

The NAFLD Fibrosis Score: A Noninvasive System That Identifies Liver Fibrosis in Patients with NAFLD

Paul Angulo,¹ Jason M. Hui,² Giulio Marchesini,³ Elisabetta Bugianesi,⁴ Jacob George,² Geoffrey C. Farrell,² Felicity Enders,⁵ Sushma Saksena,⁶ Alastair D. Burt,⁶ John P. Bida,⁵ Keith Lindor,¹ Schuyler O. Sanderson,⁷ Marco Lenzi,⁸ Leon A. Adams,¹ James Kench,⁹ Terry M. Therneau,⁵ and Christopher P. Day⁶

Patients with nonalcoholic fatty liver disease (NAFLD) and advanced liver fibrosis are at the highest risk for progressing to end-stage liver disease. We constructed and validated a scoring system consisting of routinely measured and readily available clinical and laboratory data to separate NAFLD patients with and without advanced fibrosis. A total of 733 patients with NAFLD confirmed by liver biopsy were divided into 2 groups to construct ($n = 480$) and validate ($n = 253$) a scoring system. Routine demographic, clinical, and laboratory variables were analyzed by multivariate modeling to predict presence or absence of advanced fibrosis. Age, hyperglycemia, body mass index, platelet count, albumin, and AST/ALT ratio were independent indicators of advanced liver fibrosis. A scoring system with these 6 variables had an area under the receiver operating characteristic curve of 0.88 and 0.82 in the estimation and validation groups, respectively. By applying the low cutoff score (-1.455), advanced fibrosis could be excluded with high accuracy (negative predictive value of 93% and 88% in the estimation and validation groups, respectively). By applying the high cutoff score (0.676), the presence of advanced fibrosis could be diagnosed with high accuracy (positive predictive value of 90% and 82% in the estimation and validation groups, respectively). By applying this model, a liver biopsy would have been avoided in 549 (75%) of the 733 patients, with correct prediction in 496 (90%). **Conclusion:** a simple scoring system accurately separates patients with NAFLD with and without advanced fibrosis, rendering liver biopsy for identification of advanced fibrosis unnecessary in a substantial proportion of patients. (HEPATOLOGY 2007;45:846-854.)

Paralleling the increasing prevalence of obesity, diabetes mellitus, and the metabolic syndrome in the general population, nonalcoholic fatty liver disease (NAFLD) has become the most common cause of chronic liver disease worldwide.¹⁻⁴ One in 3 adult Americans¹ and 1 in 4 or 5 adult Italians² suffer from NAFLD. NAFLD

also has reached epidemic proportions among populations typically considered at low risk, with a prevalence of 15% in China³ and 14% in Japan.⁴ The clinical implications of this alarming prevalence of NAFLD are derived from the fact that NAFLD may progress to cirrhosis, liver failure, and HCC.⁵⁻⁸

Within the spectrum of NAFLD, simple bland steatosis often remains stable for a number of years and will probably never progress in many patients.^{9,10} A subset of patients, however, particularly those with more advanced fibrosis, are at a higher risk for progressing to decompensated cirrhosis, portal hypertension, HCC, or death if liver transplantation is not accomplished.^{5-8,11-14} In contrast to patients with bland steatosis, patients with increased liver fibrosis require close follow-up with surveillance for the development of esophageal varices and HCC and enrollment into treatment trials. Thus, identifying the presence and severity of liver fibrosis in patients with NAFLD is of major importance in guiding the subsequent management of patients with this liver condition.

In the absence of decompensated cirrhosis, liver biopsy remains the only reliable means to determine prognosis

Abbreviations: BMI, body mass index; HDL, high-density lipoprotein; HOMA, homeostatic model assessment; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; ROC, receiver operating characteristics curve.

From the ¹Division of Gastroenterology and Hepatology, ²Department of Health Sciences Research, and ⁷Division of Anatomic Pathology, Mayo Clinic College of Medicine, Rochester, MN; ³Storr Liver Unit and ⁹Department of Anatomical Pathology, Westmead Hospital and Westmead Millennium Institute, University of Sydney, Australia; ⁴Department of Medicine and ⁸Pathology, Alma Mater Studiorum University, Bologna, Italy; ⁵Division of Gastroenterology, University of Turin, Ospedale San Giovanni Battista, Turin, Italy; ⁶School of Clinical Medical Sciences, University of Newcastle and Freeman Hospital, Newcastle upon Tyne, UK.

Received May 19, 2006; accepted November 2, 2006.

Address reprint requests to: Paul Angulo, M.D., Miles and Shirley Fiterman Center for Digestive Diseases, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905. E-mail: angulohernandez.paul@mayo.edu; fax: 507-266-4531.

Copyright © 2007 by the American Association for the Study of Liver Diseases.

Published online in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/hep.21496

Potential conflict of interest: Nothing to report.

based on the severity of fibrosis. However, liver biopsy is an expensive and invasive procedure associated with a number of complications and prone to sampling error. Because of all this, efforts have been made to identify noninvasive indicators of liver fibrosis in patients with NAFLD. Noninvasive approaches for assessing the severity of fibrosis in NAFLD have included a combination of clinical features and routine laboratory investigations¹⁵⁻¹⁸ as well as some less readily available serum markers of fibrosis.¹⁹⁻²² These noninvasive approaches are, however, either insufficiently accurate in their prediction of liver fibrosis or their diagnostic accuracy has been evaluated in only a limited number of patients with NAFLD; further, most have not been validated in a separate population of NAFLD patients. A model using routinely measured clinical and laboratory variables that accurately predicts the severity of liver fibrosis in patients with NAFLD has therefore yet to be developed and validated. Hence, the purpose of this study was (1) to develop a simple noninvasive scoring system aimed at separating patients with NAFLD with and without advanced liver fibrosis by using routinely determined and easily available clinical and biochemical variables, and (2) to validate the results in a separate cohort of patients. The NAFLD fibrosis score we developed is a clinically applicable and useful method to separate patients with and without prognostically significant NAFLD.

Materials and Methods

Patient Population. A total of 733 patients with well-characterized and liver biopsy–confirmed NAFLD were included in this study. They were untreated, consecutively biopsied patients seen at the Mayo Clinic in Rochester, MN ($n = 356$); Newcastle, UK ($n = 158$); Sydney, Australia ($n = 123$); and Italy ($n = 96$) during 2000-2003. The study was approved by appropriate regulatory bodies at all centers.

The diagnosis of NAFLD was based on the following criteria: (1) elevated aminotransferases (AST and/or ALT); (2) liver biopsy showing steatosis in at least 10% of hepatocytes; and (3) appropriate exclusion of liver disease of other etiology including alcohol-induced or drug-induced liver disease, autoimmune or viral hepatitis, and cholestatic or metabolic/genetic liver disease. These other liver diseases were excluded using specific clinical, biochemical, radiographic, and/or histological criteria. Aminotransferase levels had been elevated for a mean of 17.5 months (median 9.0 months, interquartile range 6, 20). All patients had a negative history of ethanol abuse as indicated by a weekly ethanol consumption of <140 g in women and <210 g in men. History of alcohol consump-

tion was specifically investigated by interviewing the patients, and in almost all cases, a close relative. In the Newcastle center, alcohol levels in urine were measured randomly to rule out patients who abused alcohol. Patients with clinical or imaging evidence of decompensated cirrhosis (i.e., portal systemic encephalopathy, variceal bleeding, ascites) were specifically excluded from this study because they most likely had cirrhotic-stage NAFLD regardless of what a model may predict.

Clinical and laboratory data were collected on the date a diagnostic liver biopsy was performed. A complete medical history and physical examination was undertaken in all patients. Body mass index (BMI) was calculated using the formula: weight (in kilograms)/height (in meters²). Waist circumference (to the nearest half centimeter) was measured at the midpoint between the lower border of the rib cage and the iliac crest. Laboratory evaluation included routine liver biochemistry (ALT and AST levels, total bilirubin, albumin, alkaline phosphatase, and gamma glutamyl transpeptidase); complete blood count; total cholesterol, HDL cholesterol, and total triglycerides; fasting glucose; fasting insulin; ferritin levels; viral serology for hepatitis B and C infection; autoantibodies; alpha 1 antitrypsin levels and phenotype; and ceruloplasmin levels.

The degree of insulin resistance was determined by the homeostatic model assessment (HOMA) using the formula²³: insulin resistance = (insulin \times glucose)/22.5. HOMA correlates well with the “gold standard” hyperinsulinemic euglycemic clamp technique, but HOMA is more appropriate for studies with large patient populations. Components of the metabolic syndrome were recorded including central obesity (waist circumference >102 cm for men and >88 cm for women; or ≥ 90 cm in Asian men and ≥ 80 cm in Asian women), hyperglycemia (fasting blood glucose ≥ 110 mg/dl or previously diagnosed type 2 diabetes), hypertriglyceridemia (triglycerides ≥ 150 mg/dl or under treatment for this lipid abnormality), hypertension (blood pressure $\geq 130/\geq 85$ or treatment of previously diagnosed hypertension), and low HDL cholesterol (<40 mg/dl in men or <50 mg/dl in women). The presence of diabetes mellitus (fasting glucose ≥ 126 mg/dl or treatment with antidiabetic drugs), obesity (BMI ≥ 30 kg/m², or ≥ 25 kg/m² in Asians), and overweight (BMI 25-29.9 kg/m², or 23-24.9 kg/m² in Asians) was also recorded. The AST/platelet ratio was calculated because it has been previously associated with the severity of fibrosis in patients with chronic hepatitis C infection²⁴ using the formula: AST (xULN) \times 100/platelet count (10⁹/l).

Liver Histology. Liver biopsies were routinely stained with hematoxylin and eosin, Masson’s trichrome, and

special stains for iron and copper. Liver biopsies were read by a single liver pathologist in each participating center, and the stage of fibrosis was scored based on the 5-point scale proposed by Brunt et al. and recently modified by Kleiner et al.²⁵ Briefly, stage 0 = absence of fibrosis; stage 1 = perisinusoidal or portal; stage 2 = perisinusoidal and portal/periportal; stage 3 = septal or bridging fibrosis; and stage 4 = cirrhosis. To control for biopsy size, the length of the biopsy was measured with a hand ruler, and the number of portal areas on one cross-section was counted.

Statistical Analysis. The main endpoint of the study was to predict the presence or absence of advanced fibrosis (stages 3-4) by a combination of simple and clinically relevant variables. Data from each of the 4 countries were randomly separated into 2/3 and 1/3 of patients for model building and model validation, respectively. Hence, data on 480 patients were used to build a model, whereas data on 253 patients were used to validate the model. Univariate descriptive statistic was used to compare patients with and without significant fibrosis. All variables were included in a multivariate backward stepwise logistic regression analysis to identify variables independently associated with presence or absence of advanced fibrosis. The variables in the resulting model were assessed for all 2-way interactions, as well as interaction with other variables suggested by clinical knowledge. The results of the multivariate analysis were adjusted for site. Those variables with $P < 0.05$ by multivariate analysis were used to construct a scoring system to predict advanced fibrosis. The overall diagnostic accuracy of the scoring system was determined by calculating the area under the receiver operating characteristic (ROC) curve (the c-statistic) and its 95% confidence intervals. Validation was performed (1) in the validation dataset ($n = 253$) and (2) in the full dataset ($n = 733$). In both cases, cross-validation was used with 20 subgroups, so that at most 5% of the data under consideration was excluded at any one time. By employing cross-validation, the possibility of an unusually positive or negative validation subset could be assessed. ROC curve estimates from cross-validation analysis were found with the jackknife method, which provided a simple overall ROC curve from the 20 cross-validation groups. Using the ROC curve for the final model, 2 cutoff points were selected, so that the positive predictive value (PPV) and negative predictive value (NPV) for advanced fibrosis were at least 90%, on the assumption that false results of less than 10% are clinically acceptable. The diagnostic accuracy of the 2 cutoff points was determined by calculating sensitivity, specificity, PPV, NPV, and likelihood ratios. For each decile of probability of significant fibrosis, the full dataset was used to identify both the average value

of the new scoring system and a 95% confidence interval (CI). All analyses were carried out using the statistical analysis software SAS Release 8.2 (SAS Institute Inc., Cary, NC).

Results

Characteristics of the Patient Population. Table 1 summarizes the clinical, laboratory, and liver biopsy data of the patient population. The 733 patients had an age range from 11 to 81 years, about half were female, and most were Caucasians. More than half were obese as defined by BMI, had central obesity as defined by waist circumference, or suffered from hypertriglyceridemia or low HDL cholesterol. About a third or more patients suffered from hyperglycemia/diabetes or hypertension. Of the total cohort, 244 (33%) patients did not have fibrosis on liver biopsy, 290 (40%) had stage 1-2 fibrosis, and 199 (27%) had advanced (stage 3-4) fibrosis.

Liver Biopsy Size and Correlation with Advanced Fibrosis. The mean (\pm SD) length of the liver biopsy was 18.7 ± 8.5 mm (median 17 mm, interquartile range 14, 22) in the total population, and similar between estimation (19 ± 8.4 mm, median 17 mm, interquartile range 14, 23) and validation (18.1 ± 8.8 mm; median 16 mm, interquartile range 13, 22) groups ($P = 0.8$) (Table 1). The number of portal areas was 10 ± 4.5 (median 9.0, interquartile range 6, 15) in the total patient population, and not significantly different between estimation (9.9 ± 4.5 ; median 9, interquartile range 6, 15) and validation (10.1 ± 4.5 ; median 9, interquartile range 6, 15) groups ($P = 0.8$). The liver biopsy length correlated significantly with the number of portal areas ($r = 0.72$, $P < 0.001$) in the total patient population, and in the estimation ($r = 0.70$, $P < 0.001$) and validation ($r = 0.74$, $P < 0.001$) groups. Advanced fibrosis, however, did not correlate significantly with biopsy length ($r = 0.04$, $P = 0.4$) or number of portal areas ($r = 0.04$, $P = 0.6$) in the total patient population. Similarly, advanced fibrosis did not correlate significantly with biopsy length ($r = 0.01$, $P = 0.9$) or number of portal areas ($r = 0.1$, $P = 0.2$) in the estimation or validation ($r = 0.04$, $P = 0.6$ and $r = 0.3$, $P = 0.8$, respectively) groups. Further, the biopsy length and number of portal areas were not significantly associated with advanced fibrosis in univariate or multivariate analyses (Table 2).

Predictors of Fibrosis. Table 2 shows the univariate comparison and the results of the multivariate analysis performed in the 480 patients comprising the estimation group. By multivariate analysis, 5 variables remained significant including age, BMI, AST/ALT ratio, platelet

Table 1. Characteristics of the Patient Population

Variable	All patients (n = 733)	Estimation Group (n = 480)	Validation Group (n = 253)
Age (years)	47.7 ± 13.2	47.7 ± 13	47.7 ± 13.6
Gender (male)	390 (53%)	267 (56%)	123 (49%)
Race (Caucasian/other)	659 (90%)	427 (90%)	232 (92%)
BMI (kg/m ²)	32.2 ± 6.1	31.9 ± 5.8	32.8 ± 6.7
BMI: normal/overweight/obese (n)	47/243/443	27/176/277	20/67/166
Waist circumference (cm)	101.8 ± 16.6	101.7 ± 15.6	102 ± 15.7
Central obesity	437 (60%)	284 (59%)	153 (60%)
Waist-to-hip ratio	0.96 ± 0.07	0.96 ± 0.07	0.96 ± 0.08
ALT (U/l)	87 ± 72	88 ± 74	85 ± 66
AST (U/l)	60 ± 50	60 ± 52	58 ± 46
AST/ALT ratio	0.84 ± 0.83	0.81 ± 0.73	0.90 ± 1.0
Albumin (g/dl)	4.3 ± 0.5	4.3 ± 0.5	4.3 ± 0.5
Total bilirubin (mg/dl)	0.8 ± 0.6	0.8 ± 0.6	0.8 ± 0.4
AST/platelet ratio	0.94 ± 0.9	0.97 ± 0.97	0.88 ± 0.73
Platelet count (×10 ⁹ /l)	235 ± 84	232 ± 82	241 ± 86
GGT (U/l)	96 ± 105	91 ± 106	92 ± 105
HOMA	5.53 ± 5.91	5.37 ± 6.07	5.84 ± 5.6
Glucose (mg/dl)	116 ± 50	115 ± 47	119 ± 55
Diabetes mellitus (yes)	219 (30%)	138 (29%)	81 (32%)
Hyperglycemia (yes)	283 (39%)	176 (37%)	107 (42%)
Triglycerides (mg/dl)	211 ± 161	205 ± 150	222 ± 179
Hypertriglyceridemia (yes)	438 (60%)	280 (58%)	158 (62%)
Total cholesterol (mg/dl)	209 ± 50	209 ± 48	210 ± 53
Low HDL-cholesterol (yes)	372 (51%)	236 (49%)	136 (54%)
Hypertension (yes)	220 (30%)	144 (30%)	76 (30%)
Biopsy length (mm)	18.7 ± 8.5	19 ± 8.4	18.1 ± 8.8
Portal areas (n)	10 ± 4.5	9.9 ± 4.5	10.1 ± 4.5
Fibrosis Stage (0/1/2/3/4)	244/190/100/95/104	167/128/60/58/67	77/62/40/37/37
Mayo Clinic	110/99/31/49/67	75/68/19/28/44	35/31/12/21/23
United Kingdom	74/34/19/16/15	47/20/12/9/8	27/14/7/7/7
Australia	27/41/17/17/21	22/28/10/13/14	5/13/7/4/7
Italy	33/16/33/11/1	23/12/19/8/1	10/4/14/5/0

NOTE. The table shows the mean ± SD for continuous variables, number (%) for binary variables, and number per group for categorical variables. The AST/platelet ratio was calculated using the formula:²³ AST (× ULN) × 100/platelet count (10⁹/l). Central obesity is defined as waist circumference >102 cm for men and >88 cm for women; or ≥90 cm in Asian men and ≥80 cm in Asian women; hyperglycemia, fasting blood glucose ≥110 mg/dL or previously diagnosed type 2 diabetes; hypertriglyceridemia, triglycerides ≥150 mg/dl or under treatment for this lipid abnormality; hypertension, blood pressure ≥130/≥85 or treatment of previously diagnosed hypertension; and low HDL-cholesterol, <40 mg/dl in men or <50 mg/dl in women; diabetes mellitus, fasting glucose ≥126 mg/dl or treatment with antidiabetic drugs; obesity, BMI ≥30 kg/m², or ≥25 kg/m² in Asians; and overweight, BMI 25-29.9 kg/m², or 23-24.9 kg/m² in Asians.

count, and hyperglycemia (Table 2). No statistically significant interactions were identified.

Model Building. In our model building process, we found a tendency for albumin or the AST/ALT ratio to be excluded because these 2 variables were negatively correlated ($r = -0.34$) leading to slight variability in model selection. In the multivariate analysis (Table 2), albumin had a $P = 0.07$. The addition of albumin to the other significant 5 variables did not have a major effect on model accuracy as indicated by an area under the ROC curve of 0.88 when albumin was included and 0.85 when albumin was eliminated. However, a model with the 5 significant variables plus albumin significantly increased the proportion of patients correctly identified with or without significant fibrosis by 13%. On this basis, we decided to keep albumin as one of the variables in the final model. Using these 6 variables, we constructed a scoring

system (risk score formula) to distinguish between patients with (F3-F4) and without (F0-F2) advanced fibrosis. This scoring system had an area under the ROC curve of 0.88 ± 0.02 (95% CI = 0.85, 0.92) (Fig. 1). The regression formula (risk score) for prediction of severity of fibrosis based on these 6 variables is: NAFLD fibrosis score = $-1.675 + 0.037 \times \text{age (years)} + 0.094 \times \text{BMI (kg/m}^2) + 1.13 \times \text{IFG/diabetes (yes = 1, no = 0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelet (} \times 10^9/\text{l)} - 0.66 \times \text{albumin (g/dl)}$.

Using the area under the ROC curve, 2 cutoff points were selected to identify the presence (greater than 0.676) and absence (lower than -1.455) of significant fibrosis (Fig. 1).

By applying the low cutoff point (score below -1.455), 273 (77%) of the 355 patients without significant fibrosis were correctly identified, whereas 22

Table 2. Variables Associated with Presence of Advanced Fibrosis (Stage 3-4) in the Estimation Group (n = 480)

Variable	Unadjusted (Univariate)			Adjusted (Multivariate)		
	Odds Ratio	95% CI (low, high)	P Value	Odds Ratio	95% CI (low, high)	P Value
Age (per year)	1.07	1.05, 1.09	<0.0001	1.04	1.01, 1.07	0.007
Gender (M)	0.42	0.28, 0.64	<0.0001			
Race (Caucasian)	0.71	0.34, 1.40	0.3			
BMI (continuous)	1.10	1.06, 1.14	<0.0001	1.10	1.04, 1.16	<0.001
Obesity	2.69	1.74, 4.25	<0.0001			
BMI (categories)						
normal (reference)	1					
overweight	2.35	0.78, 10.2	0.3			
obese	5.70	1.96, 23.95	<0.001			
Waist circumference(continuous)	1.01	0.99, 1.02	0.3			
Central obesity	1.22	0.80, 1.87	0.4			
Waist-to-hip ratio	0.21	0.002, 20.5	0.5			
AST/ALT ratio	6.89	4.00, 12.31	<0.0001	2.70	1.33, 5.62	0.007
Albumin (g/dl)	0.15	0.09, 0.25	<0.0001	0.51	0.25, 1.05	0.073
Bilirubin (mg/dl)	1.68	1.21, 2.40	0.002			
Platelet count ($\times 10^9/l$)	0.98	0.98, 0.99	<0.0001	0.987	0.98, 0.99	<0.001
AST/platelet	2.23	1.72, 3.00	<0.0001			
Hyperglycemia	5.86	3.79, 9.19	<0.0001	3.12	1.77, 5.51	<0.001
Diabetes mellitus	4.19	2.71, 6.50	<0.0001			
Hypertriglyceridemia	0.65	0.43, 0.99	0.03			
Low HDL cholesterol	2.10	1.35, 3.25	0.0003			
Hypertension	2.61	1.70, 4.04	<0.0001			
HOMA-IR (continuous)	1.11	1.05, 1.18	0.0001			
Biopsy length (mm)	1.003	0.97, 1.04	0.9			
Portal areas (n)	1.03	0.92, 1.14	0.6			

NOTE. Based on the results of the univariate logistic models, and in order to avoid overworking the multivariate model, we chose within each set of variables that provide the same clinical information (i.e., BMI continuous versus obesity versus categories of BMI versus waist circumference versus central obesity versus waist-to-hip ratio; and hyperglycemia versus diabetes) the one variable that reduced the deviance (or variation in a logistic model) the most. By doing so, the variable that improved prediction of advanced fibrosis was included in the multivariate model. These variables were BMI continuous and hyperglycemia. The odds ratio obtained by multivariate analysis are adjusted for site.

(7%) of 295 patients with a low cutoff point were incorrectly staged (Table 3). Thus, using this low cutoff point, the absence of advanced fibrosis could be excluded with high accuracy (negative predictive value of 93%).

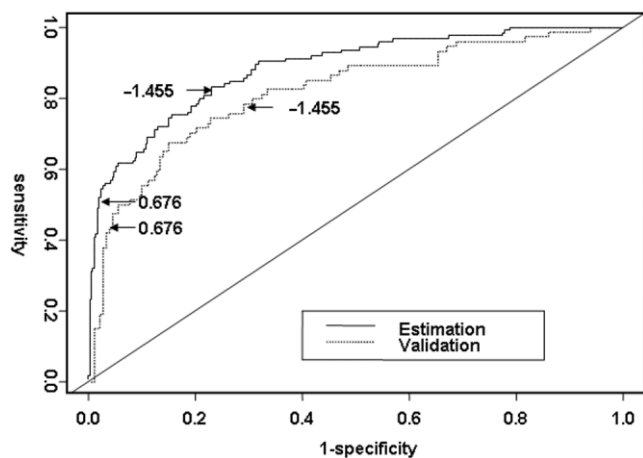


Fig. 1. ROC curves of the scoring system in the estimation ($n = 480$) and validation ($n = 253$) groups combining 6 variables (age, BMI, hyperglycemia/diabetes, AST/ALT ratio, platelet count, and albumin) to distinguish between NAFLD patients with and without advanced fibrosis. The area under the ROC curve for the estimation and validation groups is 0.88 ± 0.02 (95% confidence intervals, 0.85, 0.92) and 0.82 ± 0.03 (95% confidence intervals, 0.76, 0.88) respectively.

By applying the high cutoff point (score above 0.676), 64 (50%) of 125 patients with advanced fibrosis were correctly identified, whereas only 7 (10%) of the 71 with a high cutoff point were incorrectly staged (Table 3). Using this high cutoff point, the presence of advanced fibrosis could be diagnosed with high accuracy (positive predictive value of 90%).

Overall, in the estimation group, the model predicted the presence or absence of advanced fibrosis in $(295 + 71)/480 = 76\%$ of patients with a correct prediction in $337/366$ or 92% [or 70% ($337/480$) of the total]. The incorrect prediction rate in the estimation group was only $(22 + 7)/366 = 7.9\%$. Thus, by applying the model to the estimation group, a liver biopsy would have been avoided in 366 (76%) patients and would be performed in only 114 (24%) of the 480 patients identified as “indeterminate”.

Validation of Results. The diagnostic accuracy of the scoring system in separating patients with and without advanced fibrosis was cross-validated in a separate set of 253 patients. The area under the ROC curve remained high in the validation set [0.82 ± 0.03 (95% CI = 0.76, 0.88)], and also after 20-fold cross-validation [0.84 ± 0.02 (95% CI = 0.81, 0.88)] (Figs. 1 and 2, respectively).

Table 3. Predictive Value of the Scoring System Obtained from the Estimation Group (n = 480)

	Low cutoff point (< -1.455)	Indeterminate ($-1.455-0.676$)	High cutoff point (> 0.676)	Total
Total	295	114	71	480
No significant fibrosis (stage 0-2)	273	75	7	355
Significant fibrosis (stage 3-4)	22	39	64	125
Sensitivity	82%		51%	
Specificity	77%		98%	
Positive predictive value	56%		90%	
Negative predictive value	93%		85%	
Likelihood ratio (+)	3.567		25.966	
Likelihood ratio (-)	0.229		0.498	
Interpretation	Absence of significant fibrosis (93% certainty)		Presence of significant fibrosis (90% certainty)	

NOTE. Prevalence of advanced fibrosis of 26% in the estimation group.

By applying the low cutoff point (score below -1.455), 127 (71%) of the 179 patients without advanced fibrosis were correctly identified, whereas 17 (12%) of 144 with a low cutoff point were incorrectly staged (Table 4). Thus, using this low cutoff point, the absence of advanced fibrosis could be excluded with high accuracy (negative predictive value of 88%).

By applying the high cutoff point (greater than 0.676), 32 (43%) of the 74 patients with advanced fibrosis were correctly identified, whereas only 7 (18%) of the 39 patients with a high cutoff point were incorrectly staged (Table 4). By using this high cutoff point, the presence of advanced fibrosis could be diagnosed with high accuracy (positive predictive value of 82%).

Overall, in the validation group, the model identified presence or absence of advanced fibrosis in $(144 + 39)/253 = 72\%$ of patients with a correct prediction in 159/

183 = 87% [or 63% (159/253) of the total]. The incorrect prediction rate in the validation group was only $(17 + 7)/183 = 13.1\%$. Thus, by applying the model to the validation group, a liver biopsy would have been avoided in 183 (72%) patients and would be performed in only 70 (28%) of the 253 patients identified as "indeterminate".

Predictive Values of the Model for Different Prevalence of Significant Fibrosis. The prevalence of advanced fibrosis in the 4 centers was 12.5% (Italy), 19.6% (Newcastle), 30.9% (Sydney), and 29.8% (Mayo). Therefore, we calculated positive and negative predictive values of the 2 cutoff points using a wide range of prevalence of advanced fibrosis varying from 5% to 50%. The NPV of the low cutoff point to rule out advanced fibrosis remained high ($\geq 87\%$, Table 5). The PPV of the high cutoff point to diagnose advanced fibrosis also remained high, particularly for prevalence of advanced fibrosis of 10% or more ($\geq 78\%$, Table 5). Thus, these 2 cutoff points may be useful to predict the severity of liver fibrosis in patients with NAFLD seen in medical centers with different prevalences of advanced fibrosis. The estimated range of the NAFLD fibrosis score is shown in Table 6. Across the full range of the probability of stage 0-2 fibrosis (0% to 100%), the NAFLD fibrosis score is translated to a negative, positive, or indeterminate result.

Discussion

In this study, we developed and validated a simple noninvasive scoring system composed of routinely measured and easily available clinical and laboratory variables to discriminate between the presence or absence of advanced fibrosis in patients with NAFLD. This index, which we call the "NAFLD fibrosis score", was accurate in distinguishing the severity of fibrosis. Using values below the lower or above the higher cutoff points, a prediction of absence or presence of advanced fibrosis was made in 549

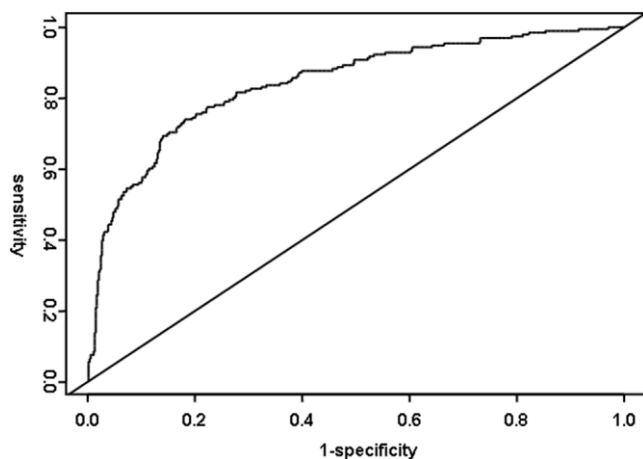


Fig. 2. ROC curve of 20-fold cross-validation on all data (n = 733). Data was randomly divided into 20 balanced subsets. One subset was removed and a model was generated using multivariate backward stepwise logistic regression as described in "Statistical Analysis". The generated model was then used to fit each observation in the removed subset. This was repeated until all 20 subsets were fit. The area under the ROC curve is 0.84 ± 0.02 (95% CI 0.81, 0.88).

Table 4. Predictive Value of the Scoring System Obtained from the Validation Group (n = 253)

	Low Cutoff Point (< -1.455)	Indeterminate ($-1.455-0.676$)	High Cutoff Point (> 0.676)	Total
Total	144	70	39	253
No significant fibrosis (stage 0-2)	127	45	7	179
Significant fibrosis (stage 3-4)	17	25	32	74
Sensitivity	77%		43%	
Specificity	71%		96%	
Positive predictive value	52%		82%	
Negative predictive value	88%		80%	
Likelihood ratio (+)	2.652		11.058	
Likelihood ratio (-)	0.324		0.591	
Interpretation	Absence of significant fibrosis (88% certainty)		Presence of significant fibrosis (82% certainty)	

NOTE. Prevalence of advanced fibrosis of 29% in the validation group.

(75%) of the 733 patients of the total cohort, and this prediction was correct in 496 (90%) of these 549 individuals. Only 184 (25%) patients of the total cohort of 733 were considered "indeterminate". This implies that by applying the NAFLD fibrosis score, liver biopsy could have been avoided in 75% (549 of 733) of patients in the total cohort.

The potential diagnostic accuracy of serum markers of fibrosis has been evaluated by others.¹⁹⁻²¹ The European liver fibrosis group assessed the combination of age and serum levels of hyaluronic acid, aminoterminal propeptide of type 3 collagen, and tissue inhibitor of matrix metalloproteinase 1 in predicting advanced fibrosis in patients with a wide range of liver disease.¹⁹ The proposed algorithm had an acceptable accuracy overall, but only 61 of the 912 patients studied had NAFLD, a number that is too small to derive meaningful conclusions for the NAFLD population. Further, the lack of availability of these serum markers of fibrosis in most centers makes it difficult to apply the proposed scoring system on a daily basis. A French group recently reported their experience with the FibroTest, a combination of 5 serum markers, in the prediction of advanced fibrosis (stages 3-4) in 267

patients with NAFLD.²² In that study,²² the FibroTest value was in between the proposed cutoffs of 0.30 and 0.70 in 88 patients, and thus, unable to predict the presence or absence of advanced fibrosis in 33% of the patients, a proportion slightly higher than the 25% of indeterminate cases in our series.

The combination of serum levels of hyaluronic acid and type 6 collagen 7S domain²⁰ and serum YKL-40²¹ levels also have been proposed as a diagnostic tool on the basis of small studies that lacked a validation group. More recently, liver stiffness measured by the ultrasonographic-based FibroScan has been proposed as a useful noninvasive method for the identification of advanced liver fibrosis in patients with chronic hepatitis C infection,^{26,27} A recent prospective study of 2,114 FibroScan examinations, however, found presence of BMI greater than 28 kg/m² to be the only independent factor associated with failure of FibroScan examination for the identification of liver fibrosis.²⁸ BMI of 28 kg/m² or greater is almost a universal finding in patients with NAFLD and thus the FibroScan's utility in fibrosis quantification in patients with NAFLD needs further evaluation.

Table 5. Predictive Values of the Cutoff Points for Different Prevalences of Advanced Fibrosis (n = 733)

Prevalence of Significant Fibrosis (%)	Lower Cutoff Value (< -1.455)		Higher Cutoff Value (> 0.676)	
	PPV (95% CI)	NPV (95% CI)	PPV (95% CI)	NPV (95% CI)
5	19 (14-25)	98 (97-100)	64 (53-75)	97 (95-98)
10	33 (26-39)	97 (95-99)	78 (68-87)	94 (91-96)
15	42 (35-49)	96 (93-98)	84 (76-93)	91 (88-94)
20	49 (42-56)	94 (91-97)	88 (80-95)	88 (85-91)
25	55 (47-62)	93 (90-96)	90 (83-97)	86 (82-89)
30	59 (52-66)	92 (88-95)	91 (85-98)	83 (80-87)
35	63 (56-70)	90 (87-94)	92 (86-99)	81 (77-85)
40	66 (59-73)	89 (85-93)	93 (88-99)	79 (75-83)
45	68 (62-75)	88 (84-92)	94 (89-100)	77 (73-81)
50	71 (64-77)	87 (83-90)	95 (89-100)	75 (71-79)

Table 6. Range of the NAFLD Fibrosis Score

NAFLD Fibrosis Score	Probability of Fibrosis (Stage 0-2) and 95% CI
Negative result (>0.676)	
14.35	0 (0.0, 0.0)
2.71	10 (7, 17)
1.84	20 (14, 27)
1.17	30 (24, 37)
Indeterminate result (< -1.455->0.676)	
0.63	40 (34, 47)
0.12	50 (44, 56)
-0.37	60 (55, 65)
-0.92	70 (66, 74)
Positive result (< - 1.455)	
-1.57	80 (76,.83)
-2.56	90 (87, 92)
-6.45	100 (99, 100)

Our results suggest that by applying the NAFLD fibrosis score, a liver biopsy to determine severity of fibrosis would be required in only 25% of patients with NAFLD, that is, those identified as “indeterminate”. Most importantly, considering that most patients with NAFLD seen in clinical practice do not have advanced fibrosis [73% (534/733) of our cohort], the lower cutoff point was particularly accurate in ruling out the presence of advanced fibrosis; the NPV was 93% and 88% in the estimation and validation groups, respectively, and ranged from 87% to 98% for the prevalence of advanced fibrosis of 5% to 50%. Among our 733 patients, 439 (60%) had a negative diagnosis of advanced fibrosis (score below -1.455), and thus a liver biopsy would have been avoided by applying the NAFLD fibrosis score; of these 439, 400 (91%) indeed had stage 0-2 fibrosis.

Our study has several unique features. First, we included the largest cohort of patients with liver biopsy-proven NAFLD ever reported. Second, patients were untreated, consecutively biopsied individuals with NAFLD seen in different areas of the world, and thus, the population includes a large variety of ages and ethnic backgrounds, and a wide-ranging prevalence of fibrosis severity. Third, our predictive model consists of objective clinical and readily available laboratory variables that are routinely determined in patients with NAFLD in clinical practice, and no additional tests are necessary. Fourth, given the uneven distribution of fibrosis in the liver in patients with NAFLD,²⁹ fibrosis severity predicted by our model may be a better reflection of fibrosis severity in the whole organ than fibrosis stage determined by liver biopsy.

Our study has some limitations. First, we included patients from different centers in the world that have a particular interest in studying NAFLD, and thus, some referral bias cannot be ruled out. Second, although we used a well-defined and acceptable scoring system to stage

liver fibrosis, liver biopsies were read by independent liver pathologists at each center, and we were not able to quantify the effect on our results of some intraobserver and interobserver variability in fibrosis staging. However, quantification of fibrosis in patients with NAFLD has the lowest intraobserver and interobserver variability as compared to any other histological features.²⁵ Further, it has been suggested that the level of pathologist experience has more influence on agreement than the characteristics of the biopsy specimens.³⁰ Third, 90% of our patient population was Caucasian and 98% were 21 years of age or older. Fourth, we acknowledge that the severity of fibrosis in our study was determined by means of a percutaneous liver biopsy, which is prone to sampling error.

Due to these limitations, our results need to be validated in independent patient populations by other investigators. Also, further studies are necessary to determine the potential utility of our model in children and adolescents and in non-Caucasian patients with NAFLD as well as in patients with persistently normal aminotransferases, and in following fibrosis progression. In addition, further studies are needed to determine the potential benefits of diagnosing advanced fibrosis with our model, such as reinforcement of lifestyle measures and enrollment of patients with advanced fibrosis into screening programs for early detection of HCC and esophageal varices.

In summary, we demonstrate that a NAFLD fibrosis score constructed from routine clinical and laboratory variables can accurately predict the presence or absence of advanced fibrosis in NAFLD, rendering liver biopsy unnecessary in the vast majority of patients. It has to be determined, however, whether the addition of serum markers of fibrosis or imaging modalities increases the diagnostic accuracy of the NAFLD fibrosis score.

References

- Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *HEPATOLOGY* 2004;40:1387-1395.
- Bedogni G, Miglioli L, Masutti F, Tiribelli C, Marchesini G, Bellentani S. Prevalence of and risk factors for nonalcoholic fatty liver disease: The Dionysos nutrition and liver study. *HEPATOLOGY* 2005;42:44-52.
- Fan JG, Zhu J, Li XJ, Chen L, Li L, Dai F, et al. Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *J Hepatol* 2005;43:508-514.
- Nomura H, Kashiwagi S, Hayashi J, Kajiyama W, Tani S, Goto M. Prevalence of fatty liver in a general population of Okinawa, Japan. *Jpn J Med* 1988;27:142-149.
- Caldwell SH, Oelsner DH, Iezzoni JC, Hespenheide EE, Battle EH, Driscoll CJ. Cryptogenic cirrhosis: clinical characterization and risk factors for underlying disease. *HEPATOLOGY* 1999;29:664-669.
- Bugianesi E, Leone N, Vanni E, Marchesini G, Brunello F, Carrucci P, et al. Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology* 2002; 123:134-140.

7. Adams LA, Lymp JF, St Sauver J, Sanderson SO, Lindor KD, Feldstein A, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterology*. 2005;129:113-121.
8. Sanyal AJ, Banas C, Sargeant C, Luketic VA, Sterling RK, Stravitz RT, et al. Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C. *HEPATOLOGY* 2006;43:682-689.
9. Teli MR, James OF, Burt AD, Bennett MK, Day CP. The natural history of nonalcoholic fatty liver: a follow-up study. *HEPATOLOGY* 1995;22:1714-1719.
10. Dam-Larsen S, Franzmann M, Andersen IB, Christoffersen P, Jensen LB, Sorensen TIA, et al. Long term prognosis of fatty liver: risk of chronic liver disease and death. *Gut* 2004;53:750-755.
11. Mendes FD, Suzuki A, Angulo P. Prevalence and indicators of portal hypertension in nonalcoholic fatty liver disease. *Gastroenterology* 2004;126:A759.
12. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Non alcoholic fatty liver disease: A spectrum of clinical and pathological severity. *Gastroenterology* 1999;116:1413-1419.
13. Ekstedt M, Franzen LE, Mathiesen UL, Thorelius L, Holmqvist M, Bode-mar G, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. *HEPATOLOGY* 2006;44:865-873.
14. Hui JM, Kench JG, Chitturi S, Sud A, Farrell GC, Byth K, et al. Long-term outcomes of cirrhosis in nonalcoholic steatohepatitis compared with hepatitis C. *HEPATOLOGY* 2003;38:420-427.
15. Angulo P, Keach JC, Batts KP, Lindor KD. Independent predictors of liver fibrosis in patients with nonalcoholic steatohepatitis. *HEPATOLOGY* 1999;30:1356-1362.
16. Ratziu V, Giral P, Charlotte F, Bruckert E, Thibault V, Theodorou I, et al. Liver fibrosis in overweight patients. *Gastroenterology*. 2000;118:1117-1123.
17. Dixon JB, Bhathal PS, O'Brien PE. Nonalcoholic fatty liver disease: predictors of nonalcoholic steatohepatitis and liver fibrosis in the severely obese. *Gastroenterology* 2001;121:91-100.
18. Shimada M, Hashimoto E, Kaneda H, Noguchi S, Hayashi N. Nonalcoholic steatohepatitis: risk factors for liver fibrosis. *J Hepatol* 2002;24:429-438.
19. Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, et al. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004;127:1704-1713.
20. Sukugagwa H, Nakayoshi T, Kobashigawa K, Yamashiro T, Maeshiro T, Miyagi S, et al. Clinical usefulness of biochemical markers of liver fibrosis in patients with nonalcoholic fatty liver disease. *World J Gastroenterol* 2005;11:255-259.
21. Johansen JS, Christoffersen P, Moller S, Price PA, Henriksen JH, Garbarsch C, et al. Serum YKL-40 is increased in patients with hepatic fibrosis. *J Hepatol* 2000;32:911-920.
22. Ratziu V, Massard J, Charlotte F, Messous D, Imbert-Bismut F, Bonyhay L, et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterology* 2006;6:6.
23. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostatic model assessment: insulin resistance and beta-cell function from fasting glucose and insulin concentrations in man. *Diabetologia* 1985;28:412-419.
24. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *HEPATOLOGY* 2003;38:518-526.
25. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *HEPATOLOGY* 2005;41:1313-1321.
26. Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *HEPATOLOGY* 2005;41:48-54.
27. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343-350.
28. Foucher J, Castera L, Bernard PH, Adhoue X, Laharie D, Bertet J, et al. Prevalence and factors associated with failure of liver stiffness measurement using FibroScan in a prospective study of 2114 examinations. *Eur J Gastroenterol Hepatol* 2006;18:411-412.
29. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005;128:1898-1906.
30. Roussellet MC, Michalak S, Dupre F, Croue A, Bedosa P, Saint-Andre JP, et al. Source of variability in histological scoring of chronic viral hepatitis. *HEPATOLOGY* 2005;41:257-264.