Original article

Age and total ribavirin dose are independent predictors of relapse after interferon therapy in chronic hepatitis C revealed by data mining analysis

Masayuki Kurosaki1, Naoki Hiramatsu2, Minoru Sakamoto3, Yoshiyuki Suzuki4, Manabu Iwasaki5, Akihiro Tamori6, Kentaro Matsuura7, Sei Kakinuma8, Fuminaka Sugauchi9, Naoya Sakamoto8, Mina Nakagawa9, Hiroshi Yatsuhashi10, Namiki Izumi1*

1Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan
2Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan
3First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan
4Department of Hepatology, Toranomon Hospital, Tokyo, Japan
5Department of Computer and Information Science, Seikei University, Tokyo, Japan
6Department of Hepatology, Osaka City University Medical School, Osaka, Japan
7Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan
8Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan
9Department of Gastroenterology, Nagoya Koseiin Medical Welfare Center, Nagoya, Japan
10Clinical Research Center, National Nagasaki Medical Center, Nagasaki, Japan

*Corresponding author e-mail: nizumi@musashino.jrc.or.jp

Background: This study aimed to define factors associated with relapse among responders to pegylated interferon (PEG-IFN) plus ribavirin (RBV) therapy in chronic hepatitis C.

Methods: A cohort of genotype 1b chronic hepatitis C patients treated with PEG-IFN plus RBV and who had an undetectable HCV RNA by week 12 (n=951) were randomly assigned to model derivation (n=636) or internal validation (n=315) groups. An independent cohort (n=598) were used for an external validation. A decision tree model for relapse was explored using data mining analysis.

Results: The data mining analysis defined five subgroups of patients with variable rates of relapse ranging from 13% to 52%. The reproducibility of the model was confirmed by internal and external validations (r²=0.79 and 0.83, respectively). Patients with undetectable HCV RNA at week 4 had the lowest risk of relapse (13%), followed by patients <60 years with undetectable HCV RNA at week 5–12 who received ≥3.0 g/kg of body weight of RBV (16%). Older patients with a total RBV dose <3.0 g/kg had the highest risk of relapse (52%). Higher RBV dose beyond 3.0 g/kg was associated with further decrease of relapse rate among patients <60 years (up to 11%) but not among older patients whose relapse rate remained stable around 30%.

Conclusions: Data mining analysis revealed that time to HCV RNA negativity, age and total RBV dose was associated with relapse. To prevent relapse, ≥3.0 g/kg of RBV should be administered. Higher dose of RBV may be beneficial in patients <60 years.

Introduction

The currently recommended therapy for chronic hepatitis C is a combination of pegylated interferon (PEG-IFN) plus ribavirin (RBV) [1]. This therapy is effective in 50% of patients with HCV genotype 1b [2,3]. The most reliable predictor of sustained virological response (SVR) is the response during early weeks of therapy. A satisfactory response to therapy in the early weeks is associated with a high rate of SVR [4–8]. A basic concept of response-guided therapy is to modify the duration of therapy according to the time to HCV RNA negativity. Extended therapy may be given to patients with delayed virological response [9–13]. Modification of duration of therapy or drug dose may also be necessary in patients with early virological
response (EVR), because approximately 20% of these patients experience relapse after the completion of 48 weeks of therapy. Recent reports have revealed that single nucleotide polymorphisms located near the \( IL28B \) gene are strongly associated with SVR or a null response to PEG-IFN plus RBV therapy [14–16]. However, single nucleotide polymorphisms located near the \( IL28B \) gene are not associated with relapse after EVR [17]. Identification of risk factors for relapse among patients with virological response may lead to more individualized therapy and improved SVR rate.

Decision tree analysis, a core component of data mining analysis, is a method that explores data to develop predictive models [18]. This method has been originally used in business and recently in medical fields [19–25]. Decision tree analysis was successfully used to build a predictive model of EVR [26] and SVR to PEG-IFN plus RBV combination therapy in chronic hepatitis C [17,27,28]. The results of the analysis are presented as a tree structure, which is easy to understand and use in clinical practice. Patients can be allocated into subgroups by simply following the flowchart form of the decision tree [29].

In the present study, we used decision tree analysis to identify predictors of relapse among patients who achieved EVR to PEG-IFN plus RBV therapy, and to define a more individualized therapeutic strategy beyond response-guided therapy.

**Methods**

**Patients**

This is a multicentre retrospective cohort study involving Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Osaka University, Nagoya City University, Yamanashi University, Osaka City University, and their related hospitals. The inclusion criteria were chronic hepatitis C patients treated with PEG-IFN-\( \alpha \)2b plus RBV, genotype 1b, pretreatment HCV RNA titre >100 KIU/ml as confirmed by quantitative PCR; Cobas Amplicor HCV Monitor version 2.0; Roche Diagnostic Systems, Pleasanton, CA, USA), an undetectable HCV RNA level within week 12 after the start of therapy, no coinfection with HBV or HIV, and no other causes of liver disease. Patients were treated with PEG-IFN-\( \alpha \)2b (1.5 µg/kg) subcutaneously every week plus a daily weight-adjusted RBV dose (600 mg for patients weighing <60 kg, 800 mg for patients weighing 60–80 kg and 1,000 mg for patients weighing >80 kg). Dose reduction or discontinuation of PEG-IFN and RBV was considered based on the recommendations of the package inserts and the discretion of physicians at each university and hospital. The standard duration of therapy was set at 48 weeks, but extension of duration was allowed and implemented at the discretion of each physician. The duration of therapy was extended beyond 48 weeks in 118 patients (mean duration was 56.3 weeks, ranging from 49 to 72 weeks). Although the exact reason for the prolonged treatment in each case was not available, one reason may be that each physician tried to achieve high adherence of RBV by extending the duration of therapy. Another reason may be the late time point of HCV RNA negativity even within early virological response. Among 118 patients, time to HCV RNA negativity was between 9 to 12 weeks in 56% of patients.

A total of 951 patients fulfilled the study criteria. The baseline characteristics and representative laboratory test results are listed in Table 1. For analysis, patients were randomly assigned to either the model derivation (636 patients) or internal validation (315 patients) groups. There were no significant differences in the clinical backgrounds between these two groups. For external validation of the model, we collaborated with another multicentre study group consisting of 29 medical centres and hospitals belonging to the National

---

**Table 1. Background of study population**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>54.9 (10.8)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, n (%)</td>
</tr>
<tr>
<td></td>
<td>Female, n (%)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>23.2 (3.3)</td>
</tr>
<tr>
<td>Albumin, g/dl</td>
<td>4.1 (1.8)</td>
</tr>
<tr>
<td>Creatinine, mg/dl</td>
<td>0.7 (0.2)</td>
</tr>
<tr>
<td>AST, IU/l</td>
<td>60.6 (46.2)</td>
</tr>
<tr>
<td>ALT, IU/l</td>
<td>80.7 (77.2)</td>
</tr>
<tr>
<td>GGT, IU/l</td>
<td>52.0 (60.0)</td>
</tr>
<tr>
<td>White blood cell count, cells/µl</td>
<td>4,993 (1,363)</td>
</tr>
<tr>
<td>Haemoglobin, g/dl</td>
<td>15.9 (52.6)</td>
</tr>
<tr>
<td>Platelets, 10⁹/µl</td>
<td>174.4 (6.1)</td>
</tr>
<tr>
<td>HCV RNA, KIU/ml</td>
<td>1,655 (1,455)</td>
</tr>
<tr>
<td>Fibrosis stage</td>
<td>–</td>
</tr>
<tr>
<td>F1-2, n (%)</td>
<td>626 (66)</td>
</tr>
<tr>
<td>F3-4, n (%)</td>
<td>98 (10)</td>
</tr>
<tr>
<td>NA, n (%)</td>
<td>227 (24)</td>
</tr>
<tr>
<td>Time to HCV RNA negativity 4/8/12 weeks</td>
<td>–</td>
</tr>
<tr>
<td>4 Weeks, n (%)</td>
<td>233 (24)</td>
</tr>
<tr>
<td>8 Weeks, n (%)</td>
<td>386 (41)</td>
</tr>
<tr>
<td>12 Weeks, n (%)</td>
<td>332 (35)</td>
</tr>
<tr>
<td>Treatment duration, weeks</td>
<td>42 (13)</td>
</tr>
<tr>
<td>Total RBV dose, g/kg body weight</td>
<td>3.1 (1.3)</td>
</tr>
<tr>
<td>Total PEG-IFN dose, µg/kg body weight</td>
<td>62.5 (38.6)</td>
</tr>
<tr>
<td>Outcome</td>
<td>–</td>
</tr>
<tr>
<td>Relapse, n (%)</td>
<td>238 (25)</td>
</tr>
<tr>
<td>SVR, n (%)</td>
<td>713 (75)</td>
</tr>
</tbody>
</table>

Total n=951. Data are expressed as mean (sd) unless otherwise indicated. ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, \( \gamma \)-glutamyltransferase; NA, not available; PEG-IFN, pegylated interferon; RBV, ribavirin; SVR, sustained virological response.
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 (32%).

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 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 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 \(\geq 4.0\) g/kg of body weight of RBV. By contrast, among patients \(\geq 60\) 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 \(\geq 3.0\) g/kg of body weight.

Patients \(\geq 60\) 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 \(\geq 4.0\) 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 \(\geq 60\) years (Figure 3B); in addition, patients \(\geq 60\) years had a higher relapse rate compared with younger patients after stratification by

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

<table>
<thead>
<tr>
<th>Factor</th>
<th>OR</th>
<th>95% CI</th>
<th>(P)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-RVR</td>
<td>4.07</td>
<td>2.57–6.43</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total RBV dose &lt;3.0 g/kg body weight</td>
<td>2.19</td>
<td>1.58–3.03</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Creatinine &lt;0.7 g/dl</td>
<td>1.67</td>
<td>1.22–2.29</td>
<td>0.001</td>
</tr>
<tr>
<td>Age (\geq 60) years</td>
<td>2.37</td>
<td>1.73–3.24</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

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 μg/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 \( \geq 3.0 \) 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 \( \geq 60 \) years was 9% and 18%, respectively. Collectively, the rate of relapse for patients \(< 60 \) years who received \( \geq 3.0 \) g/kg or \(< 3.0 \) g/kg of body weight of RBV was 2% and 11%, respectively, whereas the rate of relapse for patients \( \geq 60 \) years who received \( \geq 3.0 \) 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 (\( \geq 60 \) 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 \( \geq 3.0 \) 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 \( \geq 80\% \) of the planned dose of PEG-IFN plus RBV for \( \geq 80\% \) 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 \( \geq 80\% \) 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 (\( \geq 60 \) 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 \( \geq 3.0 \) 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 \( < 3.0 \) g/kg of body weight was 9% and 18%, respectively. The rate of relapse in patients \( \geq 60 \) years who received an RBV dose of \( < 3.0 \) g/kg of body weight was 16%, and 41% versus 16%, respectively. The rate of relapse in patients \( \geq 60 \) years who received an RBV dose of \( \geq 3.0 \) g/kg of body weight was 2% and 11%, respectively, whereas the rate of relapse for patients \( \geq 60 \) years who received \( \geq 3.0 \) g/kg or \(< 3.0 \) g/kg of body weight of RBV was 12% and 20%, respectively.
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


