Liver INTERNATIONAL
OFFICIAL JOURNAL OF THE INTERNATIONAL ASSOCIATION FOR THE STUDY OF THE LIVER

Liver International ISSN 1478-3223

VIRAL HEPATITIS

Liver fibrosis staging through a stepwise analysis of non-invasive markers (FibroSteps) in patients with chronic hepatitis C infection

Samer S. El-Kamary[1,2,3], Mona M. Mohamed[4], Maissa El-Raziky[5], Michelle D. Shardell[1], Olfat G. Shaker[6], Wafaa A. ElAkel[5] and Gamal Esmat[5]

1 Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA
2 Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA
3 Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD, USA
4 Department of Zoology, Faculty of Science, Cairo University, Giza, Egypt
5 Endemic Medicine and Hepatology Department, Faculty of Medicine, Cairo University, Cairo, Egypt
6 Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Cairo University, Cairo, Egypt

Keywords
chronic hepatitis – fibrosis markers – hepatitis C virus – liver fibrosis – logistic regression

Correspondence
Samer El-Kamary, MBChB, MS, MPH, Department of Epidemiology and Public Health, Department of Pediatrics, Center for Vaccine Development, University of Maryland School of Medicine, 660 W Redwood St, HH102B Baltimore, MD 21201, USA
Tel: +1 410 706 2228
Fax: +1 410 706 8013
e-mail: selkamar@epi.umaryland.edu

Received 29 April 2012
Accepted 13 February 2013

DOI:10.1111/liv.12139

Abstract

Background: Non-invasive fibrosis markers can distinguish between liver fibrosis stages in lieu of liver biopsy or imaging elastography. *Aims:* To develop a sensitive, non-invasive, freely-available algorithm that differentiates between individual liver fibrosis stages in chronic hepatitis C virus (HCV) patients. *Methods:* Chronic HCV patients (n = 355) at Cairo University Hospital, Egypt, with liver biopsy to determine fibrosis stage (METAVIR), were tested for preselected fibrosis markers. A novel multistage stepwise fibrosis classification algorithm (FibroSteps) was developed using random forest analysis for biomarker selection, and logistic regression for modelling. FibroSteps predicted fibrosis stage using four steps: Step 1 distinguished no (F0)/mild fibrosis(F1) vs. moderate(F2)/severe fibrosis(F3)/cirrhosis(F4); Step 2a distinguished F0 vs. F1; Step 2b distinguished F2 vs. F3/F4; and Step 3 distinguished F3 vs. F4. FibroSteps was developed using a randomly-selected training set (n = 234) and evaluated using the remaining patients (n = 118) as a validation set. *Results:* Hyaluronic Acid, TGF-β1, α2-macroglobulin, MMP-2, Apolipoprotein-A1, Urea, MMP-1, alpha-fetoprotein, haptoglobin, RBCs, haemoglobin and TIMP-1 were selected into the models, which had areas under the receiver operating curve (AUC) of 0.973, 0.923 (Step 1); 0.943, 0.872 (Step 2a); 0.916, 0.883 (Step 2b) and 0.944, 0.946 (Step 3), in the training and validation sets respectively. The final classification had accuracies of 94.9% (95% CI: 91.3–97.4%) and 89.8% (95% CI: 82.9–94.6%) for the training and validation sets respectively. *Conclusions:* FibroSteps, a freely available, non-invasive liver fibrosis classification, is accurate and can assist clinicians in making prognostic and therapeutic decisions. The statistical code for FibroSteps using R software is provided in the supplementary materials.

Liver biopsy is still considered as the gold standard for determining the degree of liver fibrosis (1), despite being an invasive procedure with morbidity of 0.3–0.6% and mortality of 0.05% (2). Furthermore, because of heterogeneity in the distribution of pathological changes in the liver (3), liver biopsy may accurately stage fibrosis in only 80% of cases and miss advanced fibrosis in as many as 30% of patients (4). Moreover, patients may not be willing, or have contraindications against, repeated biopsies to monitor progression.

To overcome these shortfalls, several non-invasive methods for staging liver fibrosis have been developed using serum markers, and more recently, imaging techniques, (5–10) many of which are specific for chronic hepatitis C virus (HCV) infection (5, 6, 11–13). Early studies focused on using biomarkers to detect only very severe fibrosis or cirrhosis,[reviewed in (8)] whereas recent efforts have focused on distinguishing fibrosis stage across the severity spectrum (14–16).

Imaging techniques to measure elasticity and stiffness in the liver tissue, such as ultrasound (FibroScan® Echosens, Paris, France) and magnetic resonance elastography were developed more recently to overcome some of these shortfalls (17, 18). While both methods have excellent inter-rater reliability, FibroScan® can be less reliable in older and obese patients. (17, 19–22) Like many of the best-performing fibrosis markers that are proprietary, both of these imaging techniques are

expensive and can be cost prohibitive in resource-limited settings.

Attaining an accurate fine-grained classification is important for clinicians to determine patients' rate of fibrosis progression while waiting for treatment, or to check for signs of fibrosis regression after successful therapy. In this study, we aim to develop a freely available, open-source, multi-stage stepwise classification model (FibroSteps) using a set of feasible, commercially available laboratory tests and fibrosis markers to predict hepatic fibrosis stages. We develop and validate the classification in chronic HCV-infected patients in Egypt where an estimated 14.7% of the population or 11 million people are antiHCV-positive, the highest prevalence in the world (23, 24).

Methods

Study population

We enrolled chronic HCV patients from both sexes, ages 18–60 years old, who were scheduled to receive antiviral therapy at the Endemic Medicine and Hepatology Department, Faculty of Medicine, Cairo University, Egypt, after providing written informed consent. HCV diagnosis was established by a positive HCV Ribonucleic Acid (RNA) using an in-house direct reverse transcriptase polymerase chain reaction (RT-PCR) assay (25). Patients were excluded if they had HIV, chronic hepatitis B, decompensated cirrhosis, or previously received interferon therapy. All enrolled patients had a pretreatment liver biopsy within the prior 2 months and a blood sample was drawn to test for fibrosis biomarkers. This study was performed in accordance with ethical standards of the Helsinki declaration of 1975, as revised in 1983, and the protocol was approved by the institutional review boards of Cairo University Faculty of Medicine, and the University of Maryland School of Medicine, Baltimore, Maryland, USA.

Liver fibrosis predictors

Social, demographical and HCV-related risk factors were determined by interviewing patients using a standardized questionnaire as previously described (5). A combination of routine clinical laboratory tests were measured using standard methods including platelet count, white blood cells (WBCs), red blood cells (RBCs), haemoglobin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALK), serum albumin, prothrombin concentration, total bilirubin, alpha-foetoprotein (AFP), thyroid stimulating hormone (TSH), creatinine, urea and blood glucose.

Fibrosis markers

Fibrosis biomarkers selected for this study were matrix metalloproteinase-1 (MMP-1, Calbiochem, La Jolla, California: Cat No. QIA55), MMP-2 (Calbiochem, La Jolla, California: Cat No. QIA63), hyaluronic acid (HA, Corgenix Inc, Broomfield, Colorado, USA), tissue growth factor beta-1 (TGF-β1, Quantikine, R&D Systems, Inc. Minneapolis, USA), tissue inhibitor of metalloproteinase-1 (TIMP-1, Calbiochem, La Jolla, California: Cat No. QIA49), α2-macroglobulin (α2M, Biocientífica S.A. Buenos Aires, Argentina), haptoglobin (Hapg, Biocientífica S.A. Buenos Aires, Argentina), and apolipoprotein-A1 (ApoA1, Biocientífica S.A. Buenos Aires, Argentina). We selected these specific markers based on their use in several studies for predicting histological staging in patients with chronic HCV infection (5, 26–35).

Histological staging and grading

The needle-biopsy liver specimens were at least 1.5 cm long with a minimum of six portal tracts. They were read by a single hepatopathologist, a professor in the pathology department at the Cairo University Faculty of Medicine, who is also a consultant pathologist for the Egyptian Ministry of Health National Hepatitis C Treatment Program. He was unaware of the patient's clinical or laboratory data; and classified the fibrosis according to the METAVIR scoring system following the criteria of the METAVIR Cooperative Study Group (36). Liver fibrosis was scored as follows: F0, no fibrosis; F1, portal fibrosis without septa (mild); F2, portal fibrosis with rare septa (moderate); F3, numerous septa without cirrhosis (severe); and F4, cirrhosis.

Statistical analysis

We developed a step-wise model-building and classification framework consisting of four steps: Step 1 distinguished F0/F1 (no or mild fibrosis) from F2/F3/F4 (clinically significant fibrosis); Step 2a distinguished F0 from F1; Step 2b distinguished F2 from F3/F4; and Step 3 distinguished F3 from F4 (Fig. 1). We *a priori* chose this step-wise algorithm to mimic the clinical decision-making context, and to allow the biomarkers and their function to differ by step. The statistical analysis was comprised of four phases: 1) variable selection for each step, 2) model-building for each step, 3) stage classification and 4) validation. We divided the dataset ($n = 355$) into a training set ($n = 237$) and a validation set ($n = 118$), a 2/3:1/3 split. The training and validation sets were compared using the Wilcoxon rank-sum test for continuous variables and Pearson's chi-square test for categorical variables. We also performed a descriptive analysis in the training set by comparing the METAVIR fibrosis stages using the Kruskal–Wallis test for continuous variables and Fisher's exact test for categorical variables. $P < 0.05$ was considered statistically significant for all tests. All analyses were performed using R statistical software version 2.15.0 (www.r-project.org). Statistical code using the free R software is provided

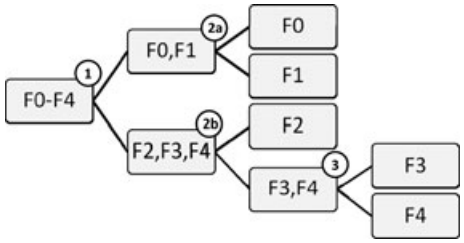

Fig. 1. The final model consisted of four steps: Step 1 differentiated between no/mild fibrosis and clinically significant fibrosis (F0, F1 vs. F2, F3, F4); Step 2a differentiated between no and mild fibrosis (F0 vs. F1); Step 2b differentiated between moderate and severe fibrosis/cirrhosis (F2 vs. F3 or F4); and Step-3 between severe fibrosis and cirrhosis (F3 vs. F4).

in the supplementary materials section to enhance accessibility to FibroSteps, particularly in resource-limited settings.

Variable selection

For each step, we performed a nonparametric random forest analysis (37) of the training set to select candidate biomarkers. A random forest is an ensemble of classification and regression trees (CART) (38). CART recursively partitions a dataset into mutually exclusive nodes by dichotomizing variables, where participants within a region are as similar as possible with respect to probabilities for outcome class (in this case fibrosis stage) membership. In a random forest, a tree is grown using a boot-strap sample of the data, and each node split is based on a random subsample of candidate variables. We chose this approach to avoid strong modelling assumptions that may be prone to mis-specification. We grew 1000 trees and ranked the variables according to their magnitude improvement in classification accuracy (called variable importance) after accounting for other variables and potential multi-way interactions. We erred on the side of inclusivity by retaining all variables that resulted in at least 0.5% improvement in accuracy with the option of potentially dropping some weak predictors at the model-building phase. Next, we reran the random forest algorithm with 1) the retained variable set, 2) the retained variable set minus the least important retained variable, and 3) the retained variable set plus the most important variable that was not retained. If the second random forest resulted in the highest accuracy, we reran the random forest algorithm minus the second-least important retained variable, and repeated this approach until the maximum accuracy was reached. If, however, the third random forest resulted in the highest accuracy, we reran the random forest algorithm adding the second-most important predictor that was not retained, and we repeated until the highest accuracy was reached. To perform this analysis we used the RandomForest package in R statistical software.

Table 1. Demographics, liver biopsy findings and laboratory test results for patients in the training and validation groups (*N = 355*)

Characteristic	Training Set *n* = 237	Validation Set *n* = 118	*P*-value*
Age (years), mean (SD)	39.6 (9.9)	40.7 (9.6)	0.35
Male sex, *n* (%)	186 (78.5)	90 (76.3)	0.74
METAVIR Fibrosis Score, *n* (%)			
No fibrosis (F0)	18 (7.6)	15 (12.7)	0.59
Portal fibrosis without septa (F1)	93 (39.2)	41 (34.7)	
Portal fibrosis with rare septa (F2)	62 (26.2)	29 (24.6)	
Numerous septa without cirrhosis (F3)	44 (18.6)	22 (18.6)	
Cirrhosis (F4)	20 (8.4)	11 (9.3)	
Biomarkers, mean (SD)			
α2-macroglobulin (g/L)	2.11 (0.95)	2.18 (0.88)	0.27
ApoA1 (g/L)	1.57 (0.66)	1.58 (0.69)	0.84
Hyaluronic Acid (ng/mL)	104.6 (114.1)	91.90 (83.1)	0.83
Haptoglobin (g/L)	0.95 (0.47)	0.96 (0.51)	0.43
MMP-1 (ng/mL)	4.36 (4.19)	4.74 (4.45)	0.54
MMP-2 (ng/mL)	498.8 (257.0)	461.2 (214.2)	0.27
TGF-β1 (ng/L)	78.5 (24.4)	78.7 (29.4)	0.82
TIMP-1 (ng/mL)	210.2 (65.8)	210.1 (60.9)	0.68
Laboratory tests, mean (SD)			
Urea (mg/dL)	33.18 (8.55)	34.39 (8.60)	0.23
Platelet count (10^3/μL)	215.2 (60.2)	206.4 (57.5)	0.20
Albumin (g/L)	42.86 (4.61)	42.54 (3.80)	0.29
TSH (mU/L)	2.36 (5.91)	1.81 (1.15)	0.74
Total Bilirubin (umols/L)	9.46 (11.80)	8.30 (3.32)	0.84
ALT (U/L)	73.61 (49.60)	82.59 (69.39)	0.16
AST (U/L)	59.75 (38.36)	62.81 (39.85)	0.32
Alkaline phospatase (U/L)	109.1 (58.6)	103.8 (51.2)	0.42
Alpha-foetoprotein (ng/mL)	9.21 (27.39)	7.68 (12.22)	0.82
Glucose (mg/dL)	94.45 (19.05)	91.71 (14.03)	0.29
RBCs (10^6/μL)	5.23 (0.47)	5.32 (0.46)	0.12
Haemoglobin (g/dL)	14.56 (1.46)	14.43 (1.54)	0.42
WBC (10^3/μL)	6.49 (2.10)	6.41 (1.80)	0.91
Creatinine (mg/dL)	0.90 (0.28)	0.89 (0.17)	0.91
Prothrombin concentration (%)	86.46 (10.68)	88.44 (11.15)	0.06

**P*-value from Wilcoxon rank-sum test (continuous variables) and chi-square test (categorical variables).

SD, Standard Deviation; α2-macroglobulin, alpha-2 macroglobulin; ApoA1,apolipoproteinA1; MMP-1, matrix metalloproteinase-1; MMP-2, matrix metalloproteinase-2; TGF-β1, tissue growth factor beta-1; TIMP-1, tissue inhibitor of metalloproteinase-1; TSH, thyroid stimulating hormone; ALT, alanine aminotransferase; AST, aspartate aminotransferase; RBC, red blood cells; WBC, white blood cells.

Model building

Next, for each step, to explore the functional form of the selected variables, we fit a multivariate adaptive regression spline (MARS) using the final set of variables from the random forest analysis (39). MARS is a generalization of CART that can model spline terms,

continuous terms, transformations, and higher-order terms. We added the following transformations of the selected variables as candidate predictors: natural log, natural log-natural log, square-root, square and cube. We allowed interactions up to three-way interactions. The final set of terms, including splines, interactions and transformations, were then included in a logistic regression model fit using maximum likelihood. We chose this approach to leverage the flexibility of MARS (which can assess many candidate predictors while mitigating model mis-specification) with the parsimony and well-known theory of logistic regression. Once again, variable importance in the MARS model was ranked, where some variables were unused. We refit the logistic regression model excluding the least important used variable in the MARS model, and refit the logistic regression model including the most-important unused variable in the MARS model. We continued until we found the model with the smallest Akaike's Information Criterion (AIC). We targeted AIC to ensure parsimony (to avoid over-fitting). Therefore, some variables retained in the variable selection phase had the potential to be dropped.

Discrimination was assessed at each step using likelihood ratio chi-square tests (*P*-value, smaller is better), area under the receiver operating characteristic curve (AUC, range 0–1, larger is better), and R^2 (range 0–1, larger is better). Calibration was assessed at each step using the Brier score (range 0–1, smaller is better) and the Hosmer–Lemeshow goodness of fit test (*P*-value, larger is better). Discrimination of all five stages using the four models was assessed using the Obuchowski index (range 0–1, larger is better) (40). Confidence intervals were calculated using the bootstrap with 1000 samples. In R, we used the earth package to perform MARS, the pROC package to calculate AUCs, and the rms package to calculate likelihood ratio and Hosmer-Lemeshow tests, R^2, and Brier scores.

Classification

Once the model building was complete, we used the predicted probabilities from each step to calculate the probabilities of F0, F1, F2, F3 and F4 for each participant. We classified participants according to the following algorithm: If any of these probabilities exceeded 0.9, then the patient was classified into the corresponding stage. For example if P(F0) >0.9, then the patient was classified as F0, where P(Fx) denotes the probability of Fx. If none of the above single-stage probabilities

Table 2. Biomarkers by METAVIR score in training set (n = 237)

Characteristic, mean (SD)*	F0 (n = 18)	F1 (n = 93)	F2 (n = 62)	F3 (n = 44)	F4 (n = 20)	*P*-value**
Age (years)	35.7 (8.9)	37.4 (9.7)	41.7 (8.5)	41.7 (11.1)	41.4 (10.7)	0.01
Male sex, n (%)	14 (77.8)	73 (78.5)	52 (83.9)	30 (68.2)	17 (85.0)	0.39
α2-macroglobulin (g/L)	1.78 (1.00)	1.75 (0.84)	2.41 (1.11)	2.64 (0.68)	1.99 (0.40)	<0.001
Apoliprotein-A1 (g/L)	1.63 (0.40)	1.76 (0.77)	1.53 (0.65)	1.35 (0.50)	1.22 (0.10)	<0.001
Hyaluronic Acid (ng/mL)	24.26 (17.77)	42.29 (60.48)	119.78 (85.16)	184.97 (100.64)	242.46 (191.58)	<0.001
Haptoglobin (g/L)	0.81 (0.24)	1.06 (0.53)	0.76 (0.38)	0.95 (0.46)	1.15 (0.42)	<0.001
MMP-1 (ng/mL)	3.47 (1.95)	5.71 (4.89)	3.76 (3.38)	2.90 (3.20)	3.90 (4.81)	<0.001
MMP-2 (ng/mL)	330.3 (293.7)	435.3 (238.2)	512.6 (232.3)	539.4 (206.2)	813.9 (207.5)	<0.001
TGF-β1 (ng/L)	73.45 (22.57)	68.64 (19.35)	93.01 (15.93)	90.53 (28.58)	57.22 (21.42)	<0.001
TIMP-1 (ng/mL)	170.5 (48.0)	194.1 (67.0)	210.8 (53.5)	235.9 (60.5)	263.0 (72.1)	<0.001
Urea (mg/dL)	33.06 (9.53)	31.22 (7.78)	35.74 (9.09)	32.88 (8.26)	35.06 (8.38)	0.02
Platelet count (10^3/μL),	232.5 (53.7)	218.7 (63.2)	208.3 (57.1)	214.2 (58.0)	207.3 (66.5)	0.52
Albumin (g/L)	44.00 (5.21)	42.92 (4.87)	43.36 (4.10)	41.72 (4.69)	42.50 (3.99)	0.28
TSH (mU/L)	1.64 (0.76)	2.44 (8.70)	1.77 (1.29)	3.30 (4.59)	2.44 (3.15)	0.06
Total Bilirubin (umols/L)	7.47 (2.90)	8.70 (5.48)	11.31 (19.90)	9.63 (11.09)	8.70 (2.77)	0.50
ALT (U/L)	56.47 (30.98)	70.96 (50.24)	68.43 (32.11)	83.65 (56.37)	95.30 (76.51)	0.36
AST (U/L)	45.40 (17.19)	59.51 (34.60)	50.75 (25.00)	71.90 (57.53)	74.95 (40.91)	0.01
Alkaline phosphatase (U/L)	111.6 (42.9)	102.7 (58.9)	121.0 (60.4)	106.6 (60.1)	105.8 (60.6)	0.24
Alpha-fetoprotein (ng/mL)	5.73 (4.15)	8.25 (30.13)	13.24 (37.23)	9.15 (12.64)	4.44 (2.51)	0.001
Glucose (mg/dL)	86.4 (12.8)	95.7 (21.4)	91.7 (13.9)	96.0 (18.5)	100.7 (24.5)	0.07
RBCs (10^6/μL)	4.91 (0.28)	5.28 (0.43)	5.23 (0.51)	5.22 (0.52)	5.29 (0.46)	0.03
Haemoglobin (g/dL)	13.9 (1.4)	14.9 (1.3)	14.2 (1.8)	14.7 (1.2)	14.7 (1.3)	0.01
WBC (10^3/μL)	7.23 (2.06)	6.34 (2.33)	6.30 (1.88)	6.66 (2.03)	6.70 (1.73)	0.22
Creatinine (mg/dL)	0.86 (0.18)	0.92 (0.39)	0.92 (0.16)	0.88 (0.16)	0.87 (0.21)	0.68
Prothrombin concentration (%)	90.8 (9.9)	86.9 (10.6)	86.2 (10.6)	85.4 (11.1)	83.6 (10.7)	0.26

*Mean (standard deviation, SD), unless otherwise specified.

***P*-value from Kruskal-Wallis test or Fisher's exact test.

α2-macroglobulin, alpha-2 macroglobulin; MMP-1, matrix metalloproteinase-1; MMP-2, matrix metalloproteinase-2; TGF-β1, tissue growth factor beta-1; TIMP-1, tissue inhibitor of metalloproteinase-1; TSH, thyroid stimulating hormone; ALT, alanine aminotransferase; AST, aspartate aminotransferase; RBC, red blood cells; WBC, white blood cells.

exceeded 0.9, then we calculated the probabilities of F0/F1, F1/F2, F2/F3, F3/F4. If any of these probabilities exceeded 0.9, then the patient was classified into the combined stage with the highest probability. If none of the above single-stage or two-stage probabilities exceeded 0.9, then we calculated the probabilities of F1 ± 1, F2 ± 1, and F3 ± 1. The participant was classified into the group with the highest of these probabilities. The proposed algorithm produced the following new fibrosis classification: F0, F0/F1, F1, F1 ± 1, F1/F2, F2, F2 ± 1, F2/F3, F3, F3 ± 1, F3/F4 and F4. Performance of the FibroSteps classification algorithm was assessed by the per cent correctly classified (accuracy). Lastly, using only Step 1, we assessed the performance of classifying patients as F0/F1 if P(F0/F1) >0.9 using the negative predictive value (i.e. the ability of the Step 1 model to rule out clinically significant fibrosis (METAVIR $\geq$ F2). Confidence intervals were calculated using exact binomial methods.

Validation

We assessed the performance of the logistic regression models and the proposed classification algorithm in the validation set using AUC, Obuchowski index, per cent accuracy, and negative predictive value. We conducted and reported this study according to the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines (41, 42).

Results

A total of 355 patients were enrolled from the Endemic Medicine and Hepatology Department, Faculty of Medicine, Cairo University, Egypt, after providing written informed consent. Table 1 shows the demographics, liver biopsy findings, the selected biomarkers and laboratory test results for the training (n = 237) and validation sets (n = 118) of patients. The two groups were well matched in baseline characteristics. The overall prevalence of fibrosis was: F0, 9.3%; F1, 37.7%; F2, 25.6%; F3, 18.6%; F4, 8.7%.

Descriptive analyses in Table 2 showed that age, α2M, ApoA1, hyaluronic acid, haptoglobin, MMP-1, MMP-2, TGF-β1, TIMP-1, Urea, AST, AFP, RBCs and haemoglobin all significantly differed across fibrosis stages ($P < 0.05$). Each of these variables was selected for at least one step using the random forest analysis, but age and AST were excluded in the final logistic regression models, because they did not improve AIC as described above.

For Step 1 (F0/F1 vs. F2/F3/F4, training n = 237), the random forest found that HA, TGF-β1, α2M, ApoA1, MMP-2, TIMP-1, MMP-1, AFP, age and urea resulted in >0.5% improvement in accuracy, in order of greatest to smallest improvement. However, the final logistic regression model excluded TIMP-1 and age, leaving eight biomarkers for Step 1. The variables, coefficients and standard errors for this and the subsequent models (Step 2a, Step 2b, and Step 3) are shown in Table 3. Fig. 2 (top-left panel) shows that the model had very

Table 3. FibroSteps logistic regression coefficient estimates for each Step

Step	Coefficient	Estimate	SE
Step 1, F0/F1 vs. F2/F3/F4	Intercept	0.106	1.352
	Ln(TGF-β1) × Ln(ApoA1)	10.525	2.929
	Sqrt(α2M) × Ln(MMP-1)	−0.414	0.240
	Sqrt(α2M) × Ln(Urea)	0.734	0.253
	[Ln(HA) − 4.25] × [Ln(HA) $\geq$ 4.25]	8.333	1.584
	[Ln(HA) − 5.00] × [Ln(HA) $\geq$ 5.00]	−10.089	2.652
	[3.05 − Ln(AFP)] × [Ln(AFP) $\leq$ 3.05]	−1.126	0.335
	[1.75 −Ln(Ln(MMP-2))] × [Ln(Ln(MMP-2)) $\leq$ 1.75]	−18.360	8.177
	[Ln(ApoA1) − 0.07] × [Ln(ApoA1) $\geq$ 0.07]	−15.443	6.086
	[Ln(ApoA1) − 0.44] × [Ln(ApoA1) $\geq$ 0.44]	8.578	4.199
	Ln(ApoA1)	−40.948	13.358
Step 2a, F0 vs. F1	Intercept	120.252	313.648
	[Ln(Hapg)]^2	2.559	2.061
	[Ln(MMP-2)]^2	34.936	26.004
	[Ln(MMP-2)]^3	−2.062	1.478
	[Ln(RBC)]^3	49.294	24.388
	Ln(RBC) × Ln(HB)	−138.024	69.908
	Ln(HB)	234.148	113.297
	Ln(MMP-2)	−194.779	151.293
	Ln(Hapg)	2.972	2.203
Step 2b, F2 vs. F3/F4	Intercept	42.338	35.198
	[Ln(HA) − 4.84] × [Ln(HA) $\geq$ 4.84]	5.917	1.960
	[Ln(HA) − 5.42] × [Ln(HA) $\geq$ 5.42]	−7.688	3.842
	[Ln(TGF-β1) − 4.15] × [Ln(TGF-β1) $\geq$ 4.15]	−15.384	3.618
	[Ln(TGF-β1) − 4.44] × [Ln(TGF-β1) $\geq$ 4.44]	18.888	4.780
	Ln(MMP-1) × Ln(Hapg)	1.086	0.637
	[Ln(MMP-2)]^3	0.109	0.075
	Ln(HA)	−0.178	0.417
	Ln(MMP-1)	−0.484	0.309
	Ln(Hapg)	0.580	0.829
	Ln(TIMP-1)	1.694	1.125
	Ln(MMP-2)	−12.006	8.859
Step 3 F3 vs. F4	Intercept	−39.221	22.308
	Ln(MMP-2)	4.431	1.753
	Ln(TGF-β1)	6.150	5.650
	α2M	−2.20	0.941
	TGF-β1	−0.152	0.095

SE, standard error; Ln, Base-e log; Sqrt, square-root; TGF-β1, tissue growth factor beta-1; ApoA1, apolipoproteinA1; α2M, alpha-2 macroglobulin; MMP-1, matrix metalloproteinase-1; HA, hyaluronic acid; AFP, Alpha-foetoprotein; MMP-2, matrix metalloproteinase-2; Hapg, haptoglobin; RBC, red blood cells; HB, haemoglobin; TIMP-1, tissue inhibitor of metalloproteinase-1.

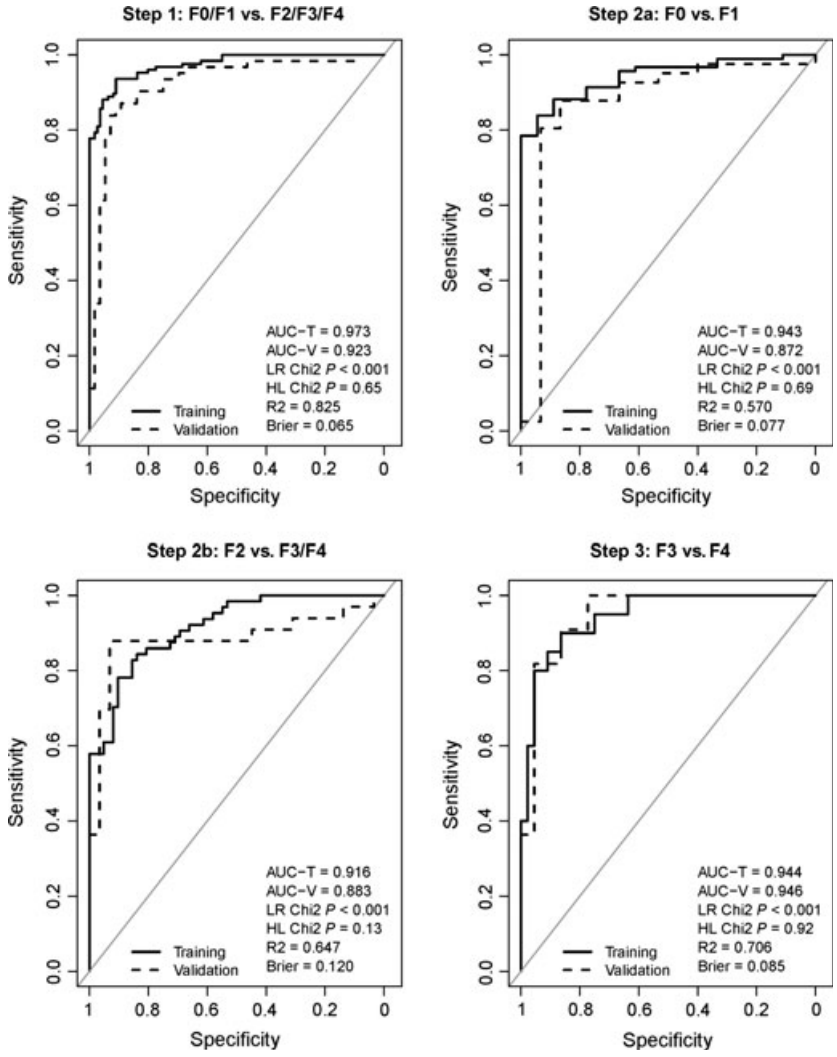

Fig. 2. Area under the receiver operating curve (AUC) diagnostic values of the FibroSteps logistic regression models for the training and validation groups predicting Step 1, no/mild (F0, F1) fibrosis vs. clinically significant fibrosis (F2, F3, F4); Step 2a, no (F0) vs. mild (F1) fibrosis; Step 2b, moderate (F2) vs. severe fibrosis/cirrhosis (F3, F4); and Step 3, severe fibrosis (F3) vs. cirrhosis (F4). AUC, Area under the curve. AUC-T, AUC in the training set; AUC-V, AUC in the validation set; LR, likelihood ratio; HL, Hosmer–Lemeshow; Chi^2, Chi-square.

high discrimination and calibration with AUCs of 0.973 (95% CI: 0.955, 0.987) and 0.923 (95% CI: 0.866, 0.971) in the training and validation sets respectively.

For Step 2a (F0 vs. F1, training with F0 or F1, n = 111), four biomarkers: RBCs, haemoglobin, MMP-2 and haptoglobin were selected using random forest and logistic regression. Fig. 2 (top-right panel) shows that this model had high discrimination and calibration with AUCs of 0.943 (95% CI: 0.897, 0.979) and 0.872 (95% CI: 0.724, 0.976) in the training and validation sets respectively.

For Step 2b (F2 vs. F3/F4, training with F2, F3, or F4, n = 126), six biomarkers: TGF-β1, HA, MMP-1, haptoglobin, TIMP-1, MMP-2 were selected using random forest and logistic regression. Fig. 2 (bottom-left panel) shows that the model also had high discrimination and calibration with AUCs of 0.916 (95% CI: 0.865, 0.957) and 0.883 (95% CI: 0.780, 0.969) in the training and validation sets respectively.

Lastly, for Step 3 (F3 vs. F4, training with F3 or F4 n = 64), three biomarkers: α2M, TGF-β1, MMP-2 were selected using random forest and logistic regression. Fig. 2 (bottom-right panel) shows that the model had very high discrimination and calibration with AUCs of 0.944 (95% CI: 0.881, 0.989) and 0.946 (95% CI: 0.860, 0.999) in the training and validation sets respectively.

The Obuchowski index for all five METAVIR stages was 0.983 (95% CI: 0.974, 0.989) and 0.947 (95% CI: 0.904, 0.977) in the training and validation sets respectively. Table 4 shows the true METAVIR stage and

Table 4. Accuracy of FibroSteps in classifying liver fibrosis stage

Dataset	Classification	METAVIR Fibrosis Stage F0	F1	F2	F3	F4	% Accuracy (95% CI)
Training	F0	2	1	0	0	0	94.9 (91.3, 97.4)
	F0/F1	11	18	1	0	0	
	F1	0	43	1	1	0	
	F1 ± 1	5	2	4	1	1	
	F1/F2	0	14	9	1	0	
	F2	0	0	12	0	0	
	F2 ± 1	0	13	5	5	3	
	F2/F3	0	0	26	17	0	
	F3	0	0	0	7	0	
	F3 ± 1	0	2	4	5	4	
	F3/F4	0	0	0	7	8	
	F4	0	0	0	0	4	
Validation	F0	3	0	0	0	0	89.8 (82.9, 94.6)
	F0/F1	8	4	0	1	0	
	F1	2	19	1	0	0	
	F1 ± 1	2	2	1	0	0	
	F1/F2	0	6	7	1	0	
	F2	0	1	8	3	0	
	F2 ± 1	0	7	2	3	1	
	F2/F3	0	2	8	9	0	
	F3	0	0	0	2	0	
	F3 ± 1	0	0	2	1	2	
	F3/F4	0	0	0	2	5	
	F4	0	0	0	0	3	

classified fibrosis stage (FibroSteps) for both the training and validation sets. The proposed classification algorithm resulted in 94.9% accuracy (95% CI: 91.3%, 97.4%) in the training set and 89.8% (95% CI: 82.9%, 94.6%) in the validation set. Lastly, when only Step 1 was performed to possibly rule out significant fibrosis (F2/F3/F4), classifying patients as F0/F1 if P(F0/F1)>0.9, had negative predictive values of 96.2% (95% CI: 89.3%, 99.2%) and 94.7% (95% CI: 82.2%, 99.3%) in the training and validation sets respectively.

Discussion

Chronic HCV liver disease poses a considerable public health burden, with increased morbidity and mortality. Therefore, monitoring liver injury severity is necessary to determine the time and necessity of antiviral therapy (1, 43). However, the difficulty and expense of performing repeated liver biopsies, and the cost and know-how required for imaging techniques makes the need for noninvasive methods even more important. (1–4, 17, 18) To our knowledge, a total of fourteen validated fibrosis marker tests have been published between 1991 and 2008, of which five are proprietary (FibroTest, FibroSpect II, ELF, FibroMeter and HepaScore), and nine have not been patented (PGA index, AP index, Bonacini index, Pohl score, Forns index, APRI index, MP3 index, FIB4 and FibroIndex) (8). We developed a freely-available, open-source multistage, stepwise classification (FibroSteps) with the potential to determine liver fibrosis stage in HCV-infected patients with liver disease. A major benefit is that patients with no/mild fibrosis would be spared the trauma and/or cost of additional interventions. This feature is particularly important in resource-limited settings, where imaging elastography and proprietary tests may be cost prohibitive.

The results obtained from this study demonstrated that a combination of twelve biochemical markers (α2M, ApoA1, hyaluronic acid, haptoglobin, MMP-1, MMP-2, TGF-β1, TIMP-1, Urea, AFP, RBCs, and haemoglobin) analysed in a stepwise manner have high accuracy to diagnose the five stages of fibrosis according to the METAVIR score. This set of biomarkers including α2M (31, 35), ApoA1 (28), hyaluronic acid (5), haptoglobin (28), MMP-1 (29), MMP-2 (29), TGF- β1 (26, 27), TIMP-1 (30), urea (31, 34), and AFP (32, 33), either reflect underlying liver function or have a role in promoting or inhibiting hepatic fibrosis, particularly in HCV-infected patients (8, 31).

The proposed stepwise modelling and classification algorithm showed excellent discrimination, calibration, accuracy and good internal crossvalidation. Moreover, the approach is practical because it facilitates efficient biomarker testing. If only clinically significant fibrosis (F2/F3/F4) needs to be ruled out then only eight biomarkers (Step 1, of which urea is a routine inexpensive test) will need to be tested. However, if all five stages need to be determined then 12 unique biomarkers will be tested (three of which are urea, RBC and haemoglobin—inexpensive, routine laboratory tests), but classifying a patient as F0 or F1 requires 11 biomarkers (Steps 1 and 2a), of which urea, RBC and haemoglobin are routine, and classifying a patient as F2, F3, or F4 requires 10 biomarkers (Steps 1, 2b, and 3), of which urea is a routine test. It is worth noting that the performance of the FibroSteps classification is similar to a previously published approach that used six fibrosis tests (some proprietary) involving 11 biomarkers and FibroScan (14), and another published classification combining FibroMeter (six biomarkers) and FibroScan (44).

While these results are encouraging, a limitation is that this study sample had few patients with extremes of staging (F0 and F4). This is not unusual, given that patients without fibrosis (F0, no fibrosis) and cirrhosis (F4, advanced illness) are less likely to be referred for therapy. However, the larger sample size of those with (F1, F2, F3) allowed us to develop a more robust model to distinguish intermediate levels of fibrosis. Additional study is also required to compare FibroSteps with other fibrosis marker tests and imaging elastography and to validate it in a non-Egyptian population, and among a different Egyptian study sample and prospectively over time (45).

In summary, we developed a free, open-source multistage stepwise classification model (FibroSteps) to distinguish between liver fibrosis stages. This test can

potentially substitute for a liver biopsy and imaging techniques and can serve as a very useful non-invasive marker to monitor liver fibrosis progression in HCV-infected patients, and can be used to monitor fibrosis regression after successful treatment without the need for repeated liver biopsies or imaging.

Acknowledgements

The authors would like to acknowledge Dr. Hany Khattab, Pathology Department, Faculty of Medicine, Cairo University, Egypt, for reading the liver biopsies. We also want to acknowledge Dr. Ahmed M. Hashem at the Department of Systems and Biomedical Engineering, Faculty of Engineering, Minia University, Minia, Egypt, and Dr. M. Emad Rasmy and Dr. Khaled M. Wahba at the Department of Systems and Biomedical Engineering, Faculty of Engineering, Cairo University, Giza, Egypt, for their suggestions for the data analysis, and their expertise in data mining. We also want to thank Dr. Ahmed K. Saleh, at Faculty of Science, Cairo University and the University of Maryland School of Medicine, for his preliminary exploratory analysis of the data, and Dr. Entesar H. El Sharqawy, at the Department of Hepatology, Gastroenterology and Infectious Diseases, Benha University, Benha, Egypt, for referring patients to this study, and all the staff members at Cairo University who assisted in this study and for all the subjects who agreed to volunteer for this study.

Competing interests: None.

Funding: This study was supported by funds from the US-Egypt Joint Science and Technology Board [grant number 58-3148-5-045 to SEK and GE]; the Egyptian Science and Technology Development Fund [project number 1512 to OGS]; and MDS's effort was supported by a grant from the NIA, NIH (grant number K25-AG034216-02).

References

1. Gebo KA, Herlong HF, Torbenson MS, *et al.* Role of liver biopsy in management of chronic hepatitis C: a systematic review. *Hepatology* 2002; **36**(5 Suppl 1): S161–72.
2. Cadranel JF, Rufat P, Degos F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). *Hepatology* 2000; **32**: 477–81.
3. Regev A, Berho M, Jeffers LJ, *et al.* Sampling error and intra-observer variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002; **97**: 2614–8.
4. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003; **38**: 1449–57.
5. Esmat G, Metwally M, Zalata KR, *et al.* Evaluation of serum biomarkers of fibrosis and injury in Egyptian patients with chronic hepatitis C. *J Hepatol* 2007; **46**: 620–7.
6. Pinzani M, Vizzutti F, Arena U, Marra F. Technology Insight: noninvasive assessment of liver fibrosis by biochemical scores and elastography. *Nat Clin Pract Gastroenterol Hepatol* 2008; **5**: 95–106.
7. Manning DS, Afdhal NH. Diagnosis and quantification of fibrosis. *Gastroenterology* 2008; **134**: 1670–81.
8. Poynard T, Afdhal NH. Perspectives on fibrosis progression in hepatitis C: an a la carte approach to risk factors and staging of fibrosis. *Antivir Ther* 2010; **15**: 281–91.
9. Kirk GD, Astemborski J, Mehta SH, *et al.* Assessment of liver fibrosis by transient elastography in persons with hepatitis C virus infection or HIV-hepatitis C virus coinfection. *Clin Infect Dis* 2009; **48**: 963–72.
10. Sandrin L, Fourquet B, Hasquenoph JM, *et al.* Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol* 2003; **29**: 1705–13.
11. Halfon P, Bourliere M, Penaranda G, *et al.* Accuracy of hyaluronic acid level for predicting liver fibrosis stages in patients with hepatitis C virus. *Comp Hepatol* 2005; **4**: 6.
12. Montazeri G, Estakhri A, Mohamadnejad M, *et al.* Serum hyaluronate as a non-invasive marker of hepatic fibrosis and inflammation in HBeAg-negative chronic hepatitis B. *BMC Gastroenterol* 2005; **5**: 32.
13. Sakugawa H, Nakayoshi T, Kobashigawa K, *et al.* Clinical usefulness of biochemical markers of liver fibrosis in patients with nonalcoholic fatty liver disease. *World J Gastroenterol* 2005; **11**: 255–9.
14. Boursier J, Bertrais S, Oberti F, *et al.* Comparison of accuracy of fibrosis degree classifications by liver biopsy and noninvasive tests in chronic hepatitis C. *BMC Gastroenterol* 2011; **11**: 132.
15. Leroy V, Halfon P, Bacq Y, *et al.* Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008; **41**: 1368–76.
16. Poynard T, Imbert-Bismut F, Munteanu M, *et al.* Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004; **3**: 8.
17. Ganne-Carrie N, Ziol M, de Ledinghen V, *et al.* Accuracy of liver stiffness measurement for the diagnosis of cirrhosis in patients with chronic liver diseases. *Hepatology* 2006; **44**: 1511–7.
18. Lewin M, Poujol-Robert A, Boelle PY, *et al.* Diffusion-weighted magnetic resonance imaging for the assessment of fibrosis in chronic hepatitis C. *Hepatology* 2007; **46**: 658–65.
19. de Ledinghen V, Vergniol J. Transient elastography (FibroScan). *Gastroenterol Clin Biol* 2008; **32**(6 Suppl 1): 58–67.
20. Boursier J, Konate A, Gorea G, *et al.* Reproducibility of liver stiffness measurement by ultrasonographic elastometry. *Clin Gastroenterol Hepatol* 2008; **6**: 1263–9.
21. Castera L, Foucher J, Bernard PH, *et al.* Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010; **51**: 828–35.
22. Fraquelli M, Rigamonti C, Casazza G, *et al.* Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* 2007; **56**: 968–73.
23. El-Zanaty F, Way A. *Egypt Demographic and Health Survey 2008*. Cairo, Egypt: Ministry of Health, El-Zanaty and Associates and Macro International, 2009.
24. Saleh DA, Shebl F, Abdel-Hamid M, *et al.* Incidence and risk factors for hepatitis C infection in a cohort of women in rural Egypt. *Trans R Soc Trop Med Hyg* 2008; **102**: 921–8.

25. Abdel-Hamid M, Edelman DC, Highsmith WE, Constantine NT. Optimization, assessmentand proposed use of a direct nested reverse transcription-polymerase chain reaction protocol for the detection of hepatitis C virus. *J Hum Virol* 1997; **1**: 58–65.
26. Bissell DM, Roulot D, George J. Transforming growth factor beta and the liver. *Hepatology* 2001; **34**: 859–67.
27. Friedman SL. Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J Biol Chem* 2000; **275**: 2247–50.
28. Imbert-Bismut F, Ratziu V, Pieroni L, *et al.* Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001; **357**: 1069–75.
29. Murawaki Y, Ikuta Y, Kawasaki H. Clinical usefulness of serum tissue inhibitor of metalloproteinases (TIMP)-2 assay in patients with chronic liver disease in comparison with serum TIMP-1. *Clin Chim Acta* 1999; **281**: 109–20.
30. Murphy FR, Issa R, Zhou X, *et al.* Inhibition of apoptosis of activated hepatic stellate cells by tissue inhibitor of metalloproteinase-1 is mediated via effects on matrix metalloproteinase inhibition: implications for reversibility of liver fibrosis. *J Biol Chem* 2002; **277**: 11069–76.
31. Cales P, Oberti F, Michalak S, *et al.* A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005; **42**: 1373–81.
32. Chen CH, Lin ST, Kuo CL, Nien CK. Clinical significance of elevated alpha-fetoprotein (AFP) in chronic hepatitis C without hepatocellular carcinoma. *Hepatogastroenterology* 2008; **55**: 1423–7.
33. Chu CW, Hwang SJ, Luo JC, *et al.* Clinical, virologic and pathologic significance of elevated serum alpha-fetoprotein levels in patients with chronic hepatitis C. *J Clin Gastroenterol* 2001; **32**: 240–4.
34. Shangraw RE, Jahoor F. Effect of liver disease and transplantation on urea synthesis in humans: relationship to acid-base status. *Am J Physiol* 1999; **276**(5 part 1): G1145–52.
35. Tiggelman AM, Boers W, Moorman AF, *et al.* Localization of alpha 2-macroglobulin protein and messenger RNA in rat liver fibrosis: evidence for the synthesis of alpha 2-macroglobulin within Schistosoma mansoni egg granulomas. *Hepatology* 1996; **23**: 1260–7.
36. 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.
37. Breiman L. Random forests. *Mach Learn* 2001; **45**: 5–32.
38. Breiman L. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group, 1984.
39. Friedman J. Multivariate adaptive regression splines. *Ann Stat* 1991; **19**: 1–141.
40. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol* 2005; **12**: 1198–204.
41. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; **138**: 40–4.
42. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; **138**: W1–12.
43. El-Kamary SS, Jhaveri R, Shardell MD. All-cause, liver-related, and nonliver-related mortality among HCV-infected individuals in the general US population. *Clin Infect Dis* 2011; **53**: 150–7.
44. Boursier J, de Ledinghen V, Zarski JP, *et al.* A new combination of blood test and fibroscan for accurate noninvasive diagnosis of liver fibrosis stages in chronic hepatitis C. *Am J Gastroenterol* 2011; **106**: 1255–63.
45. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res* 2008; **14**: 5977–83.