Hospital Organization (Japan). A dataset collected from 598 patients who were treated with PEG-IFN-α2b plus RBV and had undetectable HCV RNA within week 12 were used for external validation. 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

Haematological tests, blood chemistry and HCV RNA titre were analysed before therapy and at least once every month during therapy. Rapid virological response (RVR) was defined as an undetectable HCV RNA level at week 4, and complete early virological response (cEVR) was defined as an undetectable HCV RNA level at week 5 through week 12 after the start of therapy. SVR was defined as an undetectable HCV RNA level 24 weeks after the completion of therapy. Detection of HCV RNA level was based on qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor; Roche Diagnostic Systems). A database of pretreatment variables included haematological tests (haemoglobin level, white blood cell count and platelet count), blood chemistry tests (serum levels of creatinine, albumin, aspartate aminotransferase, alanine aminotransferase, γ-glutamyltransferase, total cholesterol, triglycerides and HCV RNA titre), stage of histological fibrosis and patient characteristics (age, sex and body mass index). Post-treatment variables included time to HCV RNA negativity, calculated total RBV dose (g/kg of body weight), and calculated total PEG-IFN dose (μg/kg of body weight).

Statistical analysis

The 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. Logistic regression models with backward selection procedures were used for multivariable analysis of factors associated with relapse. IBM SPSS software version 18.0 (SPSS Inc., Chicago, IL, USA) was used for analysis. For the decision tree analysis [30], the data mining software IBM SPSS Modeler 14 (SPSS Inc.) was used, as reported previously [17,26–28]. The decision tree analysis, the core component of the data mining, belongs to a family of non-parametric regression methods based on binary recursive partitioning of data. In this analysis, the software automatically explored the database to determine optimal split variables to build a decision tree structure. A statistical search algorithm evaluate the model derivation group to determine the optimum variables and cutoff values and to yield the most significant division of patients into two subgroups that were as homogeneous as possible for the probability of relapse. Once patients were divided into 2 subgroups, the analysis was automatically repeated on each subgroup in the same way until either no additional significant variable was detected or the number of patients was <20. Finally all patients were classified into particular subgroups that are homogeneous with respect to the probabilities of relapse.

Results

The decision tree model for the prediction of relapse

The overall rate of relapse was 26% in the model derivation group. The decision tree analysis selected three variables that are associated with relapse: time to HCV RNA negativity, age and total RBV dose (Figure 1). Time to HCV RNA negativity was selected as the best predictor of relapse. The rate of relapse was 13% for patients with RVR compared to 30% for patients with cEVR. Among patients with cEVR, age was selected as the variable of second split. Patients <60 years had a lower probability of relapse (22%) compared with those ≥60 years (41%). The total RBV dose was selected as the third variable of split with an optimal cutoff of 3.0 g/kg of body weight. The rate of relapse was lower in patients who received ≥3.0 g/kg of body weight of RBV compared to patients who received <3.0 g/kg of body weight (among patients <60 years rates were 16% versus 32% and among patients ≥60 years rates were 26% versus 52%, respectively).

According to this decision tree, the patients were divided into five groups with different rates of relapse ranging from 13% to 52%. Patients with RVR had the lowest risk of relapse. Among patients with cEVR, patients <60 years who received ≥3.0 g/kg of body weight of RBV also had a low risk of relapse (16%). By contrast, patients who received <3.0 g/kg of body weight of RBV had higher than the average risk of relapse, especially in patients ≥60 years (52%).

Validation of the decision tree model

The decision tree model was validated using an internal validation group that was not included in the model derivation. The rates of relapse for each subgroup of patients were correlated closely between the model derivation and the internal validation group (r^2=0.79; Figure 2A). When validated using an external validation group, the rates of relapse for each subgroup of patients were again correlated closely between the model derivation and the external validation group. (r^2=0.83; Figure 2B).

Multivariable logistic regression analysis for factors associated with relapse

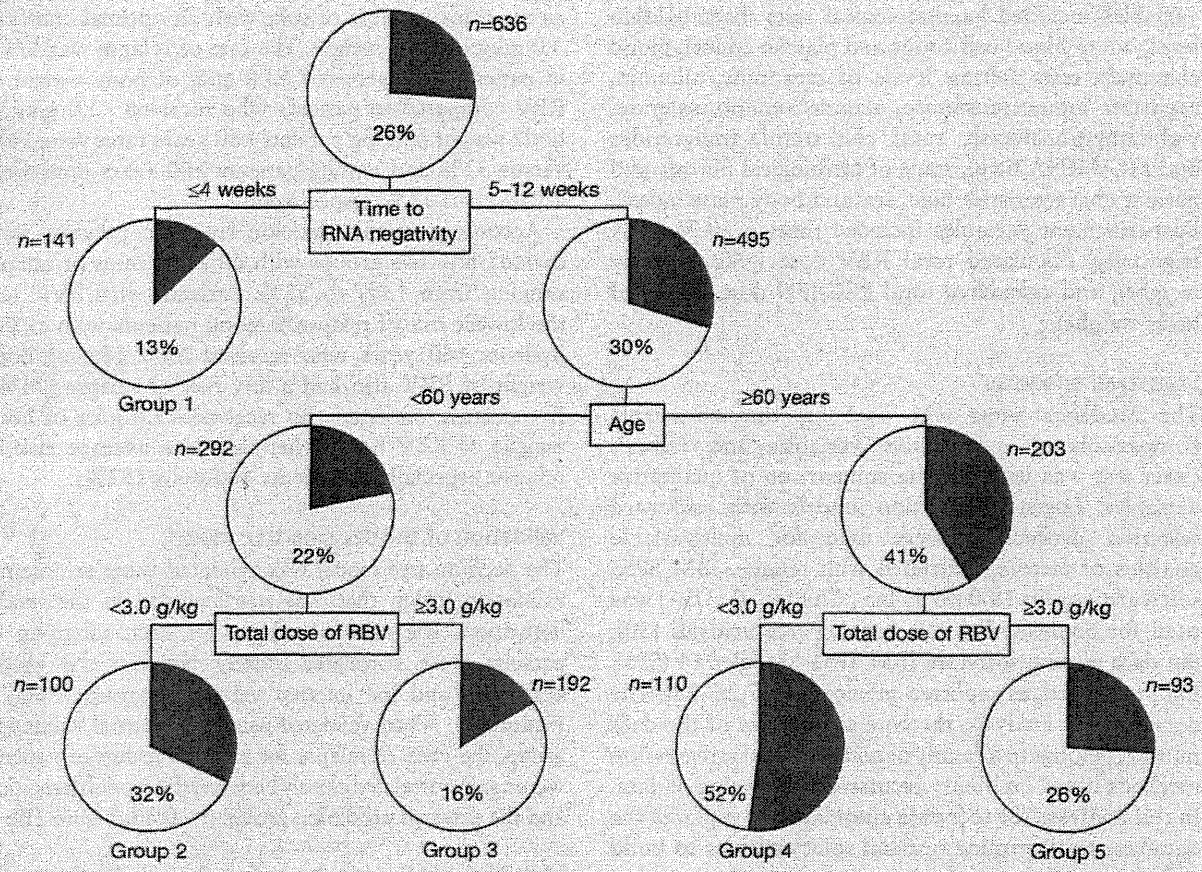
Univariable and multivariable analysis was performed using the combined population of model derivation and internal validation group. Univariable analysis found

that age, sex, serum levels of creatinine, haemoglobin, platelet count, HCV RNA titre, time to HCV RNA negativity, total PEG-IFN dose and total RBV dose were associated with relapse. Duration of therapy was not associated with reduction in relapse rate. Multivariable analysis including these factors showed that age, total RBV dose, serum level of creatinine, and time to HCV RNA negativity were independent predictors of relapse (Table 2). Creatinine was not selected as a splitting variable in data mining analysis probably due to the limitation to stop the analysis when the number of patients was <20. Using the combined population of model derivation and internal validation group, patients in each subgroup of decision tree model were further stratified by creatinine levels and the effect of creatinine level on relapse was analysed. Among patients with RVR, the rate of relapse did not differ between patients with creatinine levels of <0.7 g/dl and ≥0.7 g/dl and were 12% and 12%, respectively. Among patients with cEVR, the rate of relapse was higher in patients with creatinine levels of <0.7 g/dl compared to those with creatinine levels of ≥0.7 g/dl and were 39% versus 23%, respectively, for patients <60 years who received <3.0 g/kg of body weight of RBV, 19% versus 14% for patients <60 years who received ≥3.0 g/kg of body weight of RBV, 58% versus 41% for patients ≥60 years who received <3.0 g/kg of body weight of RBV, and 42% versus 26% for patients ≥60 years who received ≥3.0 g/kg of body weight of RBV.

Effect of age and total RBV dose on relapse among patients with cEVR

The effect of total RBV dose on relapse was analysed among patients with cEVR in a combined group of

Figure 1. The decision-tree model of relapse among patients with rapid virological response or complete early virological response

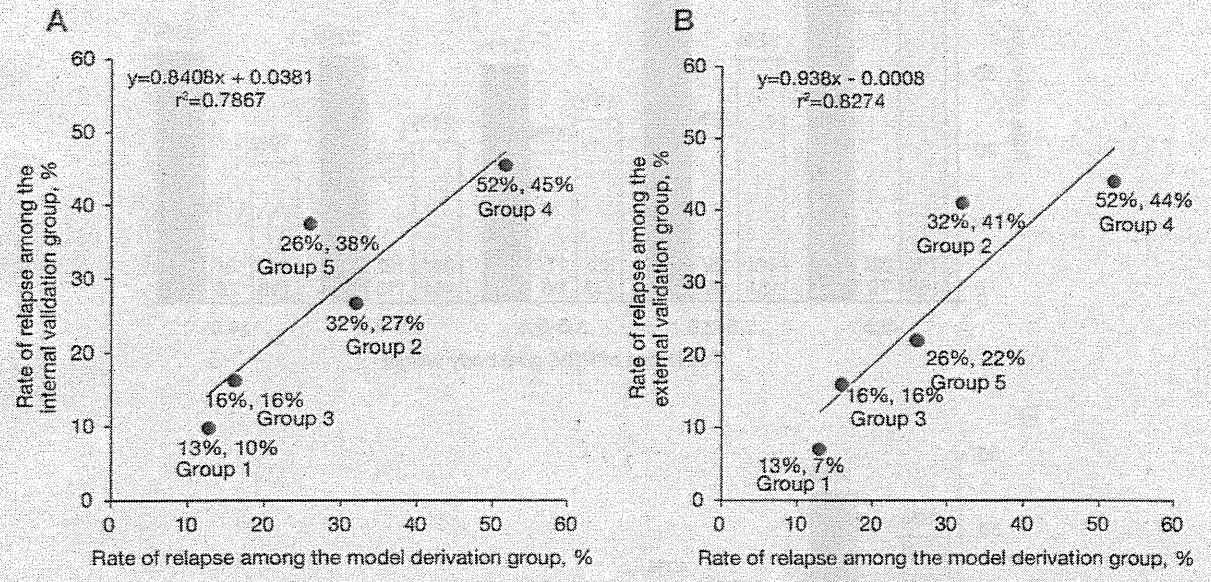


Boxes indicate the factors used for splitting and the cutoff values for the split. Pie charts indicate the rate of relapse for each group of patients after splitting. Terminal groups of patients discriminated by the analysis are numbered from 1 to 5. The rate of relapse was higher than average (>26%) in subgroups 2 and 4, where total ribavirin (RBV) dose was <3 g/kg of body weight.

model derivation and internal validation (n=718). The relapse rate decreased with an increase in RBV dose (Figure 3A). When patients were stratified into two groups according to age, the relapse rate decreased with an increase in RBV dose in patients <60 years. The relapse rate was lowest (11%) in patients <60 years who received ≥4.0 g/kg of body weight of RBV. By contrast, among patients ≥60 years, the relapse rate decreased with an increase in RBV dose up to 3.0 g/kg of body weight, but remained relatively stable despite a further increase in the RBV dose beyond 3.0 g/kg of body weight. The rate of relapse was 31% to 33% in patients who received ≥3.0 g/kg of body weight. Patients ≥60 years had higher relapse rate compared with patients <60 years after stratification by RBV dose (P=0.044 for RBV <2.5 g/kg, P=0.009 for RBV 2.5–2.9 g/kg, P=0.150 for RBV 3.0–3.4 g/kg, P=0.036 for RBV 3.5–3.9 g/kg and P=0.006 for RBV ≥4.0 g/kg).

To exclude the effect of the duration of therapy, patients who received 42–54 weeks of therapy were selected (n=544). Again, the relapse rate decreased with an increase in RBV dose in patients <60 years but remained stable despite a further increase in the RBV dose beyond 3.0 g/kg of body weight in patients ≥60 years (Figure 3B); in addition, patients ≥60 years had a higher relapse rate compared with younger patients after stratification by

Figure 2. Internal and external validation of the decision-tree model: subgroup-stratified comparison of the rate of relapse between the model derivation and validation groups



Each patient in the internal and external validation population was allocated to groups 1 to 5 following the flowchart of the decision tree. The rates of relapse were then calculated for each group and a graph was plotted. The rate of relapse in the (A) internal and (B) external validation groups are shown. The rates of relapse are shown as percentages below data points: the value on the left is from the model derivation group and on the right is from the validation group. The rates of relapse in each group of patients correlated closely between the model derivation group and the validation group (correlation coefficient: r^2=0.79 and 0.83, respectively).

Table 2. Multivariable analysis of factors associated with relapse among patients with RVR/cEVR

Factor	OR	95% CI	P-value
No-RVR	4.07	2.57–6.43	<0.0001
Total RBV dose <3.0 g/kg body weight	2.19	1.58–3.03	<0.0001
Creatinine <0.7 g/dl	1.67	1.22–2.29	0.001
Age ≥60 years	2.37	1.73–3.24	<0.0001

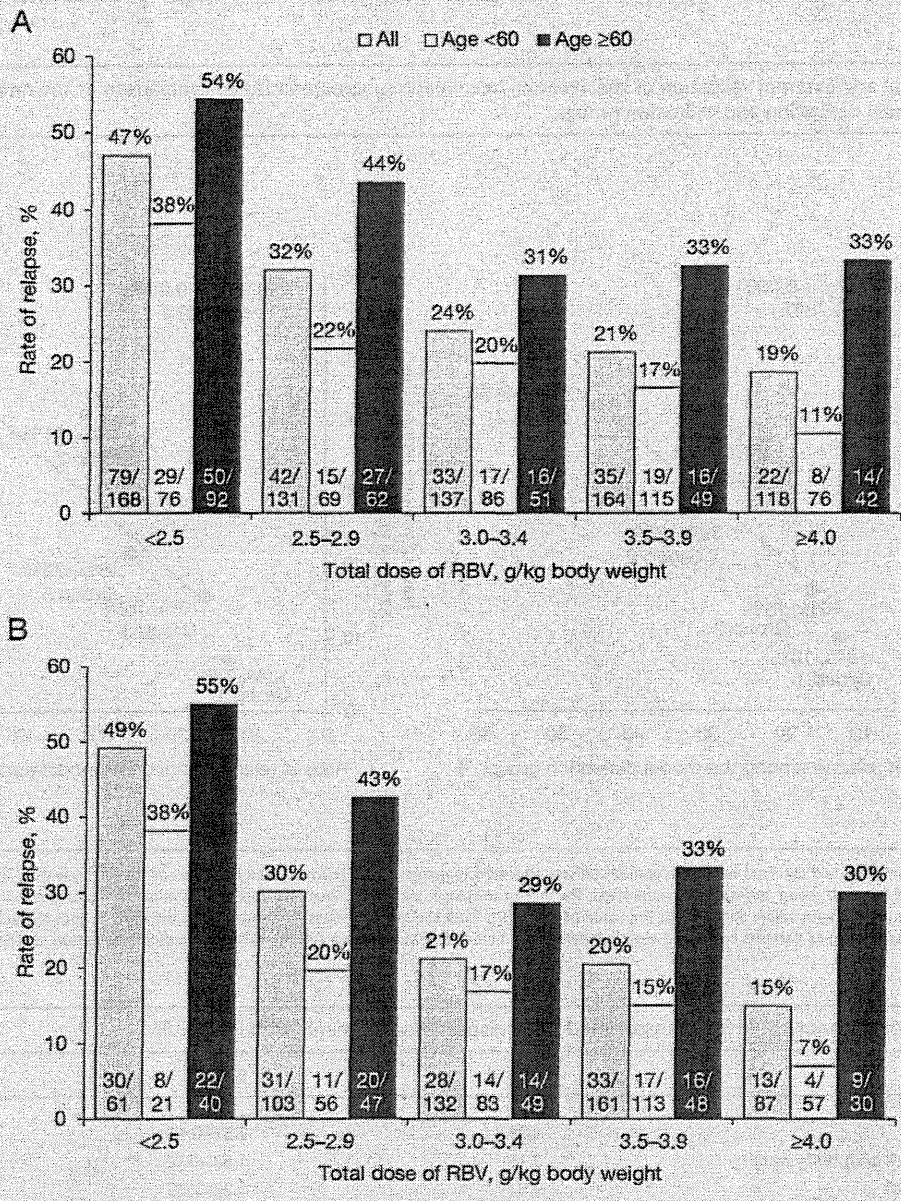
cEVR, complete early virological response (HCV-RNA-positive at week 4, but negative at week 12); RBV, ribavirin; RVR, rapid virological response (HCV-RNA-negative at week 4).

RBV dose (P=0.283 for RBV <2.5 g/kg, P=0.017 for RBV 2.5–2.9 g/kg, P=0.127 for RBV 3.0–3.4 g/kg, P=0.011 for RBV 3.5–3.9 g/kg and P=0.009 for RBV ≥4.0 g/kg).

Total dose of RBV was associated with relapse independently of PEG-IFN dose. The cutoff value of 58 zμg/kg of PEG-IFN was selected, which corresponds to the 80% of 1.5 μg/kg dose for 48 weeks. In patients who received <58 μg/kg of body weight of PEG-IFN, the rate of relapse for patients who received ≥3.0 g/kg or <3.0 g/kg of body weight of RBV was 24% and 42%, respectively. In patients who received ≥58 μg/kg of body weight of PEG-IFN, the rate of relapse for patients who received ≥3.0 g/kg or <3.0 g/kg of body weight of RBV was 21% and 38%, respectively.

The data mining analysis procedure did not select further split variables among RVR patients. However,

Figure 3. Correlation between the rate of relapse and total RBV dose among patients with cEVR after stratification by age



Association between the total ribavirin (RBV) dose and the rate of relapse among patients with complete early virological response (cEVR) is shown. (A) Higher dose of RBV was associated with reduced rate of relapse. (B) These associations were also confirmed in selected patients who received 42–54 weeks of therapy.

when analysed separately, the rate of relapse was also associated with age and total RBV dose among patients with RVR. The rate of relapse for patients who received ≥3.0 g/kg or <3.0 g/kg of body weight of RBV was 5% and 14%, respectively. The rate of relapse for patients <60 and ≥60 years was 9% and 18%, respectively. Collectively, the rate of relapse for patients <60 years who received ≥3.0 g/kg or <3.0 g/kg of body weight of RBV was 2% and 11%, respectively, whereas the rate of relapse for patients ≥60 years who received ≥3.0 g/kg or <3.0 g/kg of body weight of RBV was 12% and 20%, respectively.

Discussion

The result of the present study shows that older age and insufficient dose of RBV are significant and independent risk factors for relapse among patients with cEVR to PEG-IFN plus RBV. Older patients (≥60 years) who received a total RBV dose <3.0 g/kg of body weight had the highest risk of relapse (52%), whereas younger patients who received a total RBV dose ≥3.0 g/kg of body weight had the lowest risk of relapse (16%). The rate of relapse decreased depending on the total RBV dose in younger patients, but remained stable in older patients despite a further increase in the RBV dose beyond 3.0 g/kg of body weight. These findings imply that the target dose of total RBV can be set at 3.0 g/kg of body weight in patients who achieved cEVR, and further increase in RBV dose up to 4.0 g/kg of body weight or greater may be recommended in patients <60 years.

The associations between the drug adherence and virological response had been reported with inconsistent results. In an earlier study, patients who received >80% of the planned dose of PEG-IFN plus RBV for >80% of the planned duration of therapy had a higher rate of SVR compared to those who received a lesser dose (51% versus 34%) [31]. Consistent results were obtained in a study reporting that patients who received >80% of the planned dose of PEG-IFN and RBV within the first 12 weeks of therapy had a higher rate of EVR compared with those who received a lesser dose of both drugs (80% versus 33%) [4]. By contrast, a large-scale multicentre study showed that reducing the PEG-IFN dose during the first 20 weeks reduced SVR; however, reducing RBV did not affect SVR as long as RBV was not prematurely discontinued [32]. The reason for these inconsistencies is unclear. One reason may be the differences in the backgrounds of patients enrolled in the study, and hence the last study was limited to patients with advanced fibrosis and prior non-responders to PEG-IFN therapy. Because the probability of SVR is affected by virological response and relapse after response, the effect of drug dosing should be analysed separately with respect to these two factors.

In the present study, we focused on factors predictive of relapse after early virological response. According to the decision tree model, relapse was less likely in patients with RVR compared with cEVR. Among patients with cEVR, older patients (≥60 years) had a higher risk of relapse compared to younger patients (41% versus 22%). In addition, our results emphasized the effect of RBV dose for the prevention of relapse. In our study, a total RBV dose of ≥3.0 g/kg of body weight was repeatedly associated with a suppressed rate of relapse in the model derivation and validation groups. The rate of relapse in patients <60 years who received an RBV dose of <3.0 versus ≥3.0 g/kg of body weight in the model derivation, internal validation and external validation groups were 32% versus 16%, 27% versus 16%, and 41% versus 16%, respectively. The rate of relapse in patients ≥60 years who received an RBV dose of <3.0 versus ≥3.0 g/kg of body weight in the model derivation, internal validation and external validation groups were 52% versus 26%, 45% versus 38%, and 44% versus 22%, respectively. It has been reported that the rate of relapse is suppressed in 48 weeks of IFN plus RBV combination therapy compared to IFN monotherapy, indicating that RBV contributes to the increase in SVR by reducing relapse [2,3]. Another study, focused on the associations between the drug dose reduction and relapse in patients with virological response, found that maintaining RBV dose ≥12 mg/kg/day during 48 weeks of treatment, which can be translated into a total dose of 4.0 g/kg of body weight, suppressed relapse [33]. Results of the present study are in accordance with this report.

The importance of drug dosing on reduction in relapse is also supported by the findings that extending therapy from 48 to 72 weeks in patients with delayed virological response improved SVR rates by reducing relapse [9–13]. Apart from these clinical studies, in the real world of clinical practice, duration of therapy is extended – even in patients with cEVR – at the physician's discretion. The relationship between duration of therapy or RBV dose, and relapse among patients with cEVR and treated with various lengths of therapy has not been examined. In the combined group of our study, extending the duration of therapy was not associated with a reduction in relapse rate. Rather, the rate of relapse decreased depending on the total RBV dose. These findings suggest that acquiring a sufficient total RBV dose, either within 48 weeks or by extending the duration of therapy, is essential to prevent relapse among patients with cEVR. The limitation of the present study was that the mean duration of therapy was only 56.3 weeks in patients whose duration of therapy was extended beyond 48 weeks. It is probable that extended duration of therapy was not long enough for the prevention of relapse. Further studies with

longer durations of therapy are necessary to confirm the effect of extended duration of therapy on reduction of relapse among patients with cEVR.

Previous reports did not consider the effects of age in setting the optimal dose of RBV. In the present study, the relapse rate decreased with an increase in RBV dose from <2.5 to 3.0–3.5 g/kg of body weight, but remained relatively stable despite a further increase in the RBV dose in older patients. Thus, a total RBV dose ≥3.0 g/kg of body weight should be the target dose for patients ≥60 years with cEVR. By contrast, ≥3.0 g/kg of body weight of RBV was associated with lower risk of relapse in patients <60 with cEVR (16% versus 32%), and a further increase in RBV dose led to a more profound reduction in relapse rates, as low as 11% in patients who received ≥4.0 g/kg of body weight. Thus, a total dose of ≥4.0 g/kg of body weight or even greater should be the target dose in patients <60 years.

In the near future, more potent therapies, such as direct antiviral agents [34,35], may become available. These drugs require RBV and PEG-IFN in combination. However, not all patients may be able to tolerate this triple combination therapy due to adverse drug reactions, such as severe anaemia or skin eruption. In particular, it may be difficult to administer a full dose of triple drugs to older patients. Thus, personalizing the PEG-IFN and RBV combination therapy based on this model may be beneficial to patients who were intolerant to triple combination therapy.

In the present study creatinine was an independent predictor of relapse by multivariable logistic regression analysis. However creatinine was not selected as a splitting variable in decision tree, which may be due to the unique property of data mining analysis. In data mining analysis, limitation is imposed to stop the analysis when the number of patients is <20. This limitation is used to avoid dividing patients into too small subgroups which lead to the generation of rules that only apply to the model derivation population and not reproduced when applied to other populations. This phenomenon is called the over-fitting of the model. Due to this limitation, the variables selected in the data mining analysis are not necessarily identical to the variables that are significant by ordinary multivariable analysis. In a separate analysis, lower level of creatinine was associated with higher rate of relapse in each subgroup of patients with cEVR. The reason for this association is not clear, but lower creatinine level may be related to more efficient clearance of RBV leading to lower serum level of RBV. Further research is needed to confirm this speculation.

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

In conclusion, we built a decision tree model for the prediction of relapse among patients with EVR to PEG-IFN plus RBV. The result of the present study shows that older age and insufficient dose of RBV are significant and independent risk factors for relapse. The target dose of total RBV can be set at 3.0 g/kg of body weight in patients who achieved cEVR. A further increase in RBV dose up to 4.0 g/kg of body weight may be warranted in patients <60 years.

Acknowledgements

This study was supported by a Grant-in-Aid from the Ministry of Health, Labor and Welfare, Japan (H20-kanen-006).

Disclosure statement

The authors declare no competing interests.

References

1. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology* 2004; 39:1147–1171.
2. Fried MW, Shiffman ML, Reddy KR, *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347:975–982.
3. Manns MP, McHutchison JG, Gordon SC, *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; 358:958–965.
4. Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 2003; 38:645–652.
5. Lee SS, Ferenci P. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther* 2008; 13 Suppl 1:9–16.
6. Namiki I, Nishiguchi S, Hino K, *et al.* Management of hepatitis C: report of the consensus meeting at the 45th annual meeting of the Japan Society of Hepatology (2009). *Hepatol Res* 2010; 40:347–368.
7. Jensen DM, Morgan TR, Marcellin P, *et al.* Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. *Hepatology* 2006; 43:954–960.
8. Yu ML, Dai CY, Huang JF, *et al.* Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: a randomized trial. *Hepatology* 2008; 47:1884–1893.
9. Berg T, von Wagner M, Nasser S, *et al.* Extended treatment duration for hepatitis C virus type 1: comparing 48 versus 72 weeks of peginterferon-alfa-2a plus ribavirin. *Gastroenterology* 2006; 130:1086–1097.

10. Sánchez-Tapias JM, Diago M, Escartin P, *et al.* Peginterferon-alfa2a plus ribavirin for 48 versus 72 weeks in patients with detectable hepatitis C virus RNA at week 4 of treatment. *Gastroenterology* 2006; 131:451–460.
11. Ferenci P, Laferl H, Scherzer TM, *et al.* Peginterferon alfa-2a/ribavirin for 48 or 72 weeks in hepatitis C genotypes 1 and 4 patients with slow virologic response. *Gastroenterology* 2010; 138:503–512.e1.
12. Buti M, Lurie Y, Zakharova NG, *et al.* Randomized trial of peginterferon alfa-2b and ribavirin for 48 or 72 weeks in patients with hepatitis C virus genotype 1 and slow virologic response. *Hepatology* 2010; 52:1201–1207.
13. Pearlman BL, Ehleben C, Saifee S. Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis c genotype 1-infected slow responders. *Hepatology* 2007; 46:1688–1694.
14. Tanaka Y, Nishida N, Sugiyama M, *et al.* Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 2009; 41:1105–1109.
15. Suppiah V, Moldovan M, Ahlenstiel G, *et al.* IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 2009; 41:1100–1104.
16. Ge D, Fellay J, Thompson AJ, *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009; 461:399–401.
17. Kurosaki M, Tanaka Y, Nishida N, *et al.* Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. *J Hepatol* 2011; 54:439–448.
18. Breiman LJH, Friedman RA, Olshen CJ, Stone CM. *Classification and regression trees*. 1980. Belmont, CA: Wadsworth.
19. Garzotto M, Park Y, Mongoue-Tchokote S, *et al.* Recursive partitioning for risk stratification in men undergoing repeat prostate biopsies. *Cancer* 2005; 104:1911–1917.
20. Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 2002; 12:243–248.
21. Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. *Surgery* 2002; 132:589–604.
22. Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the German dermatological society. *J Clin Oncol* 2004; 22:3660–3667.
23. Valera VA, Walter BA, Yokoyama N, *et al.* Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol* 2007; 14:34–40.
24. Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 2005; 11:5440–5443.
25. Baquerizo A, Anselmo D, Shackleton C, *et al.* Phosphorus ans an early predictive factor in patients with acute liver failure. *Transplantation* 2003; 75:2007–2014.
26. Kurosaki M, Matsunaga K, Hirayama I, *et al.* A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatol Res* 2010; 40:251–260.
27. Kurosaki M, Sakamoto N, Iwasaki M, *et al.* Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. *J Gastroenterol* 2011; 46:401–409.
28. Kurosaki M, Sakamoto N, Iwasaki M, *et al.* 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. *J Med Virol* 2011; 83:445–452.
29. LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res* 1995; 75:113–124.
30. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989; 8:539–550.
31. McHutchison JG, Manns M, Patel K, *et al.* Adherence to combination therapy enhances sustained response in genotype-1-infected patients with chronic hepatitis C. *Gastroenterology* 2002; 123:1061–1069.
32. Shiffman ML, Ghany MG, Morgan TR, *et al.* Impact of reducing peginterferon alfa-2a and ribavirin dose during retreatment in patients with chronic hepatitis C. *Gastroenterology* 2007; 132:103–112.
33. Hiramatsu N, Oze T, Yakushijin T, *et al.* Ribavirin dose reduction raises relapse rate dose-dependently in genotype 1 patients with hepatitis C responding to pegylated interferon alpha-2b plus ribavirin. *J Viral Hepat* 2009; 16:586–594.
34. Hézode C, Forestier N, Dusheiko G, *et al.* Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 2009; 360:1839–1850.
35. McHutchison JG, Everson GT, Gordon SC, *et al.* Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* 2009; 360:1827–1838.

Accepted 18 April 2011; published online 21 October 2011

Hepatol Int
DOI 10.1007/s12072-011-9307-6

ORIGINAL ARTICLE

Analysis of viral amino acids sequences and the IL28B SNP influencing the development of hepatocellular carcinoma in chronic hepatitis C

Mika Miura · Shinya Maekawa · Makoto Kadokura · Ryota Sueki · Kazuki Komase · Hiroko Shindo · Takako Ohmori · Asuka Kanayama · Kuniaki Shindo · Fumitake Amemiya · Yasuhiro Nakayama · Takatoshi Kitamura · Tomoyoshi Uetake · Taisuke Inoue · Minoru Sakamoto · Shunichi Okada · Nobuyuki Enomoto

Received: 23 April 2011 / Accepted: 22 July 2011

Abstract
Background and aims The association between hepatitis C virus (HCV) sequences with interleukin 28B (IL28B) single-nucleotide polymorphism (SNP) in the development of hepatocellular carcinoma (HCC) has not been well clarified.
Methods Complete HCV open-reading frame sequences were determined in 20 patients developing HCC and 23 non-HCC patients with HCV-1b infection in two distant time points. An additional 230 patients were studied cross-sectionally for core and NS5A sequences with HCC development. Among them, 98 patients with available samples were investigated for changes in viral core sequences over time. Finally, IL28B SNPs and HCC development were investigated in 228 patients.
Results During observation period (HCC for 10.8 years, and non-HCC for 11.1 years), changes in core a.a. 70 and three amino acid positions in NS5A were characteristics of the patients developing HCC. In 230 patients, Q (glutamine) or H (histidine) to R (arginine) ratio at core a.a. 70 was significantly higher in the HCC group (HCC group 43:22 vs. non-HCC group 66:99, $p = 0.001$). A change in core R70Q was observed over time in 11 patients associated with a decrease in platelets ($p = 0.005$) and albumin ($p = 0.005$), while a Q70R change was observed in 4 patients without associated changes in platelets (nonsignificant) and albumin (nonsignificant). IL28B SNP showed significant correlation with the core a.a. 70 residue. There was no evident link between IL28B SNPs and the occurrence of HCC.
Conclusions Hepatitis C virus core a.a. 70 residue is associated with liver disease progression and is independent factor for HCC development in genotype-1b infection. IL28B SNPs are related to core a.a. 70 residue, but not to HCC. The functional relevance of core a.a. 70 residue in hepatitis C pathogenesis should be further investigated.

Keywords HCV · HCC · Core · IL28B

Introduction

Hepatitis C virus (HCV) infection is a major risk factor for hepatocellular carcinoma (HCC). Chronic HCV infection can result in liver cirrhosis (LC) and HCC over the course of 20–30 years [1]. However, the rate of progression is variable; some patients remain for a long time with persistently normal ALT values, while others progress rapidly to LC and HCC.

Viral factors, host factors, and their interplay appear to play an important role in determining the progression of chronic hepatitis C to LC and HCC. In terms of viral factors, most previous clinical studies have focused on searching for HCV regions correlated with the response to interferon (IFN)-based therapy. In those analyses, correlation between amino acid substitutions and treatment response have been reported for the IFN sensitivity

M. Miura and S. Maekawa have contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s12072-011-9307-6) contains supplementary material, which is available to authorized users.

M. Miura · S. Maekawa (✉) · M. Kadokura · R. Sueki · K. Komase · H. Shindo · T. Ohmori · A. Kanayama · K. Shindo · F. Amemiya · Y. Nakayama · T. Kitamura · T. Uetake · T. Inoue · M. Sakamoto · S. Okada · N. Enomoto
First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimokato, Chuo, Yamanashi 409-3898, Japan
e-mail: maekawa@yamanashi.ac.jp

Published online: 17 August 2011

determining region in nonstructural (NS)5A [2], a.a. 70 and 91 in core [3], the PKR/eIF-2alpha phosphorylation homology domain (PePHD) in envelope (E)2 [4], and the IFN–ribavirin resistance determining region (IRRDR) in NS5A [5].

Regarding the viral factors related to disease progression, among the various HCV proteins, the core protein has been thought widely to contribute because it has been shown experimentally to affect multiple cellular functions, in addition to the evidence from clinical studies [6–10]. Core protein modifies cellular apoptosis, oncogenic signaling, reactive oxygen species formation, lipid metabolism, transcriptional activation, transformation, and immune reactivity. Core protein has oncogenic potential in transgenic mice [11]. In contrast, fewer clinical studies to date have systematically investigated the correlation between the variability of HCV regions and disease progression. However, some of those limited clinical studies reported a correlation between amino acid substitutions in core or NS5A with disease progression [12–15]. Despite those reports, few studies to date support the correlation. Moreover, it is unclear whether those viral sequences change during disease progression or how the disease activity is modified by those viral sequences in the long course of chronic hepatitis.

On the other hand, regarding host factors, recent reports disclosed a significant correlation between polymorphisms in the IL28B gene and responses to pegylated-IFN plus ribavirin therapy for HCV patients [16–19]. This single-nucleotide polymorphism (SNP) also showed significant correlation with natural HCV clearance [20]. However, it remains unknown whether the IL28B SNP is related to disease progression or the development of HCC.

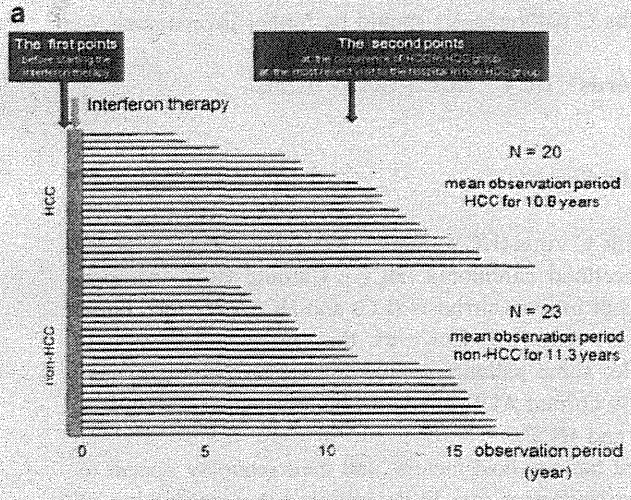
In this study, we first undertook the analysis to identify the viral regions related to disease progression and HCC development through the analysis of complete HCV open-reading frame (ORF) sequences. Because some regions in HCV core and NS5A showed characteristic changes over time in patients developing HCC during the observation period, we proceeded further to analyze the contribution of those regions to disease progression, in association with time and with the IL28B SNP.

Patients and methods

Patients

This study is based on the analysis of two groups of patients, 43 in Group 1 and 230 patients in Group 2.

In the first part, we tried to characterize and extract viral sequences specific to disease progression through the analysis of complete HCV ORFs (Fig. 1a). In particular, we focused our investigation on the changes in viral sequences over time in association with disease progression by comparing HCV sequences of two sufficiently distant time points. With this aim, we determined to investigate patients with a history of IFN therapy, because those patients often were followed long-term with preservation of old and recent sera. However, we excluded sustained



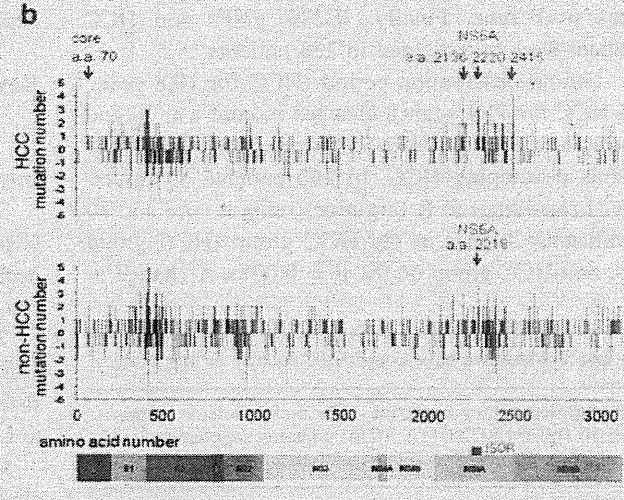


Fig. 1 **a** A total of 43 patients were analyzed for complete HCV ORF sequences. They were all non-responders to the previous IFN therapy. Twenty patients developed HCC during the observation period, while the 23 patients did not. HCV ORF sequences were determined for the paired samples and the predicted amino acid changes were compared in each patient. **b** Specific HCV amino acid changes associated with disease progression was evaluated by the analysis of the full-length viral ORF during the observation period for each patient according to the following rules: *1* +1 point for consensus to non-consensus, *2* −1 point for non-consensus to consensus, *3* 0 point for non-consensus to non-consensus. These points were added together and are shown for HCC and non-HCC patients. Patient with later HCC development (*upper panel*). Non-HCC patients (*lower panel*)

virologic response (SVR) patients because we thought that the viral clearance leads to improvement of the liver disease and, therefore, viral regions influencing the IFN response would be extracted as affecting the course of disease. Between March 1992 and April 2004, 273 consecutive patients with HCV-1b infection were given IFN monotherapy at Yamanashi University Hospital, and 133 were followed long-term. A total of 65 patients showed SVR, while 68 showed non-SVR. Among these 68 non-SVRs, 43 patients were included in the study because laboratory data and sera were available from the two distant time points (Group 1). Twenty patients developed HCC during the observation period, while the remaining 23 did not. Regarding the sera, the first time point for both groups was before starting IFN therapy, while the second points were at the occurrence of HCC for the HCC group and at the most recent visit to the hospital for the non-HCC group (Fig. 1a).

An additional 230 HCV-1b patients were recruited (Group 2) as the second study group. They were mainly outpatients at Yamanashi University Hospital and were selected randomly from those with stored sera at the time of disease diagnosis. Sixty-five had HCC, while the remaining 165 did not. They all were positive for HCV RNA at the time of study, although 74 patients had a history of IFN therapy. Parts of the core and NS5A sequences were determined at HCC onset in 65 HCC patients and at the most recent visit to the hospital in 165 non-HCC patients. Because historical sera around 10 years before were also available for 55 of these 230 patients, HCV sequence analysis was also performed for core and NS5A at those previous time points in those patients.

From these two study groups, 228 patients (68 HCCs and 160 non-HCCs) with available genomic DNAs were examined to determine the IL28B SNP.

All the patients studied all fulfilled following criteria: (1) Negative for hepatitis B surface antigen. (2) No other forms of hepatitis, such as primary biliary cirrhosis, autoimmune liver disease, or alcoholic liver disease. (3) Free of coinfection with human immunodeficiency virus. (4) A signed consent was obtained for the study protocol that had been approved by Human Ethics Review Committee of Yamanashi University Hospital.

Complete and partial HCV ORF sequence determination by direct sequencing from sera

HCV RNA extraction, complementary DNA synthesis and amplification by two-step nested PCR from serum samples were done using the specific primers for full HCV ORF or partial viral regions as described previously [15]. PCR amplicons were sequenced directly by Big Dye Terminator Version 3.1 (ABI, Tokyo, Japan) with universal M13 forward and reverse primers using an ABI prism 3130 sequencer (ABI). Generated sequence files were assembled using Vector NTI software (Invitrogen, Tokyo, Japan) and base-calling errors were corrected following inspection of the chromatogram.

IL28B SNP analysis

Human genomic DNA was extracted from peripheral blood using a blood DNA extraction kit (QIAGEN, Tokyo, Japan) according to the manufacturer's protocol. The allele typing of each DNA sample was performed by real-time PCR with a model 7500 (ABI) using FAM-labeled SNP primer for the locus rs8099917 (ABI).

Statistical analysis

Statistical differences in the parameters, including all available patients' demographic, biochemic, hematologic, and virologic data, were determined between different groups of patients by Student's t test for numerical variables and Fisher's exact probability test for categorical variables. Odds ratios and their 95% confidence intervals were used to quantify the level of association. All p values of <0.05 by the two-tailed test were considered significant throughout. Multiple logistic regression analyses were used to identify the independent variables influencing core a.a. 70 residue and HCC development. Because most variables used for the analyses were generally considered to correlate with the disease progression, we entered all the variables into the multiple logistic regression analysis even if some of them did not reach significant differences in individual univariate analysis.

Results

Comparing complete HCV amino acid sequences between patients with and without HCC

The clinical characteristics of the 43 patients (Group 1) analyzed for HCV ORF changes over time are shown in Table 1. At the start of observation, clinical characteristics did not differ significantly between the HCC group and the non-HCC group. The mean observation period was comparable between the two groups and was 10.8 years for the HCC group and 11.3 years for non-HCC group ($p = 0.745$). On the other hand, platelets ($p < 0.001$), albumin ($p < 0.001$), and AFP ($p = 0.001$) became significantly lower or higher in the HCC group at the end of observation (Table 1).

We proceeded to investigate viral amino acid changes during the course of disease in each patient to determine