin time (activity %)	104.6 ± 14.4	102.6 ± 14.8	0.363
Prothrombin time (international normalized ratio)	1.0 ± 0.1	1.0 ± 0.1	0.387
Thrombin time (%)	97.2 ± 31.3	109 ± 31.5	0.231
Activated partial thromboplastin time (s)	29.7 ± 4.4	29.1 ± 2.7	0.260
Hepaplastin test (%)	97.8 ± 20.3	95.4 ± 19.4	0.523
Fibrinogen (%)	237 ± 44	225 ± 45	0.069
Hepatitis C virus RNA (<850/≥850 KIU/mL)	130/139	70/61	0.394
Histological grade of			
Activity (A1/A2/A3)	138/107/24	62/55/14	0.714
Fibrosis (F1/F2/F3/F4)	135/74/57/3	58/40/27/6	0.131
Steatosis (0%/1-9%/10-29%/30%≤)	89/109/37/34	49/45/21/16	0.643
Hepatitis C virus RNA negative at week 12 (yes/no)	116/153	63/68	0.349

the level of GGT, with an optimal cutoff of 40 U/L, were then selected as the fourth level of split, low levels being the predictor of higher RVR/cEVR probability (60% vs. 35%).

All five factors selected as significant variables in the CART analysis were also significantly associated with RVR/cEVR by univariate analysis (Table 2). In addition, steatosis, LDL-C, age and GGT were also independently

Table 2 Factors associated with rapid or complete early virological response by univariate and multivariate logistic regression analysis

Parameter	Category	Univariate			Multivariate		
		Odds	95% CI	P-value	Odds	95% CI	P-value
Age (years)	<50 vs. ≥50	2.65	1.51-4.65	<0.001	2.03	1.04-3.97	0.039
Fibrosis stage	F1-2 vs. F3-4	2.47	1.31-4.66	0.005	1.77	0.85-3.68	0.120
Steatosis (%)	<30 vs. ≥30	4.11	1.64-10.29	0.003	2.88	1.07-7.79	0.037
Hepatitis C virus RNA (KIU/mL)	<850 vs. ≥850	1.97	1.21-3.22	0.007	1.93	1.09-3.43	0.025
Creatinine (mg/dL)	≥0.8 vs. <0.8	3.30	1.96-5.56	<0.001	3.54	1.88-6.67	<0.001
Hemoglobin (g/dL)	≥14.5 vs. <14.5	1.76	1.08-2.87	0.023	1.38	0.74-2.57	0.320
Hematcrit (%)	≥43 vs. <43	1.75	1.07-2.84	0.003			
Gamma-glutamyltransferase (IU/L)	<40 vs. ≥40	2.06	1.26-3.37	0.004	2.45	1.32-4.56	0.005
Low-density-lipid cholesterol (mg/dL)	≥100 vs. <100	2.71	1.61-4.55	<0.001	2.21	1.21-4.06	0.010
Blood sugar (mg/dL)	<120 vs. ≥120	2.00	1.02-3.95	0.045	1.42	0.64-3.13	0.390

CI, confidence interval.

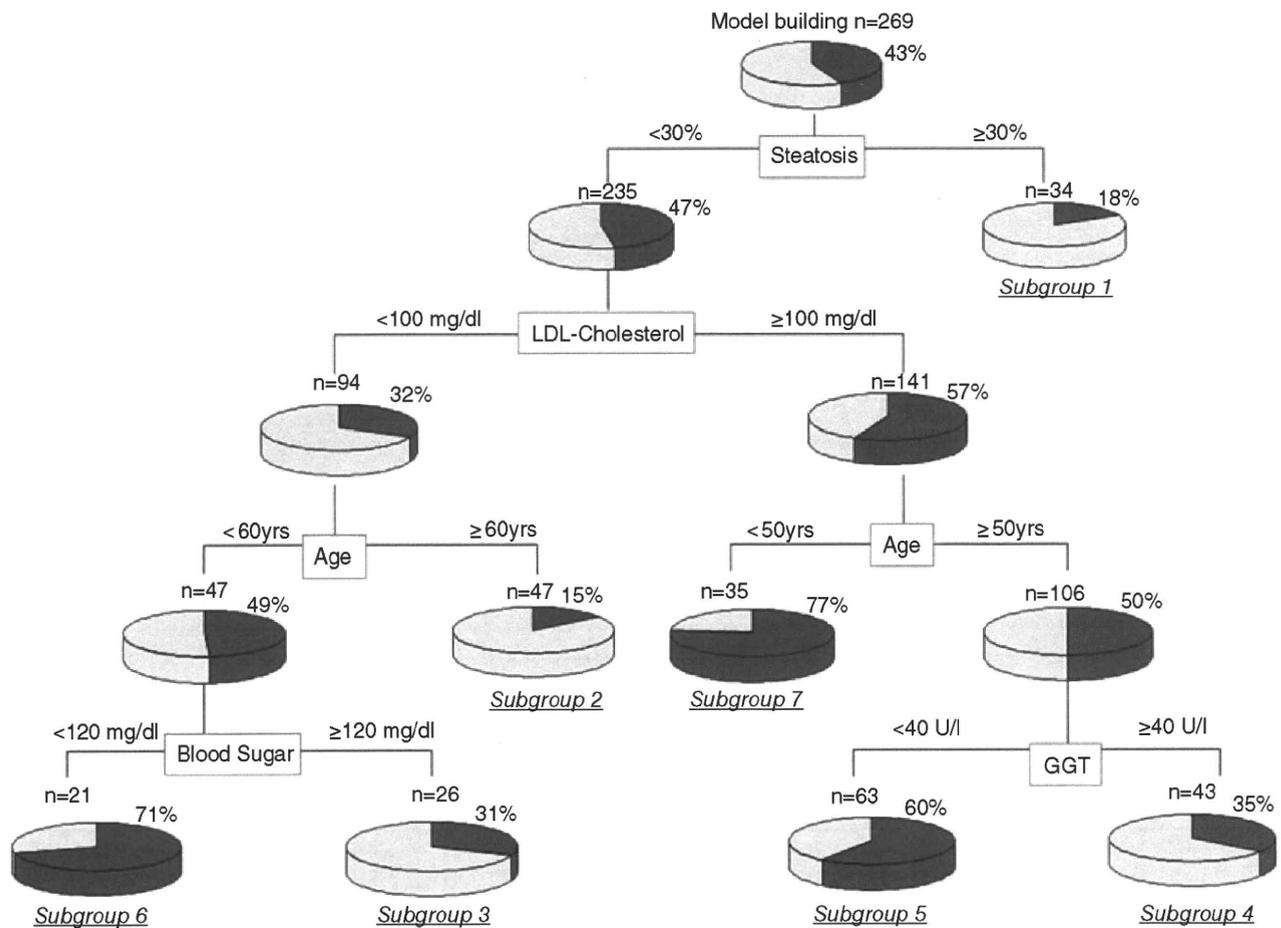


Figure 1 Classification and regression tree analysis. Boxes indicate the factors used for splitting and the cut-off value for the split. Pie charts indicate the rate of RVR/cEVR for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from one to seven. GGT, gamma-glutamyltransferase; LDL, low-density-lipoprotein.

associated with RVR/cEVR by multivariate logistic regression analysis while blood sugar was not (Table 2). On the other hand, HCVRNA and creatinine which were significantly associated with RVR/cEVR by multivariate analysis were not selected as significant variables in CART analysis.

The probabilities of RVR/cEVR for the seven subgroups derived by this process were highly variable. The subgroup whose hepatic steatosis was less than 30%, serum LDL-C was 100 mg/dL or more and of an age less than 50 years (subgroup 7) showed the highest probability of RVR/cEVR (77%), while the subgroup whose hepatic steatosis more than 30% (subgroup 1) and the subgroup whose hepatic steatosis was less than 30% but serum LDL-C was less than 100 mg/dL and of an age

greater than 60 years (subgroup 2) showed the lowest probability of RVR/cEVR (18% and 15%, respectively).

Validation of the CART analysis

The results of the CART analysis were validated with a validation dataset of 131 cases which is 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 CART tree. The rates of RVR/cEVR were 20% for subgroups 1 and 2, 29% for subgroups 3, 38% for subgroup 4, 59% for subgroup 5, 71% for subgroup 6, and 85% for subgroups 7. The rates of RVR/cEVR for each subgroup of patients were closely correlated between the model building dataset and the validation dataset (Fig. 2).

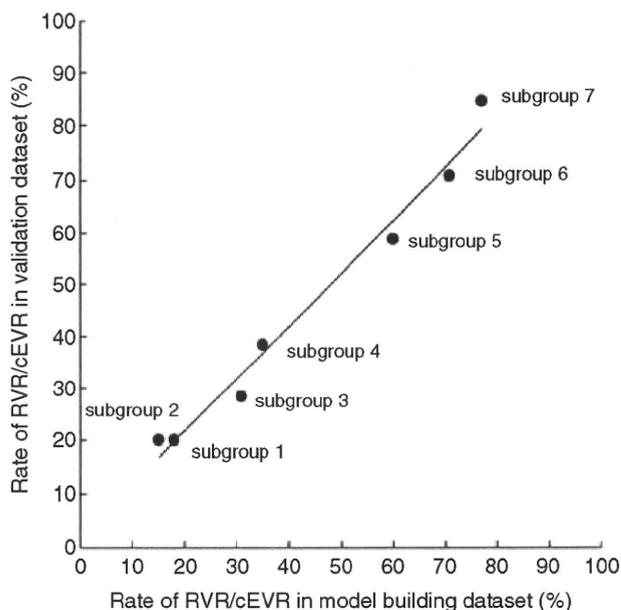


Figure 2 Validation of the classification and regression tree (CART) analysis: Subgroup stratified comparison of the rate of rapid or complete early virological response (RVR/cEVR) between the model building and validation datasets. Each patient in the validation set was allocated to subgroups 1–7 by following the flow-chart form of the CART tree and the rates of RVR/cEVR were calculated. The rate of RVR/cEVR in each subgroup was plotted. The x-axis represents the rate of RVR/cEVR in the model building datasets and the y-axis represents the rate of RVR/cEVR in the validation datasets. The rates of achieving RVR/cEVR in each subgroup of patients closely correlated between the model building dataset and the validation dataset ($r^2 = 0.987$).

Construction of 3 groups according to the probability of RVR/cEVR

If the seven subgroups were reconstructed into three groups according to their rate of RVR/cEVR, the rate of RVR/cEVR was 16% for low probability group (subgroup 1 and 2), 46% for intermediate probability group (subgroup 3, 4, and 5) and 75% for high probability group (subgroup 6 and 7; $P < 0.0001$).

Effect of adherence

Adherence of PEG-IFN and RBV was not included as a variable of analysis since the present study aimed to develop a pre-treatment model for the prediction of response. To analyze the possible effect of adherence on the result of CART analysis, three groups of patients divided by CART (low, intermediate and high probability group) were further stratified according to adherence

of PEG-IFN and RBV. Poor adherence was defined as taking less than 80% planned dose of PEG-IFN or RBV at 12 weeks, and good adherence was defined as taking more than 80% planned dose of both PEG-IFN and RBV at 12 weeks. The result is shown in Figure 3. Among patients with good adherence, the rate of RVR/cEVR was 19% for low probability group, 52% for intermediate probability group and 77% for high probability group. Among poor adherence group, the rate of RVR/cEVR was 13% for low probability group, 41% for intermediate probability group and 73% for high probability group. Collectively, even after adjustment for adherence, 3 groups of patients divided by CART analysis still had low, intermediate and high probability of achieving RVR/cEVR, respectively.

DISCUSSION

IN THE PRESENT study, we performed the CART analysis and built a simple decision tree model for the pre-treatment prediction of response to PEG-IFN plus

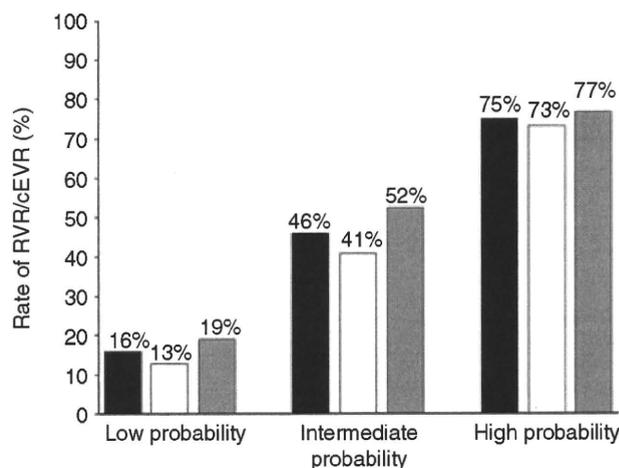


Figure 3 The rate of rapid or complete early virological response (RVR/cEVR) between the classification and regression tree (CART) groups stratified by adherence. The three groups of patients divided by CART (low, intermediate and high probability group) were further stratified according to adherence of peg-interferon (PEG-IFN) plus ribavirin (RBV). Black, white and gray boxes in the bar chart indicate total patients, patients with poor adherence (taking less than 80% planned dose of PEG-IFN or RBV at 12 weeks), and good adherence (taking more than 80% planned dose of both PEG-IFN and RBV at 12 weeks), respectively. Even after adjustment for adherence, 3 groups of patients divided by CART analysis still had low, intermediate and high probability of achieving RVR/cEVR, respectively.

RBV therapy. The analysis highlighted five host variables relevant to response: steatosis, LDL-C, age, blood sugar and GGT. Classification of patients based on these variables identified subgroups of patients with high probabilities of achieving RVR/cEVR among difficult to treat chronic hepatitis C patients. The reproducibility of the model was confirmed by the independent validation datasets. According to the result of the CART, patients were categorized into 3 groups: the rate of RVR/cEVR was 16% for low probability group, 46% for intermediate probability group and 75% for high probability group. The result of the CART analysis could be readily applicable to clinical practice because patients could be allocated to specific subgroups with a defined rate of response simply by following the flow-chart form. Although an early disappearance of serum HCV RNA is the prerequisite for achieving SVR, no reliable baseline predictors of response to PEG-IFN plus RBV therapy are established to date. Thus, this model may have the potential to support decisions in patient selection for PEG-IFN plus RBV therapy or to tailor treatment strategies for individual patients. Moreover, our result may provide a rationale for treating metabolic factors to improve the efficacy of antiviral therapy.

Among variables relevant to the prediction of RVR/cEVR, the grade of hepatic steatosis was selected as the variable of the first split. Previous studies suggested that steatosis induces resistance to IFN and RBV combination therapy^{20,21} along with underlining metabolic factors such as insulin resistance or obesity.²¹⁻²⁴ In the present study, the grade of steatosis correlated positively with BMI and serum glucose level (data not shown) suggesting the etiologic role of metabolic factors. In addition, serum glucose level was selected as a predictor of RVR/cEVR at the fourth level of split. Serum GGT, which is associated with obesity,²⁵ insulin resistance²⁶ and response to IFN therapy,²⁷⁻³⁰ was also selected as a predictor of RVR/cEVR at fourth level of splitting which may emphasize the importance of metabolic factors in therapeutic resistance. These findings raise the possibility that treatment of these metabolic factors may improve the virological response to the PEG-IFN plus RBV therapy. This hypothesis should be examined by a prospective study.

We and others have reported that steatosis, obesity and insulin resistance are associated with the progression of fibrosis,^{17,31-33} which can interfere indirectly with the effect of IFN on hepatocytes. Other possible mechanisms of resistance by steatosis or metabolic factors include dysregulation of adipocytokines³⁴ or oxidative stress which may inhibit intracellular IFN signaling

pathway.³⁵ Despite these findings, the precise mechanism of resistance is not established and further investigation is needed.

Another factor relevant in the prediction of RVR/cEVR was LDL-C. LDL-C was selected as the second factor for splitting by CART, and was an independent predictor of RVR/cEVR by logistic regression analysis. LDL-C recently has attracted attention as a novel predictor of response to IFN or PEG-IFN plus RBV.^{30,36,37} Since *in vitro* study showed that LDL-C receptor acts as a receptor for HCV and LDL-C competitively inhibit the binding of HCV,³⁸ high level of serum LDL-C may inhibit HCV entry to hepatocytes and attenuate replication. LDL-C and its receptor may be a future therapeutic target.

Not all factors selected as significant variables in the CART analysis were also significantly associated with response by standard statistical analysis: blood sugar was associated with response by univariate analysis but not by multivariate logistic regression analysis. On the other hand, HCV RNA and creatinine which were significantly associated with RVR/cEVR by multivariate analysis were not selected as significant variables in CART analysis. These differences may indicate both the unique feature and the limitations of the CART analysis. To note, blood sugar was significantly associated with RVR/cEVR within specialized subgroups of patients defined by the CART analysis: in subgroup of patients with steatosis <30%, LDL-C <100 mg/dL and younger than 60, which indicate the unique feature of the CART analysis that it could visualize significant predictors that specifically apply to selected patients. The limitation is that not all significant factors may be adopted in the decision tree since we applied the rule to stop CART procedure when the sample size was below 20. This rule was applied to avoid the generation of over-fit model which may lack universality. Therefore, it is possible that HCV RNA or creatinine may become a significant variable in the CART analysis if larger number of patients were included in the analysis. Stage of fibrosis was significantly associated with response to therapy by univariate analysis but not by multivariate analysis and not selected as a significant variable in the CART analysis. The possible reason is that advanced fibrosis is associated with older age as a confounding factor.

CART analyses are gaining acceptance in medical research in addition to biomedical field. Recent publications include the prediction of aggressive prostate cancer,⁸ diabetic vascular complications,¹⁹ prognosis of melanoma,^{7,39} response to preoperative radiotherapy for rectal tumor,⁹ prognostic groups in colorectal carcinoma,¹² and outcome after liver failure.¹¹ An advantage