Table 3. Factors Associated with HCC After IFN Therapy

Risk Factor Value	Univariate Analysis		Multivariate Analysis	
	Risk Ratio (95% CI)	P Value	Risk Ratio (95% CI)	P Value
Age (by every 10 year)	2.2 (1.8-2.7)	<0.001	3.0 (1.9-4.8)	<0.001
Sex				
Female	1		1	
Male	1.2 (0.9-1.6)	0.2	2.0 (1.0-3.8)	0.04
BMI (by every 10 kg/m^2)	2.0 (1.2-1.3)	0.005	1.1 (0.4-3.5)	0.8
Fibrosis stage				
F0/F1/F2	1		1	
F3/F4	5.4 (3.9-7.5)	<0.001	2.5 (1.2-4.9)	0.01
Degree of steatosis				
<10%	1		1	
≥10%	4.5 (3.0-6.9)	<0.001	3.5 (1.9-6.4)	<0.001
Esophagogastric varices				
No	1		1	
Yes	3.3 (2.0-5.3)	<0.001	1.6 (0.6-4.4)	0.3
Virological response				
SVR	1		1	
Non-SVR	3.3 (2.1-5.2)	<0.001	2.6 (1.2-5.5)	0.001
Genotype				
Non-1	1		1	
1	1.7 (1.2-2.5)	0.006	1.0 (0.5-2.3)	0.9
Albumin (by every 1 g/dL)	0.2 (0.1-0.3)	<0.001	0.6 (0.2-2.2)	0.3
ALT (by every 100 IU/L)	1.0 (0.9-1.0)	0.8	0.4 (0.1-1.8)	0.6
AST (by every 100 IU/L)	1.2 (1.1-1.3)	0.001	1.1 (0.6-1.8)	0.8
γ-GTP (by every 100 IU/L)	1.3 (1.1-1.6)	0.009	0.6 (0.3-1.6)	0.3
ALP (by every 100 IU/L)	1.3 (1.2-1.5)	<0.001	0.6 (0.3-1.2)	0.2
Total bilirubin (by every 1 mg/dL)	1.6 (1.3-2.1)	<0.001	1.2 (0.6-2.7)	0.6
Total cholesterol (by every 100 mg/dL)	0.3 (0.2-0.6)	<0.001	0.2 (0.1-0.6)	0.006
Triglyceride (by every 100 mg/dL)	0.8 (0.5-1.1)	0.2	0.1 (0.02-1.1)	0.08
Fasting blood sugar (by every 100 mg/dL)	1.8 (1.5-2.2)	<0.001	1.1 (1.0-1.1)	0.04
WBC (by every 100/μL)	0.1 (0.03-0.3)	<0.001	0.1 (0.01-2.2)	0.2
RBC (by every 10^6/μL)	0.5 (0.4-0.7)	<0.001	1.8 (0.7-4.4)	0.2
Platelet counts (by every 10^5/μL)	0.3 (0.2-0.4)	<0.001	0.6 (0.3-1.5)	0.3
Baseline AFP (by every 10 ng/mL)	1.0 (0.9-1.1)	0.2	1.3 (1.0-1.7)	0.04
Post IFN AFP (by every 10 ng/mL)	1.2 (1.1-1.3)	<0.001	1.9 (1.5-2.4)	<0.001
HCV load (by every 100 KIU/mL)	1.0 (0.9-1.0)	0.4	1.0 (1.0-1.1)	0.06
IFN regimen				
IFN monotherapy	1		1	
IFN + RBV (24 W)	1.2 (0.8-1.8)	0.4	1.5 (0.7-3.2)	0.3
PEG-IFN monotherapy (48 W)	1.1 (0.6-1.9)	0.8	1.5 (0.4-5.5)	0.6
PEG-IFN + RBV	0.4 (0.2-0.9)	0.03	1.0 (0.3-3.1)	0.9

Risk ratios for development of HCC were calculated by Cox proportional hazards regression analysis. AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspirate aminotransferase; BMI, body mass index; γ-GTP, gamma-glutamyltranspeptidase; HCC, hepatocellular carcinoma; IFN, interferon; PEG, pegylated; RBC, red blood cell counts; RBV, ribavirin; SVR, sustained virological response; WBC, white blood cell count.

may be at a higher risk of developing HCC. Further study is necessary to confirm this association in a clinical situation. Because several developed countries are in the midst of a growing obesity epidemic, the risk related to obesity cannot be ignored in patients with chronic hepatitis C who are treated with interferon.

Several retrospective cohort studies have been conducted to evaluate the effect of interferon on the incidence of HCC among patients with chronic hepatitis C.[8-11] Our results, obtained from one of the largest cohort studies, confirm the efficacy of viral eradication in preventing HCC. In one study conducted in a Western population, no statistically significant reduc-tion was found in the development of HCC among patients with SVR compared with those without SVR (adjusted hazard ratio, 0.46; 95% CI, 0.12-1.70; P = 0.25).[12] Because relatively few occurrences of HCC were observed in this cohort, and the duration of follow-up was shorter, the differences in HCC development between patients with and without SVR might be less pronounced.

Interestingly, our results demonstrated that the risk for HCC remains even after achieving SVR in older patients, confirming the findings of previous studies conducted with a smaller number of patients.[22,23] The cumulative incidence of HCC during the first 5 years

Table 4. Factors Associated with Development of HCC After Achieving SVR

Risk Factor	Odds Ratio (95% CI)	P-value
Univariate analysis		
Age (by every 10 year)	3.2 (1.8-5.5)	<0.001
Sex		
Female	1	
Male	3.0 (1.0-8.8)	0.04
Fibrosis stage		
F0/F1/F2	1	
F3/F4	5.9 (2.5-14.0)	<0.001
Degree of steatosis		
<10%	1	
≥10%	5.5 (2.0-15.2)	0.001
BMI (by every 10 kg/m^2)	3.2 (0.8-12.6)	0.09
ALT (by every 10 IU/L)	0.9 (0.7-1.3)	0.7
AST (by every 10 IU/L)	1.1 (0.9-1.4)	0.3
Genotype		
Non-1	1	
1	1.2 (0.6-3.0)	0.5
HCV load (by every 100 KIU/mL)	0.9 (0.8-1.0)	0.2
IFN regimen		
IFN monotherapy	1	
IFN + RBV (24 W)	0.7 (0.2-2.3)	0.5
PEG-IFN monotherapy (48 W)	0.8 (0.2-3.6)	0.8
PEG-IFN + RBV	0.3 (0.03-2.0)	0.2
Multivariate analysis		
Age (by every 10 year)	2.7 (1.5-5.1)	0.002
Sex		
Female	1	
Male	4.1 (0.9-18.9)	0.06
Fibrosis stage		
F0/F1/F2	1	
F3/F4	2.6 (0.9-7.5)	0.08
Degree of steatosis		
<10%	1	
≥10%	5.6 (1.9-16.5)	0.002

Odds ratios for SVR were calculated by logistic regression analysis.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; HCV, hepatitis C virus; IFN, interferon; HCC, hepatocellular carcinoma; PEG, pegylated; RBV, ribavirin; SVR, sustained virological response.

after completion of interferon therapy was similar between SVR and non-SVR patients in the older age group, and the risk for HCC remained for 9 years after eradication of HCV in our patients. Therefore, HCC patients with SVR who have a risk factor should be screened for at least 5-10 years after the completion of interferon therapy.

It has been reported that coffee consumption has a protective effect against hepatocarcinogenesis[24,25] and liver disease progression in patients with chronic HCV infection.[26] Because we could not review coffee consumption in all the patients and fewer data were available in the previous literature as to whether a habitual change of reducing coffee consumption occurs in older patients, it is unclear whether increased risk for HCC in older patients is an effect of this habitual change in older patients. However, the majority (68%) of Japa-

nese patients who have HCV (n = 1058) drink less than 1 cup of coffee per day, and only 7.6% consume more than 3 cups of coffee per day.[27] Therefore, it is unlikely that a habitual change in older patients affects the increased risk for hepatocarcinogenesis in older patients.

Recently, it was reported that interferon therapy might be less effective in preventing HCC among patients with chronic hepatitis C who are positive for anti-HBc antibody,[28] but this finding is still controversial.[29,30] In the present study, anti-HBc was only detected in 4 of 22 patients in whom HCC developed after viral eradication, and age distribution was similar among anti-HBc-positive and anti-HBc-negative patients. Because no significant difference in mean age was found between anti-HBc-positive and anti-HBc-negative patients in the recent study conducted in Japan,[28] it is unlikely that previous exposure to hepatitis B virus or occult hepatitis B virus infection is responsible for the difference in risk for HCC between younger and elderly patients found in the present study.

In conclusion, aging has become one of the most important risk factors for HCC. Even after stratification by stage of fibrosis, the risk for HCC after antiviral treatment was significantly higher in older patients, and HCV eradication had a smaller effect on HCC-free survival in older patients. Patients with HCV should therefore be identified at an earlier age and antiviral treatment should be initiated. The present results have potentially important clinical implications for physicians that may influence their decisions about the treatment strategy in individual patients.

References

1. Parkin DM. Global cancer statistics in the year 2000. Lancet Oncol 2001;2:533-543.
2. Di Bisceglie AM. Hepatitis C and hepatocellular carcinoma. Semin Liver Dis 1995;15:64-69.
3. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. Lancet 2003;362:1907-1917.
4. Kiyosawa K, Sodeyama T, Tanaka E, Gigo Y, Yoshizawa K, Nakano Y, et al. Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. HEPATOLOGY 1990;12:671-675.
5. Tanaka Y, Hanada K, Mizokami M, Yao AE, Shih JW, Gojobori T, et al. A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. Proc Natl Acad Sci U S A 2002;99:15584-15589.
6. U.S. Census Bureau. Population Projections. U.S. Interim Projections by Age, Sex, Race, and Hispanic Origin: 2000-2050. http://www.census.gov/population/www/projections/usinterimproj/. Page last modified: September 14, 2009.
7. Yancik R. Population aging and cancer: a cross-national concern. Cancer J 2005;11:437-441.

8. Imai Y, Kawata S, Tamura S, Yabuuchi I, Noda S, Inada M, et al. Relation of interferon therapy and hepatocellular carcinoma in patients with chronic hepatitis C. Ann Intern Med 1998;129:94-99.

9. Ikeda K, Saitoh S, Arase Y, Chayama K, Suzuki Y, Kobayashi M, et al. Effect of interferon therapy on hepatocellular carcinogenesis in patients with chronic hepatitis type C: a long-term observation study of 1,643 patients using statistical bias correction with proportional hazard analysis. HEPATOLOGY 1999;29:1124-1130.

10. Yoshida H, Shiratori Y, Moriyama M, Arakawa Y, Ide T, Sata M, et al. Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. Ann Intern Med 1999;131:174-181.

11. Bruno S, Stroffolini T, Colombo M, Bollani S, Benvegnù L, Mazzella G, et al. Sustained virological response to interferon-alpha is associated with improved outcome in HCV-related cirrhosis: a retrospective study. HEPATOLOGY 2007;45:579-587.

12. Veldt BJ, Heathcote EJ, Wedemeyer H, Reichen J, Hofmann WP, Zeuzem S, et al. Sustained virologic response and clinical outcomes in patients with chronic hepatitis C and advanced fibrosis. Ann Intern Med 2007;147:677-684.

13. Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. HEPATOLOGY 1994;19:1513-1520.

14. Hamada H, Yatsuhashi H, Yano K, Daikoku M, Arisawa K, Inoue O, et al. Impact of aging on the development of hepatocellular carcinoma in patients with posttransfusion chronic hepatitis C. Cancer 2002;95:331-339.

15. Ohishi W, Kitamoto M, Aikata H, Kamada K, Kawakami Y, Ishihara H, et al. Impact of aging on the development of hepatocellular carcinoma in patients with hepatitis C virus infection in Japan. Scand J Gastroenterol 2003;38:894-900.

16. Miki D, Aikata H, Uka K, Saneto H, Kawaoka T, Azakami T, et al. Clinicopathological features of elderly patients with hepatitis C virus-related hepatocellular carcinoma. J Gastroenterol 2008;43:550-557.

17. Kumar D, Farrell GC, Fung C, George J. Hepatitis C virus genotype 3 is cytopathic to hepatocytes: reversal of hepatic steatosis after sustained therapeutic response. HEPATOLOGY 2002;36:1266-1272.

18. Monto A, Alonzo J, Watson JJ, Grunfeld C, Wright TL. Steatosis in chronic hepatitis C: relative contributions of obesity, diabetes mellitus, and alcohol. HEPATOLOGY 2002;36:729-736.

19. Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, et al. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nat Med 1998;4:1065-1067.

20. Koike K, Moriya K. Metabolic aspects of hepatitis C viral infection: steatohepatitis resembling but distinct from NASH. J Gastroenterol 2005;40:329-336.

21. Koike K, Moriya K, Matsuura Y. Animal models for hepatitis C and related liver disease. Hepatol Res 2010;40:69-83.

22. Tokita H, Fukui H, Tanaka A, Kamitsukasa H, Yagura M, Harada H, et al. Risk factors for the development of hepatocellular carcinoma among patients with chronic hepatitis C who achieved a sustained virological response to interferon therapy. J Gastroenterol Hepatol 2005;20: 752-758.

23. Toyoda H, Kumada T, Tokuda A, Horiguchi Y, Nakano H, Honda T, et al. Yon-Ken HCV-HCC Follow-up Study Group. Long-term follow-up of sustained responders to interferon therapy, in patients with chronic hepatitis C. J Viral Hepat 2000;7:414-419.

24. Larsson SC, Wolk A. Coffee consumption and risk of liver cancer: a metaanalysis. Gastroenterology 2007;132:1740-1745.

25. Bravi F, Bosetti C, Tavani A, Bagnardi V, Gallus S, Negri E, et al. Coffee drinking and hepatocellular carcinoma risk: a meta-analysis. HEPATOLOGY 2007;46:430-435.

26. Freedman ND, Everhart JE, Lindsay KL, Ghany MG, Curto TM, Shiffman ML, et al. Coffee intake is associated with lower rates of liver disease progression in chronic hepatitis C. HEPATOLOGY 2009;50: 1360-1369.

27. Inoue M, Kurahashi N, Iwasaki M, Shimazu T, Tanaka Y, Mizokami M, et al. Effect of coffee and green tea consumption on the risk of liver cancer: cohort analysis by hepatitis virus infection status. Cancer Epidemiol Biomarkers Prev 2009;18:1746-1753.

28. Ikeda K, Marusawa H, Osaki Y, Nakamura T, Kitajima N, Yamashita Y, et al. Antibody to hepatitis B core antigen and risk for hepatitis C-related hepatocellular carcinoma: a prospective study. Ann Intern Med 2007 1;146:649-656.

29. Stroffolini T, Almasio PL, Persico M, Bollani S, Benvegnù L, Di Costanzo G, et al. Lack of correlation between serum anti-HBcore detectability and hepatocellular carcinoma in patients with HCV-related cirrhosis. Am J Gastroenterol 2008;103:1966-1972.

30. Hiraoka T, Katayama K, Tanaka J, Ohno N, Joko K, Komiya Y, et al. Lack of epidemiological evidence for a role of resolved hepatitis B virus infection in hepatocarcinogenesis in patients infected with hepatitis C virus in Japan. Intervirology 2003;46:171-176.

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 Enomoto[3] and Namiki Izumi[1]

[1]*Division of Gastroenterology and Hepatology and* [2]*Division of Pathology, Musashino Red Cross Hospital, Tokyo,* [3]*First Department of Internal Medicine, University of Yamanashi, Yamanashi, and* [4]*Department 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)

251

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 hematcrit 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, HCVRNA, creatinine, hemoglobin, GGT, LDL-C, and blood sugar: hematcrit 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 forth 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/m²)	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 (10⁴/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 (10⁴/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.

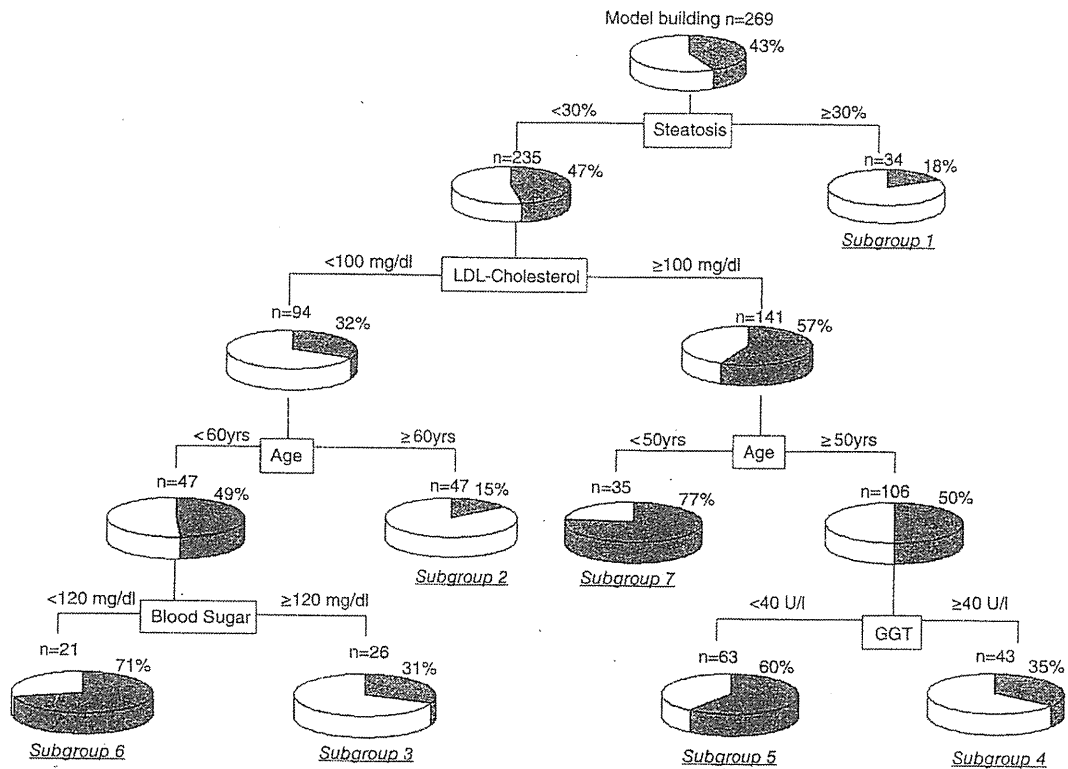© 2010 The Japan Society of Hepatology

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 flowchart 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).
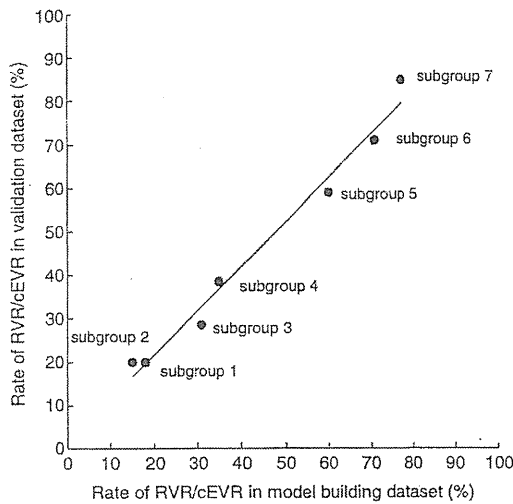
Figure 2 Validation of the classification and regression tree (CART) analysis: Subgroup stratified comparison of the rate of rapid or complete early virological response (RVR/cEVR) between the model building and validation datasets. Each patient in the validation set was allocated to subgroups 1–7 by 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
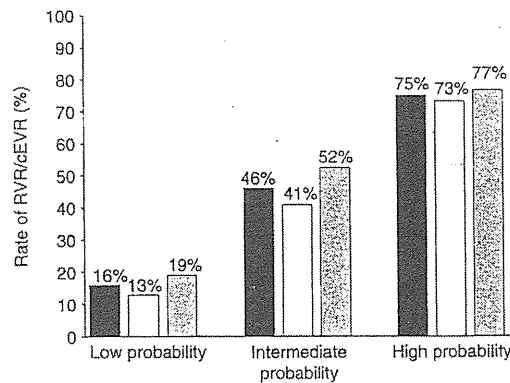

Figure 3 The rate of rapid or complete early virological response (RVR/cEVR) between the classification and regression tree (CART) groups stratified by adherence. The three groups of patients divided by CART (low, intermediate and high probability group) were further stratified according to adherence 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 resistance[26] 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 adipocytokines[34] 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, HCVRNA 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 HCVRNA 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

of CART over traditional regression models is that it can identify prognostic subgroups that are useful in clinical practice. Because the results of CART analysis are presented as a decision tree, which is intuitive, they can be readily interpreted by medical professionals without any specific knowledge of statistics. The most important consideration is that five variables used in the decision tree were clinical parameters that are readily available by the usual work-up of patients before therapy. Especially, glucose, GGT and LDL-C are simple biochemical markers that are easily measured at a low cost. Using this model, we can rapidly develop an estimate of the response before treatment, which may facilitate clinical decision making.

In conclusion, we built a pre-treatment model for the prediction of virological response in PEG-IFN plus RBV therapy. Because this decision tree model was made up of simple host factors such as steatosis, LDL-C, age, blood sugar and GGT, it can be easily applied to clinical practice. This model may have the potential to support decisions in patient selection for PEG-IFN plus RBV therapy based on the possibility of response against a potential risk of adverse events or costs, and may provide a rationale for treating metabolic factors to improve the efficacy of antiviral therapy.

ACKNOWLEDGEMENTS

REFERENCES

1 Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology* 2004; 39: 1147–71.

2 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347: 975–82.

3 Manns MP, McHutchison JG, Gordon SC *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; 358: 958–65.

4 Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 2003; 38: 645–52.

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 Breiman LJH, Friedman RA, Olshen CJ, Stone CM. *Classification and Regression Trees.* Calif: Wadsworth, 1980.

7 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. *Surg* 2002; 132: 589–602.

8 Garzotto M, Beer TM, Hudson RG *et al.* Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 2005; 23: 4322–9.

9 Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 2005; 11: 5440–3.

10 Jin H, Lu Y, Harris ST *et al.* Classification algorithms for hip fracture prediction based on recursive partitioning methods. *Med Decis Making* 2004; 24: 386–98.

11 Baquerizo A, Anselmo D, Shackleton C *et al.* Phosphorus ans an early predictive factor in patients with acute liver failure. *Transplantation* 2003; 75: 2007–14.

12 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.

13 Martin MA, Meyricke R, O'Neill T, Roberts S. Mastectomy or breast conserving surgery? Factors affecting type of surgical treatment for breast cancer – a classification tree approach. *BMC Cancer* 2006; 6: 98.

14 LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res* 1995; 75: 113–24.

15 Costanza MC, Paccaud F. Binary classification of dyslipidemia from the waist-to-hip ratio and body mass index: a comparison of linear, logistic, and CART models. *BMC Med Res Methodol* 2004; 4: 7.

16 Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; 24: 289–93.

17 Kurosaki M, Matsunaga K, Hirayama I *et al.* The presence of steatosis and elevation of alanine aminotransferase levels are associated with fibrosis progression in chronic hepatitis C with non-response to interferon therapy. *J Hepatol* 2008; 48: 736–42.

18 Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989; 8: 539–50.

19 Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 2002; 12: 243–8.

20 Akuta N, Suzuki F, Tsubota A *et al.* Efficacy of interferon monotherapy to 394 consecutive naive cases infected with hepatitis C virus genotype 2a in Japan: therapy efficacy as consequence of tripartite interaction of viral, host and interferon treatment-related factors. *J Hepatol* 2002; 37: 831–6.

21 Poynard T, Ratziu V, McHutchison J *et al.* Effect of treatment with peginterferon or interferon alfa-2b and ribavirin on steatosis in patients infected with hepatitis C. *Hepatology* 2003; 38: 75–85.

22 Bressler BL, Guindi M, Tomlinson G, Heathcote J. High body mass index is an independent risk factor for non-response to antiviral treatment in chronic hepatitis C. *Hepatology* 2003; 38: 639–44.

23 Romero-Gomez M, Del Mar Viloria M, Andrade RJ *et al.* Insulin resistance impairs sustained response rate to peginterferon plus ribavirin in chronic hepatitis C patients. *Gastroenterology* 2005; 128: 636–41.

24 Konishi I, Horiike N, Hiasa Y *et al.* Diabetes mellitus reduces the therapeutic effectiveness of interferon-alpha2b plus ribavirin therapy in patients with chronic hepatitis C. *Hepatol Res* 2007; 37: 331–6.

25 Marchesini G, Avagnina S, Barantani EG *et al.* Aminotransferase and gamma-glutamyltranspeptidase levels in obesity are associated with insulin resistance and the metabolic syndrome. *J Endocrinol Invest* 2005; 28: 333–9.

26 Fraser A, Ebrahim S, Smith GD, Lawlor DA. A comparison of associations of alanine aminotransferase and gamma-glutamyltransferase with fasting glucose, fasting insulin, and glycated hemoglobin in women with and without diabetes. *Hepatology* 2007; 46: 158–65.

27 Mazzella G, Salzetta A, Casanova S *et al.* Treatment of chronic sporadic-type non-A, non-B hepatitis with lymphoblastoid interferon: gamma GT levels predictive for response. *Dig Dis Sci* 1994; 39: 866–70.

28 Villela-Nogueira CA, Perez RM, de Segadas Soares JA, Coelho HS. Gamma-glutamyl transferase (GGT) as an independent predictive factor of sustained virologic response in patients with hepatitis C treated with interferon-alpha and ribavirin. *J Clin Gastroenterol* 2005; 39: 728–30.

29 Berg T, Sarrazin C, Herrmann E *et al.* Prediction of treatment outcome in patients with chronic hepatitis C: significance of baseline parameters and viral dynamics during therapy. *Hepatology* 2003; 37: 600–9.

30 Akuta N, Suzuki F, Kawamura Y *et al.* Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 2007; 46: 403–10.

31 Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* 2001; 33: 1358–64.

32 Ortiz V, Berenguer M, Rayon JM, Carrasco D, Berenguer J. Contribution of obesity to hepatitis C-related fibrosis progression. *Am J Gastroenterol* 2002; 97: 2408–14.

33 Muzzi A, Leandro G, Rubbia-Brandt L *et al.* Insulin resistance is associated with liver fibrosis in non-diabetic chronic hepatitis C patients. *J Hepatol* 2005; 42: 41–6.

34 Charlton MR, Pockros PJ, Harrison SA. Impact of obesity on treatment of chronic hepatitis C. *Hepatology* 2006; 43: 1177–86.

35 Di Bona D, Cippitelli M, Fionda C *et al.* Oxidative stress inhibits IFN-alpha-induced antiviral gene expression by blocking the JAK-STAT pathway. *J Hepatol* 2006; 45: 271–9.

36 Minuk GY, Weinstein S, Kaita KD. Serum cholesterol and low-density lipoprotein cholesterol levels as predictors of response to interferon therapy for chronic hepatitis C. *Ann Intern Med* 2000; 132: 761–2.

37 Gopal K, Johnson TC, Gopal S *et al.* Correlation between beta-lipoprotein levels and outcome of hepatitis C treatment. *Hepatology* 2006; 44: 335–40.

38 Agnello V, Abel G, Elfahal M, Knight GB, Zhang QX. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci USA* 1999; 96: 12766–71.

39 Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. *J Clin Oncol* 2004; 22: 3660–7.

Reproducibility and usability of chronic virus infection model using agent-based simulation; comparing with a mathematical model

Jun Itakura [a,*], Masayuki Kurosaki [a], Yoshie Itakura [a], Sinya Maekawa [b], Yasuhiro Asahina [a], Namiki Izumi [a], Nobuyuki Enomoto [b]

[a] Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan
[b] First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimogatou, Chuou-shi, Yamanashi 409-3898, Japan

ARTICLE INFO

ABSTRACT

We created agent-based models that visually simulate conditions of chronic viral infections using two software. The results from two models were consistent, when they have same parameters during the actual simulation. The simulation results comprise a transient phase and an equilibrium phase, and unlike the mathematical model, virus count transit smoothly to the equilibrium phase without overshooting which correlates with actual biology in vivo of certain viruses. We investigated the effects caused by varying all the parameters included in concept; increasing virus lifespan, uninfected cell lifespan, uninfected cell regeneration rate, virus production count from infected cells, and infection rate had positive effects to the virus count during the equilibrium period, whereas increasing the latent period, the lifespan-shortening ratio for infected cells, and the cell cycle speed had negative effects. Virus count at the start did not influence the equilibrium conditions, but it influenced the infection development rate. The space size had no intrinsic effect on the equilibrium period, but virus count maximized when the virus moving speed was twice the space size. These agent-based simulation models reproducibly provide a visual representation of the disease, and enable a simulation that encompasses parameters those are difficult to account for in a mathematical model.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

All viruses need hosts as a basis for their life. When a virus enters the host body, it invades cells and uses both its own enzymes and those of the host cells to replicate. Host cells infected by viruses launch a self-defense system known as the innate immune system (See and Wark, 2008; Naniche, 2009), which inhibits viral replication and uses the human leukocyte antigen system and cytokines to elicit an immune response. Immune cells that have received signals from host cells activate other immune cells, neutralize viruses in the serum by means of antibodies, and prevent the virus from replicating and proliferating by destroying or curing host cells. Viral infection is a disorder based on the interactions between viruses and cells.

The power relationship between these agents changes along with the progression of the disease. In the very early stages of infection, as the host defense mechanisms are immature, the virus has the ability to overwhelm the host cells, actively replicate, and proliferate. Subsequently, as the capacity of the immune system improves, the speed of viral proliferation drops and the virus count reaches a peak. Infected host cells begin to be disrupted by the immune system or virus particles, and symptoms appear as a result. If the immune system is stronger than the virus, then the viral counts decline, and, in transient viral disorders, the virus is finally eliminated and the host recovers. In chronic viral disorders, however, the power relationship between the virus and host cells reaches equilibrium, and a long-term power balance is maintained with the virus count reaching a plateau.

Mathematical models have been proposed to study the dynamics of such viral disorders, and are regarded as being of value in understanding this phenomenon (Ho et al., 1995; Nowak et al., 1996; Neumann et al., 1998). However, these models are difficult to understand for clinicians, and their applicability is somewhat limited in everyday practice. In clinical research, measurements of viral dynamics in patients for short duration have been made for human
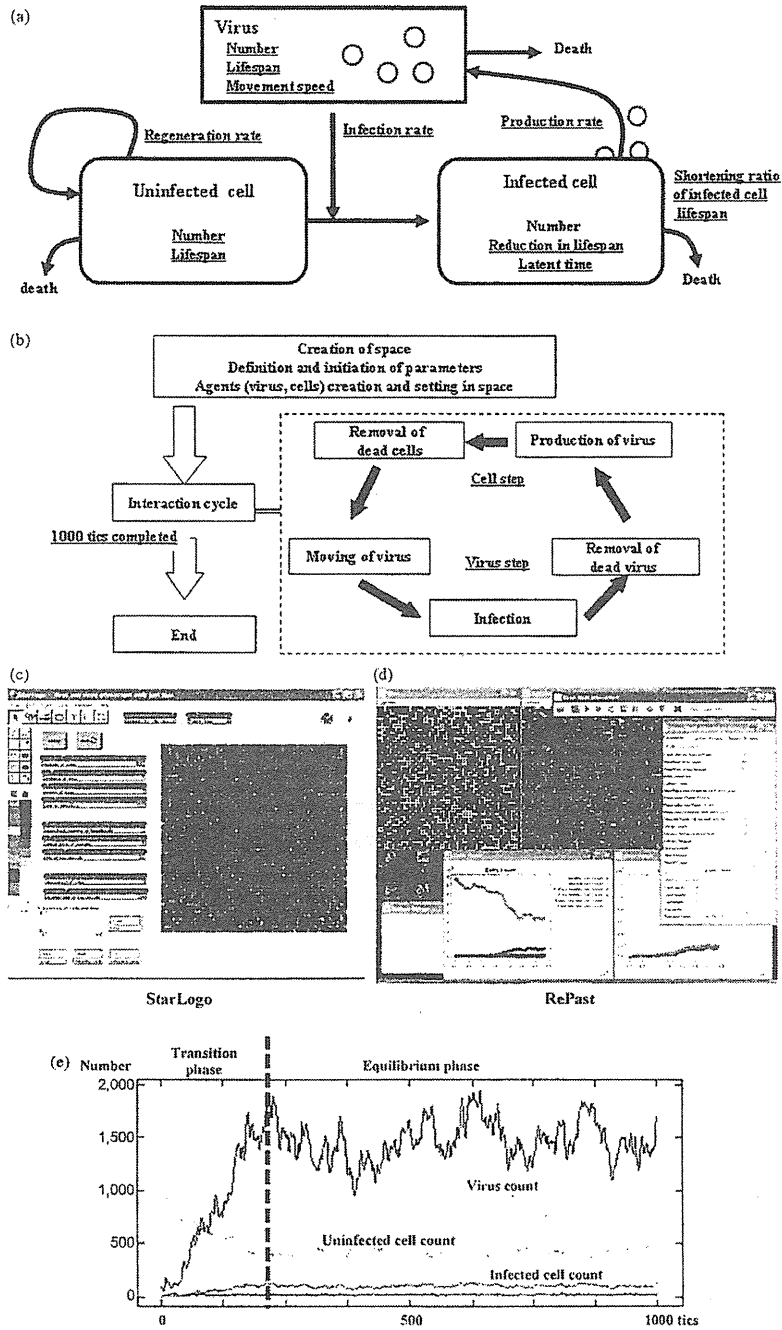
Fig. 1. Simulation design and an example of simulation results. (a) Model concept. Viruses, uninfected cells, and infected cells were treated as agents, and parameters were set for each of these and for interactions between agents (underlined). (b) Flowchart of the program. After preparing the simulation, we entered the interaction cycle, in which virus steps (such as movement) and cell steps were repeated. One cycle was counted as 1 tic, and the simulation concluded after 1000 tics. (c and d) Simulation screen using (c) StarLogo and (d) RePast. Yellow circles are viruses, green squares are uninfected cells, and orange and red indicate infected cells, with orange indicating the latent period. In StarLogo, all the agents are shown on the same screen, but in RePast, viruses and cells are shown in separate windows. (e) Example of a simulation chart in StarLogo. After the start of simulation the virus count and infected cell count increase while the uninfected cell count decreases, with equilibrium state reached after a certain number of tics.
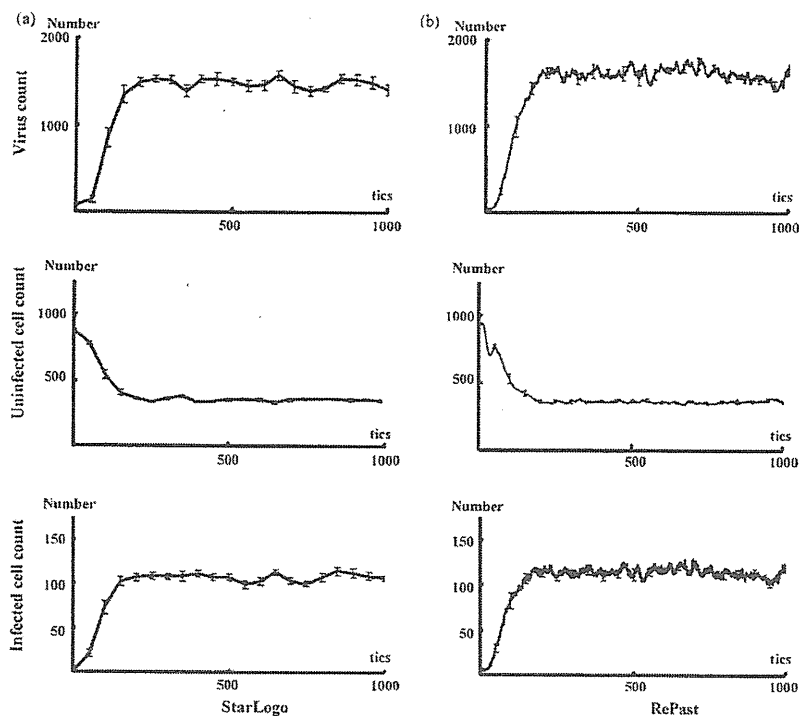
176

Fig. 2. Comparison of simulation results in (a) StarLogo and (b) RePast. The results were consistent when the parameters were made consistent. (Virus count [average ± SD]: StarLogo 1458.03 ± 173.1, RePast 1462.71 ± 178.8, p = 0.94. Uninfected cell count: 364.24 ± 30.4, 368.11 ± 33.4, p = 0.83. Infected cell count: 105.73 ± 13.0, 107.74 ± 13.0, p = 0.24. Unpaired Student's t-test.) Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; and virus reproduction rate, 5/cells/tic. The following parameter settings were taken from actual measurements: virus lifespan, 4.5 tics; uninfected cell lifespan, 49.8 tics; and infected cell lifespan, 6.7 tics.

immunodeficiency virus (HIV) (Ho et al., 1995), hepatitis B virus (HBV) (Nowak et al., 1996) and hepatitis C virus (HCV) (Neumann et al., 1998), and research is also underway on a range of models based on animal experiments and cell culture systems. As chronic viral disorders persist over long periods of time complete follow-up of viral dynamics is difficult. Furthermore, limitations of items that can be measured, such as the difficulty of measuring whole numbers of host cells, make it extremely difficult to investigate their consistency in mathematical models.

The recent ascend of dynamic-models owes much to advances in computers. Computer performance has improved markedly in recent years, not only in terms of their calculating capacity but also with regard to image displays, and models that offer a visual representation of viral disorders are now being reported (Gilbert and Bankes, 2002; Duca et al., 2007; Shapiro et al., 2008; Castiglione et al., 2007). One advantage of such visual models is that by providing a visual representation, they make understanding the disease status easy. Another benefit is that they enable parameters to be identified that are hidden as background noise in mathematical models. However, these models have some problems; it is difficult to prove the reproducibility of the simulation results derived from different languages or libraries, difficult to prove the validity of the model's concepts, and difficult to prove that the simulation results accurately reflect the reality. In this study, we created agent-based computer models that visually simulate the conditions of chronic viral infections using two software. The reproducibility of two agent-based computer models and the differences between agent-based models and the mathematical model were analyzed.

This agent-based model enabled us to investigate how each parameter included in the concept affects the conditions of chronic viral infections.

2. Methods

2.1. Selection of Software

In this study, we used two different types of softwares: StarLogo version 2.0 (http://education.mit.edu/starlogo/) supplied by MIT Media Laboratory and Recursive Porous Agent Simulation Toolkit (RePast-3.0, http://repast.sourceforge.net/) supplied by the Argonne National Laboratory. StarLogo uses Logo, one of the simplest programming languages, and has a fixed graphical user interface. RePast is a library that uses Java, another programming language, which also has a fixed graphical user interface.

Logo is an assembly language, and StarLogo carries out processing sequentially. Java is an object-oriented language, and RePast has a faster processing speed than StarLogo. In addition, StarLogo has a number of stipulations to simplify simulations, such as parameters can only be set up to five decimal places and the simulation space is also fixed as 51 × 51 square grids. RePast, on the other hand, has fewer such restrictions. Thus, it offers a higher degree of freedom in program settings than StarLogo. Taking simulation space as an example, in spite of the restrictions imposed by the underlying operating system's image display system, any number of grids can be set and a hexagonal grid could also be chosen rather than a square one. However, users must stipulate and set all parameters themselves. This means that they must first declare the shape of the grid and the number of grids they will use to fill the simulation space. Java is also more difficult to learn than Logo, and debugging and correcting the program is also more difficult. Thus, it is difficult to judge whether or not the results agree with the planned simulation.

In effect, these two different types of softwares are polar opposites. It is simple to start a simulation in StarLogo, but producing results takes time and it is difficult to carry out more complex simulations. In RePast it is difficult to compose the program and judge whether or not the planned study has actually been achieved, but the

simulation itself takes only a short time to complete and there are lesser restrictions in the construction of a simulation model.

2.2. Concept for Modeling

We applied the basic virus–host interaction mathematical model to the agent-based simulation system with slight modifications. The mathematical model was used to describe the dynamics of HIV (Ho et al., 1995), HBV (Nowak et al., 1996), and HCV (Neumann et al., 1998) and the only agents involved were host cells and viruses, without the inclusion of immune cells. The effects of the immune system are expressed by varying parameters such as lifespan of host cells and viruses.

Fig. 1a illustrates the study concept. Viruses have the ability to infect healthy host cells (uninfected cells) and the infected cells produce new viruses. Both cells and viruses have definite lifespans, and the lifespan of infected cells is usually shorter than that of uninfected cells. Uninfected cells automatically regenerate within the space, whereas infected cells only arise due to infection of uninfected cells. Viruses also lack the ability to regenerate themselves and are only produced from infected cells.

2.3. Parameter Settings

In the present study, as the StarLogo settings are circumscribed, we limited the simulation space to 51 × 51 square grids. However, we made an exception here while investigating the effects of size of space on the simulation results. The numbers of viruses, uninfected cells, and infected cells could only be set before the start of the simulation. As described in the later, our simulation ran in cycles, with 1 cycle defined as 1 tic.

In mathematical simulation models, the death rate is required as a parameter. However, in our program we set lifespans for viruses and uninfected cells. These lifespans were not uniform, but were set to have a deviation of about 10%. The lifespan of cells was shortened by infection with ratio decided beforehand.

The infection ratio was meaningful only when an infected cell and a virus coincidentally occupied the same grid, and this was used to calculate the probability of the infection occurring in that situation. The virus production rate was set as the number of viruses produced by an infected cell during 1 tic. Infected cells could be set as a parameter indicating the latent period between the time of virus infection and the time of virus replication.

In order to emulate the tissue repair capacity, we set uninfected cell regeneration rate such that grids without any cells had a specified probability of producing uninfected cells on top of themselves. As a result, the more the cell count declined within a space the more regenerated uninfected cells were produced, whereas the number of regenerated cells declined as cell count increased.

The number of grids through which a virus could move in 1 tic was set as the speed of movement, and the direction of movement was set within a range of 90° toward the top of the simulation space. The program used a circulatory method of movement: when a virus arrived at the top of the space, it was translocated to the bottom, and moved again toward the top. Cells were fixed on the grid.

2.4. Simulation Flowchart

Fig. 1b shows a flowchart of the program. First, the simulation space was produced, after which each parameter was defined and the initial settings were made. Next the agents – viruses and uninfected and infected cells – were produced. The simulation cycle was as follows. Viruses moved to a new grid, and if an uninfected cell was present, this was infected with a probability based on the infection rate. The lifespan of the virus decreased, and viruses that had completed their lifespan and those that had caused an infection were removed from the space. Infected cells produced new viruses, the lifespans of both uninfected and infected cells decreased. Then, cells that had completed their lifespan were eliminated and a new cycle began. The program was set such that the simulation ended after this cycle had repeated 1000 times. This meant that one simulation was complete after 1000 tics.

2.5. Data Collection

The RePast model was programmed such that data for each tic was saved automatically as a text file at the end of the simulation. This text file could be opened by a database software. The StarLogo model was programmed to stop the simulation and collect data after every 50 tics.

2.6. Mathematical Model

In order to compare the results of this agent-based simulation, we used a viral infection mathematical model, which we improved as follows.

$$\frac{dT}{dt} = s[2601 - (T + I)] - dT - bVT \tag{1}$$

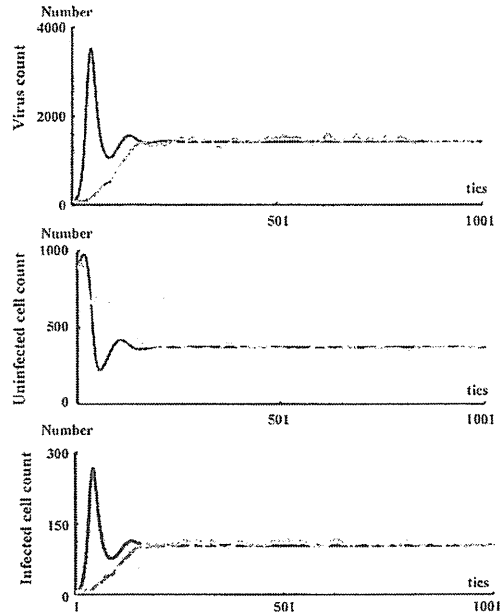$$\frac{dI}{dt} = bVT - dI \tag{2}$$

Fig. 3. Comparison of results of agent-based simulation and mathematical simulation. Both sets of results were consistent for the equilibrium phase, but differed in the shift in transition phase. Black line: mathematical model; grey line: results of simulation in RePast. Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; virus reproduction rate, 5/cells/tic; virus lifespan, 10 tics; uninfected cell lifespan, 50 tics; and cell lifespan-shortening ratio as a result of infection, 69%.

$$\frac{dV}{dt} = pI - cV \tag{3}$$

where, T is the uninfected cell count, I is the infected cell count, and V is the virus count. Uninfected cells are supplied to the space with a probability s[2601 − (T+I)], as the number of grids in this agent-based simulation model was 2601 (51 × 51). The death rate of uninfected cells is d, the death rate of infected cells is δ, and the death rate of viruses is c. The infection rate is indicated by β. Viruses are released from infected cells at a probability p.

2.7. Statistical Analysis

Statistical analyses were performed by statistical tests using the program StatView 5.0 (SAS Institute Inc.). All tests of significance were two-tailed, with p values of <0.05 considered to be significant.

3. Results

3.1. Reproducibility of Chronic Viral Infection Disease Models Using Agent-based Simulation Methods

We constructed the chronic viral infection model with StarLogo library. Fig. 1c shows the simulation screen, and Fig. 1e shows one sample result. Immediately after the start of the simulation, the virus count temporarily dropped in accordance with the onset of an infection. Subsequently, the virus count started to increase with an increase in the infected cells and a decrease in the uninfected cells. After a certain number of tics (around 300 in this example), although the virus count, infected cell count, and uninfected cell count had some fluctuation, an equilibrium state was reached. We use the following descriptive terms in this paper: the transient phase is the period during which virus growth peaks, and the equilibrium phase is the period during which an equilibrium state is
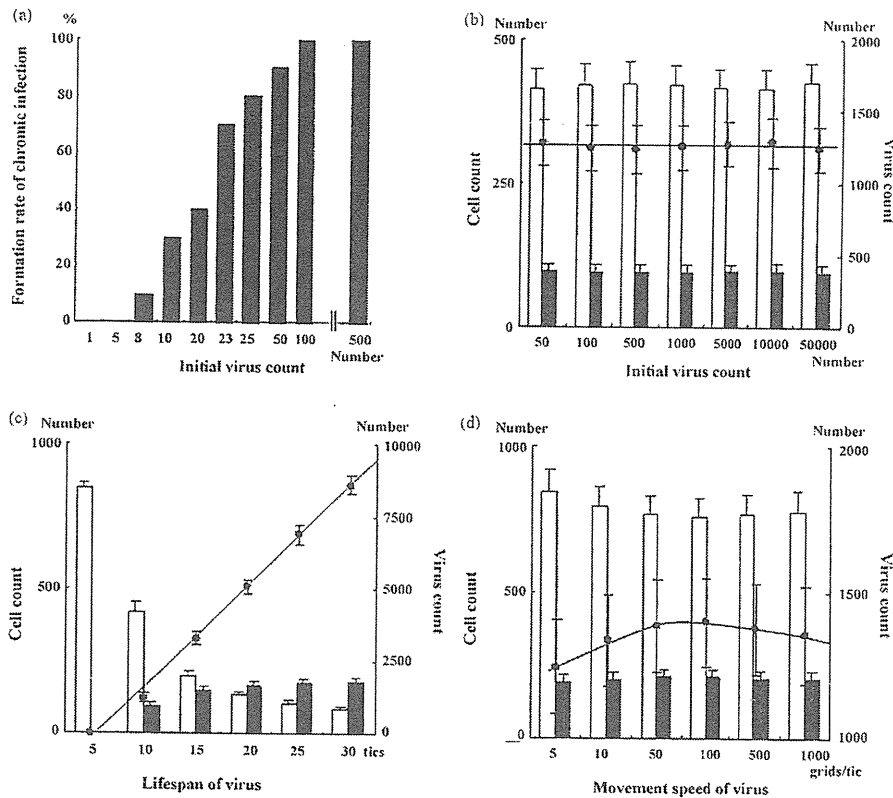
Fig. 4. Effects of changes in viral parameters. (a) The higher the initial virus count, the greater is the increase in the rate of formation of chronic infection, but (b) there was no effect on the conditions in the equilibrium phase. (c) Extending the virus lifespan increased the virus count. (d) Increasing the speed of virus movement to 100 grids/tic increased the virus count, but increasing it to 500 grids/tic had the opposite effect, with a slight declining trend. (a) Black bars: number of infections produced; (b–d) black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

established. When the simulation was performed multiple times, the features described above were maintained, and the average values for virus, infected cell, and uninfected cell counts during the equilibrium state were all consistent.

Fig. 1d shows the simulation screen of the RePast. When we attempted setting all the initial parameters to the same values as those in the StarLogo, the results were not consistent. When we recalculated the parameters from the simulation results, in RePast, the parameters were largely maintained at the levels of the settings, but in StarLogo, the lifespans of both cell types became shorter than the settings while the simulation was in progress. We made the results of both simulations consistent by using the same parameters during the actual simulation (Fig. 2a and b).

3.2. Comparison Between Agent-based Simulation Models and Mathematical Simulation Model

We investigated whether the results of a chronic viral infection disease model produced by RePast would be consistent with the results of a mathematical model. For the mathematical model, we carried out an approximate integration using a four-dimensional Runge–Kutta method to ensure that the uninfected cell count and infected cell count would be in the same class. Parameters were always fixed as constant between simulations. The simulation results were consistent for the equilibrium

phase, but transitions in virus count during the transient phase varied, with a shift to equilibrium state following two overshoots in the mathematical model, but a monotonic increase following a logistic curve in the agent-based model (Fig. 3). In the mathematical model, when the equilibrium condition was calculated with $dT/dt = dI/dt = dV/dt = 0$, the equilibrium-phase virus count, uninfected cell count, and infected cell count were very similar to those of the agent-based model (virus count: mathematical model 371.8/space, agent-based model 371.1 ± 32.4/space [average ± SD]; uninfected cell count: mathematical model 1605/space, agent-based model 1454 ± 194/space; infected cell count: mathematical model 115.9/space, agent-based model 108.3 ± 14.2/space).

3.3. Usability of the Models: Effect of Changing Parameters

We investigated the changes in the equilibrium phase brought about by changing each parameter. All the investigations below were carried out by using RePast, and we used the average values from ten simulations.

3.4. Viral Parameters

The lower the virus counts at the beginning of the simulation, the lower the probability of a chronic infection (Fig. 4a). However, the initial virus count did not have any effect on the equilibrium
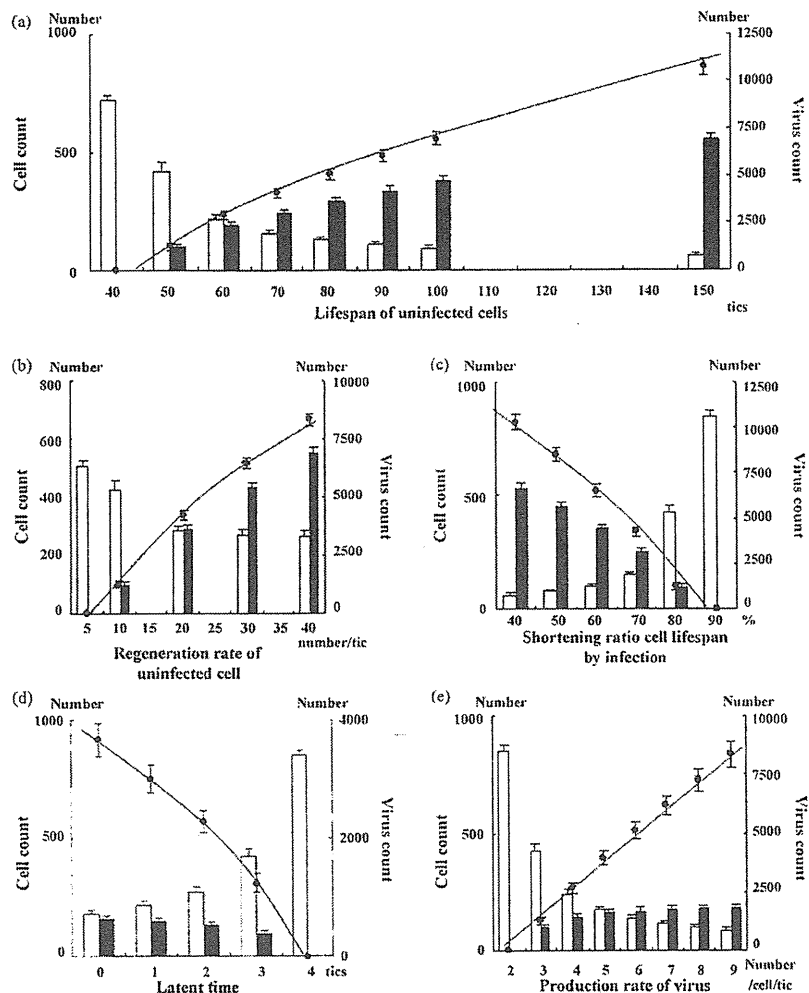
Fig. 5. Effects of changes in cell parameters. (a) Extending the uninfected cell lifespan and (b) increasing the uninfected cell regeneration rate increased the virus count. (c) Raising the lifespan-shortening ratio as a result of infection shortened the lifespan of infected cells, thereby decreasing the virus count. (d) Extending the latent period shortened the period of virus production from infected cells, thereby decreasing the virus count. (e) Increasing the virus production count resulted in a linear increase in equilibrium-phase virus count. Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

phase itself (Fig. 4b). Extending the lifespan of viruses resulted in a linear increase in equilibrium-phase virus count (Fig. 4c). Although the infected cell count increased, the rate of increase gradually declined. Changing the speed of viral movement resulted in the equilibrium-phase virus count to eventually decline after 100 grids/tic was reached, allowing movement over an area twice the size of the simulation space (Fig. 4d).

3.5. Uninfected Cell Parameters

Extending the lifespan of uninfected cells led to an increased virus count during the equilibrium phase (Fig. 5a). Increasing the uninfected cell regeneration rate also contributed to increased equilibrium-phase virus count (Fig. 5b). In both the cases, the

increases in virus count and infected cell count were not linear, but showed a tendency for the rate of increase to decline gradually.

3.6. Infected Cell Parameters

We carried out an investigation of the effects of variation in the lifespan-shortening ratio on the virus count on the assumption that cell lifespan is shortened by infection. When this ratio was increased, the virus count decreased (Fig. 5c). An extended latent period was also related to a decreased virus count (Fig. 5d). However, the virus production from infected cells led to a linear increase in the virus count (Fig. 5e).
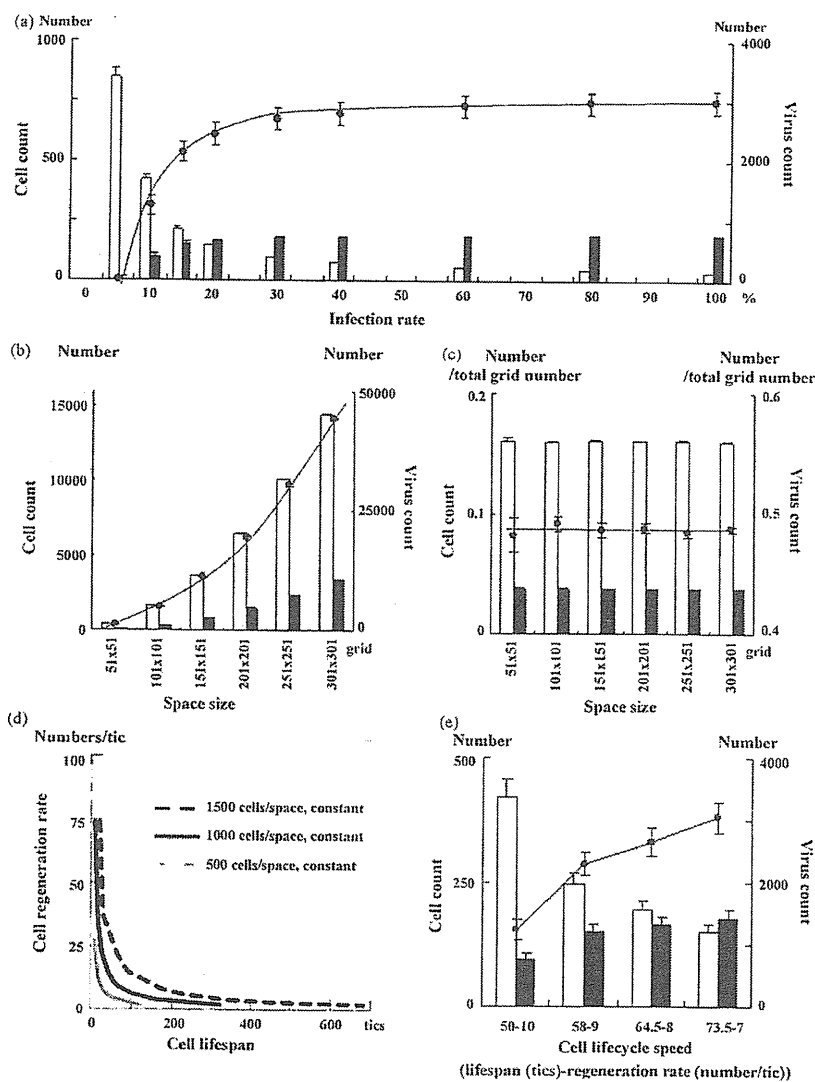
Fig. 6. (a) Increasing the infection rate increased the virus count in equilibrium periods, but the virus count did not change at infection rates of 30% or more. (b) The size of the simulation space increased not only virus count but also the cell count; however, (c) when virus and cell counts were divided by the total number of grids in the space, they were constant for all space sizes. (d) Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible to change only the cell cycle speed without altering the uninfected cell count. (e) When the cell cycle speed was reduced, the virus count increased toward the right of the graph. This may be because the effect of extending the lifespan of cells exceeds that of reducing their regeneration rate. (a–c and e) Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

3.7. Infection Rate and Space Size

Increasing the infection rate caused an increase in the virus count, but the change was minimal at an infection rate of 30% or more. The same results were seen for infected cell count, but a decrease in uninfected cell count resulted in a tendency for the infection rate to decrease by up to 60% (Fig. 6a).

The larger the space, higher the increase in both virus and cell counts (Fig. 6b). This increase was proportional to space size, how-

ever, when virus and cell counts were divided by the total number of grids in the space they were all constant (Fig. 6c).

3.8. Cell Cycle Speeds

Running a simulation with the initial virus count set to zero enables only the equilibrium condition for uninfected cells to be simulated. Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible