

2008

Prediction of Clinical Outcomes in Primary Biliary Cirrhosis by Serum Enhanced Liver Fibrosis Assay

Marlyn J. Mayo

University of Texas Southwestern Medical Center, Dallas

Julie Parkes

Applied Epidemiology Group and Liver Group, University of Southampton, Southampton, UK

Beverley Adams-Huet

University of Texas Southwestern Medical Center, Dallas

Burton Combes

University of Texas Southwestern Medical Center at Dallas

A. Mills

Virginia Commonwealth University Medical Center, Richmond

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/publichealthresources>

 Part of the [Public Health Commons](#)

Mayo, Marlyn J.; Parkes, Julie; Adams-Huet, Beverley; Combes, Burton; Mills, A.; Markin, Rodney; Rubin, Raphael; Wheeler, Donald; Contos, Melissa; West, A B.; Saldana, Sandra; Getachew, Yonas; Butsch, Robert; Luketic, Velimir; Di Bisceglie, Adrian; Bass, Nathan; Lake, John; Boyer, Thomas; Martinez, Enrique; Boyer, James; Garcia-Tsao, Guadalupe; Barnes, David; and Rosenberg, William, "Prediction of Clinical Outcomes in Primary Biliary Cirrhosis by Serum Enhanced Liver Fibrosis Assay" (2008). *Public Health Resources*. 153.

<http://digitalcommons.unl.edu/publichealthresources/153>

This Article is brought to you for free and open access by the Public Health Resources at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Health Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Marlyn J. Mayo, Julie Parkes, Beverley Adams-Huet, Burton Combes, A. Mills, Rodney Markin, Raphael Rubin, Donald Wheeler, Melissa Contos, A. B. West, Sandra Saldana, Yonas Getachew, Robert Butsch, Velimir Luketic, Adrian Di Bisceglie, Nathan Bass, John Lake, Thomas Boyer, Enrique Martinez, James Boyer, Guadalupe Garcia-Tsao, David Barnes, and William Rosenberg

Prediction of Clinical Outcomes in Primary Biliary Cirrhosis by Serum Enhanced Liver Fibrosis Assay

Marlyn J. Mayo,^{1,22} Julie Parkes,² Beverley Adams-Huet,³ Burton Combes,^{4,22} A. S. Mills,⁵ Rodney S. Markin,⁶ Raphael Rubin,⁷ Donald Wheeler,⁸ Melissa Contos,⁵ A. B. West,⁹ Sandra Saldana,¹⁰ Yonas Getachew,¹¹ Robert Butsch,³ Velimir Luketic,¹² Marion Peters,¹³ Adrian Di Bisceglie,¹⁴ Nathan Bass,¹⁴ John Lake,¹⁵ Thomas Boyer,¹⁶ Enrique Martinez,¹⁷ James Boyer,¹⁸ Guadalupe Garcia-Tsao,¹⁹ David Barnes,²⁰ and William M. Rosenberg²¹

Primary biliary cirrhosis (PBC) is sometimes diagnosed based on a positive antimitochondrial antibody in the appropriate clinical setting without a liver biopsy. Although a liver biopsy can assess the extent of liver fibrosis and provide prognostic information, serum fibrosis markers avoid biopsy complications and sampling error and provide results as a continuous variable, which may be more precise than categorical histological stages. The current study was undertaken to evaluate serum fibrosis markers as predictors of clinical progression in a large cohort of PBC patients. Serial liver biopsy specimens and serum samples were collected every 2 years in 161 PBC subjects for a median of 7.3 years. Clinical progression was defined as development of one or more of the following events: varices, variceal bleed, ascites, encephalopathy, liver transplantation, or liver-related death. Serum hyaluronic acid, tissue inhibitor of metalloproteinase 1, and procollagen III aminopeptide were measured and entered into the previously validated enhanced liver fibrosis (ELF) algorithm. The ability of ELF, histological fibrosis, bilirubin, Model for End-Stage Liver Disease (MELD), and Mayo Risk Score to differentiate between individuals who would experience a clinical event from those who would not was evaluated at different time points. Event-free survival was significantly lower in those with high baseline ELF. Each 1-point increase in ELF was associated with a threefold increase in future complications. The prognostic performance of all tests was similar when performed close to the time of the first event. However, at earlier times in the disease process (4 and 6 years before the first event), the prognostic performance of ELF was significantly better than MELD or Mayo R score. **Conclusion:** The ELF algorithm is a highly accurate noninvasive measure of PBC disease severity that provides useful long-term prognostic information. (HEPATOLOGY 2008;48:1549-1557.)

Abbreviations: AUROC, area under the receiver operating characteristic curve; CI, confidence interval; ELF, enhanced liver fibrosis; MELD, Model for End-Stage Liver Disease; PBC, primary biliary cirrhosis; ROC, receiver operating characteristic.

From the ¹University of Texas Southwestern Medical Center, Dallas, TX; the ²Applied Epidemiology Group and Liver Group, University of Southampton, Southampton, UK; the ³Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX; the ⁴University of Texas Southwestern Medical Center at Dallas, Dallas, TX; the ⁵Department of Pathology, Virginia Commonwealth University Medical Center, Richmond, VA; the ⁶Department of Pathology and Microbiology, University of Nebraska Medical Center, Lincoln, NE; the ⁷Departments of Pathology and Anatomy, Thomas Jefferson University, Philadelphia, PA; the ⁸Departments of Pathology and Cell Biology, University of South Florida, Tampa, FL; ⁹Department of Pathology, Yale University, New Haven, CT; ¹⁰Department of Surgery, Loma Linda University, Loma Linda, CA, and the Department of Internal Medicine, University of Texas Southwestern Medical Center at Dallas; the ¹¹Department of Medicine, University of Texas Southwestern Medical Center; the ¹²Virginia Commonwealth University, Richmond, VA; ¹³University of California, San Francisco, School of Medicine, San Francisco, CA; the ¹⁴Department of Internal Medicine, St. Louis University School of Medicine, St. Louis, MO; the ¹⁵Department of Medicine, University of Minnesota, Minneapolis, MN; ¹⁶Arizona Health Sciences Center, Liver Research Institute, Tucson, AZ; the ¹⁷Atlanta Gastroenterology Associates, The Liver Center, Atlanta, GA; ¹⁸Department of Internal Medicine, Yale University, New Haven, CT; ¹⁹Yale University School of Medicine and Veterans Administration CT Healthcare System, Department of Internal Medicine, Section of Digestive Diseases; the ²⁰Department of Gastroenterology/A31, The Cleveland Clinic Foundation, Cleveland, OH; the ²¹Department of Hepatology, University of Southampton, Southampton, UK; and ²²Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas.

Received February 26, 2008; accepted June 26, 2008.

Supported by the following grants: National Institutes of Health (NIH) R21 DK70995, NIH DK46602, NIH M01-RR00633, DTI UK-Texas Initiative. Address reprint requests to: Marlyn J Mayo, M.D., UT Southwestern Medical Center, 5323 Harry Hines Boulevard, Room J.5.130A, Dallas, TX 75390-9151. E-mail: marlyn.mayo@utsouthwestern.edu; fax: 214-648-0274.

Copyright © 2008 by the American Association for the Study of Liver Diseases.

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

DOI 10.1002/hep.22517

Potential conflict of interest: Nothing to report.

Some patients with primary biliary cirrhosis (PBC) will have a normal life span without complications, whereas others will progress to complications of cirrhosis, liver failure, and death.¹ Accurately predicting clinical outcomes in patients with PBC presents a challenge. Surrogate markers of disease progression are needed for multiple reasons, including providing prognostic information to patients, optimizing referral time for liver transplantation, and serving as endpoints in clinical trials.² The Model for End-Stage Liver Disease (MELD) and Mayo Prognostic Model for PBC (Mayo R score) are prognostic models that have been validated in PBC patients as predictors of death. They are widely used to determine appropriate timing of liver transplantation, but these models were not originally developed to detect disease progression at earlier stages of PBC. Liver biopsy has been the only reliable means to assess prognosis in all stages of disease. However, the accuracy of liver biopsy is limited by sampling error, subjective interpretations, limited frequency of biopsies, and categorical staging systems. Noninvasive surrogate markers are appealing because they do not pose the same risks of pain and bleeding as liver biopsy; they can be performed frequently and provide a score that is a continuous variable, potentially capable of tracking progression from mild fibrosis through to end-stage cirrhosis. Serum fibrosis markers could potentially replace liver biopsy as the test of choice for determining the position of a patient along the spectrum of disease severity. However, their prognostic accuracy compared with liver biopsy has not been assessed in PBC.

The correlation between serum fibrosis markers and histology in other chronic liver disease is good but not perfect. Thus, clinicians have been reluctant to substitute serum fibrosis markers for histological staging. Most marker panels show similar ability to distinguish significant from insignificant fibrosis in cross-sectional studies, with area under the receiver operating characteristic curve (AUROC) of approximately 0.8.³⁻⁵ However, imperfect correlation attributable to biopsy sampling error is to be expected, and only longitudinal studies with clinical endpoints can determine the prognostic accuracy of biochemical markers of disease severity. Very few studies have acquired sufficient follow-up data to evaluate clinical outcomes. Amongst 357 patients with alcoholic liver disease, those with elevated serum YKL-40 (growth factor) or N-terminal propeptide of type III procollagen had shorter survival than patients with normal serum levels of YKL-40 or N-terminal propeptide of type III procollagen.⁶ In a group of 537 treated and untreated patients with hepatitis C, fibrosis biomarkers had a similar but statistically better ability than biopsy to predict hepatitis-

related complications over a 5-year period.⁷ This study investigated the hypothesis that serum fibrosis markers would predict clinical progression as well as or better than current methods of disease staging in primary biliary cirrhosis. This is the first study to evaluate serum fibrosis markers in a large cohort of patients with primary biliary cirrhosis patients and in particular to examine the ability of serum fibrosis markers to predict clinical outcomes.

Patients and Methods

Patients. One hundred sixty-one patients with PBC were prospectively followed between 1993 and 2003 as part of a multicenter U.S. clinical trial (Primary Biliary Cirrhosis Ursodiol Methotrexate Placebo Study [PUMPS]) that was designed to investigate whether low-dose weekly methotrexate, when added to ursodiol, improved survival or delayed progression of PBC. Methotrexate was not found to affect the course of PBC,⁸ so patients from both treatment arms were combined for the purpose of the current analysis. Combining treatment arms was also later justified by a subanalysis demonstrating no effect of the treatment arm on the outcomes of this analysis. Only patients with established but not decompensated PBC were enrolled into the parent trial. Participants were required to have both a positive anti-mitochondrial antibody and either an abnormal alkaline phosphatase or at least stage 1 disease on liver biopsy. Exclusion criteria included a history of variceal bleeding, ascites, or encephalopathy.

Investigations. Subjects received a history, physical examination, and serum blood tests at entry and every 3 months. At entry and at 2-year intervals, they also underwent percutaneous liver needle core biopsy to assess histological changes, endoscopy to look for new varices, and abdominal ultrasound to look for new ascites. On the same day as these procedures, serum for long-term storage was obtained, frozen, and later used for the serum fibrosis marker assay. Thus, all biochemical, serological, and histological data were precisely synchronized. Biopsy specimens were coded with a unique subject identifier and date. They were fixed and stained with hematoxylin-eosin for assessment of cell morphology and trichrome stain for assessment of fibrosis. Each biannual biopsy was read by four independent pathologists and a mean PBC stage of 1 to 4 (explained in Table 1) was calculated. Pathologists convened on an annual basis to independently read those biopsy specimens that had been collected over the preceding 12 months. Biopsy specimens were read in isolation; that is, pathologists did not confer, and previous biopsy specimens from the same subject were not compared with newly acquired biopsy specimens. Each pathologist read

Table 1. Fibrosis Scoring Systems

Ishak Score	Characteristic
0	No fibrosis
1	Fibrous expansion of some portal areas, with or without short septae
2	Fibrous expansion of most portal areas, with or without short septae
3	Fibrous expansion of most portal areas with occasional portal to portal (P-P) and/or portal to central (P-C) bridging
4	Fibrous expansion of portal areas with marked bridging (P-P as well as P-C)
5	Marked bridging (P-P as well as P-C) with occasional nodules (incomplete cirrhosis)
6	Cirrhosis (probable or definite)

PBC Stage	Characteristic
1	No fibrosis. (There had to be some portal inflammation on the initial biopsy to be enrolled)
2	Non-bridging fibrosis. (Portal tracts expanded beyond the limiting plate with any degree of fibrosis less than stage 3.)
3	Bridging fibrosis (Two or more portal-portal bridges)
4	Cirrhosis (at least one regenerative nodule surrounded by fibrosis)

each biopsy only once, except for a single pathologist (S. Mills), who, at the request of the National Institutes of Health, also read each biopsy using the Ishak scoring system to quantify fibrosis (Table 1).⁹ Routine laboratories [bilirubin, albumin, international normalized ratio (INR), creatinine] were performed at each center throughout the trial. Model for end-stage liver disease (MELD) and Mayo R scores were calculated as described previously. Enhanced liver fibrosis (ELF) test scores, were determined on frozen sera after the close of the study. The ELF assay, which includes serum hyaluronic acid, tissue inhibitor of metalloproteinase 1, and N-terminal propeptide of type III procollagen, was previously developed and validated as a predictor of histological fibrosis in a cohort of 1921 patients with chronic liver disease, including 53 patients with either PBC or primary sclerosing cholangitis.³ All assays are heterogeneous immunoassays employing a magnetic particle separation technique developed by Bayer Diagnostics and now property of Siemens Molecular Solutions.³

Analysis. Each of the potential prognostic tests (ELF, histology, serum bilirubin, Mayo R score, and MELD) was then examined for its ability to separate patients who developed clinical progression from those who did not. Clinical progression was defined as the development of one of the following events: new varices, variceal bleed, ascites, encephalopathy, liver-related death, or liver transplantation.

Two types of analyses were performed to examine the predictive ability of the prognostic tests. In the forward-

looking analysis, the ability of each baseline test result (ELF, Ishak score, PBC stage, bilirubin, MELD, Mayo R score) to predict development of a clinical event during follow-up was analyzed. Each of the baseline test scores were divided into three subgroups (low, intermediate, and high score), and the frequency of progression in each subgroup was determined. A Kaplan-Meier time-to-event curve was generated for each subgroup. A Cox proportional hazard ratio was modeled, using ELF and the mean histological PBC stage of the four pathologists as continuous variables.

A rear-looking analysis was also performed to address the possibility that some tests might be better long-range predictors in patients with earlier disease. The rear-looking analysis was designed to account for heterogeneity of disease severity in the study population at the time of entry into the parent trial. This analysis looked retrospectively at tests performed at 0, 2, 4, 6, or 8 years before the development of the first clinical event. For example, in subjects who developed new varices, the data measured at the time of the endoscopy are designated "0 years prior" to the event. In subjects who never developed a clinical event, data were used from the last date the individual was known to be completely free of complications. At each of these time points, the ability of each test to predict whether the subject would later develop a clinical event was determined. Receiver operator curve (ROC) curves were generated to compare the diagnostic performance of each of the tests at each time point in the disease process. The AUROCs for each test were compared using a non-parametric approach,¹⁰ including both omnibus and pairwise comparisons. To adjust for multiple comparisons, a *P*-value of <0.01 was considered significant for the pairwise testing.

ELF scores were correlated with biopsy findings using the Spearman rank order correlation test. An ROC curve

Table 2. Characteristics of Study Population at Study Entry (n = 161)

Characteristic	No Event (n = 119)	Event (n = 42)	P-Value
Age	51.1 ± 8.8	53.0 ± 8.4	0.24
ELF	0.17 ± 0.72	1.10 ± 0.98	<0.0001
	0.14	1.14	
Ishak score (0-6)	2.37 ± 1.34	3.14 ± 1.18	0.0008
	2.0	3.0	
PBC stage (1-4)	1.98 ± 0.70	2.50 ± 0.87	0.0009
	1.0	1.75	
Bilirubin	0.54 ± 0.28	0.81 ± 0.53	0.002
	0.50	0.60	
MELD	6.30 ± 0.66	7.02 ± 1.55	0.001
	6.0	6.0	
Mayo risk	3.68 ± 0.64	4.15 ± 0.60	0.0001
	3.66	4.21	

Results are expressed as mean ± standard deviation and median.

Table 3. Clinical Progression According to Baseline Prognostic Test

Test	Subgroup	Subgroup Value	Clinical Event*	
			Yes	No
ELF†	Low	-1.32 to -0.08	4 (8%)	49 (92%)
	Intermediate	-0.06 to 0.669	13 (24%)	41 (76%)
	High	0.67 to 3.23	25 (46%)	29 (54%)
Ishak score	Low	0-1	1 (4%)	25 (96%)
	Intermediate	2-3	25 (25%)	74 (75%)
	High	4-6	16 (46%)	19 (54%)
PBC stage	Low	<2.0	12 (17%)	57 (83%)
	Intermediate	2.0-2.9	12 (20%)	47 (80%)
	High	3-4	18 (55%)	15 (45%)
Bilirubin	Low	≤1.0	32 (22%)	114 (78%)
	Intermediate	1.1-2.0	8 (62%)	5 (38%)
	High	2.0	2 (100%)	0 (0%)
MELD	Low	≤6.0	24 (20%)	95 (80%)
	Intermediate	7-10	17 (41%)	24 (59%)
	High	≥11	1 (100%)	0
Mayo risk	Low	<3.5	8 (14%)	48 (86%)
	Intermediate	3.5-3.9	9 (21%)	34 (79%)
	High	≥4.0	25 (40%)	37 (60%)

* $P < 0.004$ for all prognostic tests, event group versus no event group, comparison by Cochran-Armitage Trend test.

†ELF is categorized by tertiles.

was generated to examine the ability of ELF score to predict histological fibrosis. Serial assessments of ELF that were made over the 8-year follow-up duration were evaluated for change over time by estimating the slopes of the regression line from initial to final ELF determination with random coefficient models. Comparisons of the slopes between the event and no-event groups were made with linear contrasts from the random coefficient model. The modeling and fitting was performed in SAS 9.1. (SAS Institute, Cary, NC) using the PROC MIXED procedure.

Results

The median follow-up of the 161 subjects was 7.3 years (range, 5 months to 10 years). Most subjects were white (93%) and female (93%). During follow-up, 42 (26%) of the subjects developed one of the events defined as clinical progression: onset of varices ($n = 33$), variceal bleed ($n = 0$), ascites ($n = 6$), encephalopathy ($n = 1$), liver-related death ($n = 1$), or liver transplantation ($n = 1$). Although most of these subjects had the onset of varices as their first event, this was often followed by one or more of the other events.

The characteristics of the study population are shown in Table 2. The mean scores of all prognostic tests were significantly higher at study entry in the group that progressed to develop clinical events during follow-up.

Adequacy of Liver Biopsy Specimens. Of the 578 liver biopsy specimens used in this study, the median

biopsy length was 1.7 cm (minimum, 1 cm; maximum, 5 cm). Most biopsy specimens (80%) had six portal tracts per tissue core. Fewer than 1% had more than six portal tracts per biopsy. Very few (7%) had fewer than four portal tracts, and the remainder (13%) had four or five portal tracts per biopsy. The biopsy specimens were fragmented in 47% of cases. Pathologists were given the option to interpret a biopsy as "inadequate for staging," which occurred in one instance. The weighted kappa score for interpathologist variability in assessing PBC stage was 0.62.

Forward-Looking Event-Free Survival. The chance of developing a clinical event was examined in persons with either low, intermediate, or high scores on each of the prognostic tests obtained at baseline. There was a significant correlation between the baseline score of all prognostic tests and the likelihood of developing a clinical complication over the next 6 years (Table 3). Only 8% of subjects with a low baseline ELF score developed a complication during follow-up, whereas 24% of subjects with an intermediate score and 46% of subjects with high a score developed a clinical complication. In a Cox's proportional hazard model adjusting for age and treatment arm (the latter had no effect), each increase in ELF score by 1 point was associated with a threefold increase in future complications (relative risk, 2.9; 95% CI, 2.0-4.3). Every stage increase of PBC stage (1-4 point scale) led to a twofold increase in future complication rate [relative risk, 2.4; 95% confidence interval (CI), 1.6-3.6].

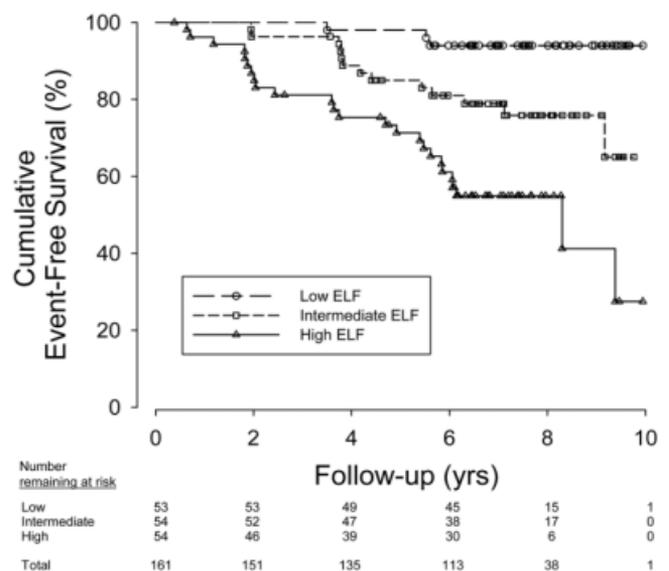


Fig. 1. Event-free survival by baseline ELF score. Kaplan-Meier curve of time to first clinical event by ELF score. Low ELF: -1.32 to -0.08; Intermediate ELF: -0.06 to 0.669; High ELF: 0.67-3.23. Event-free survival was significantly lower in patients with higher baseline ELF scores.

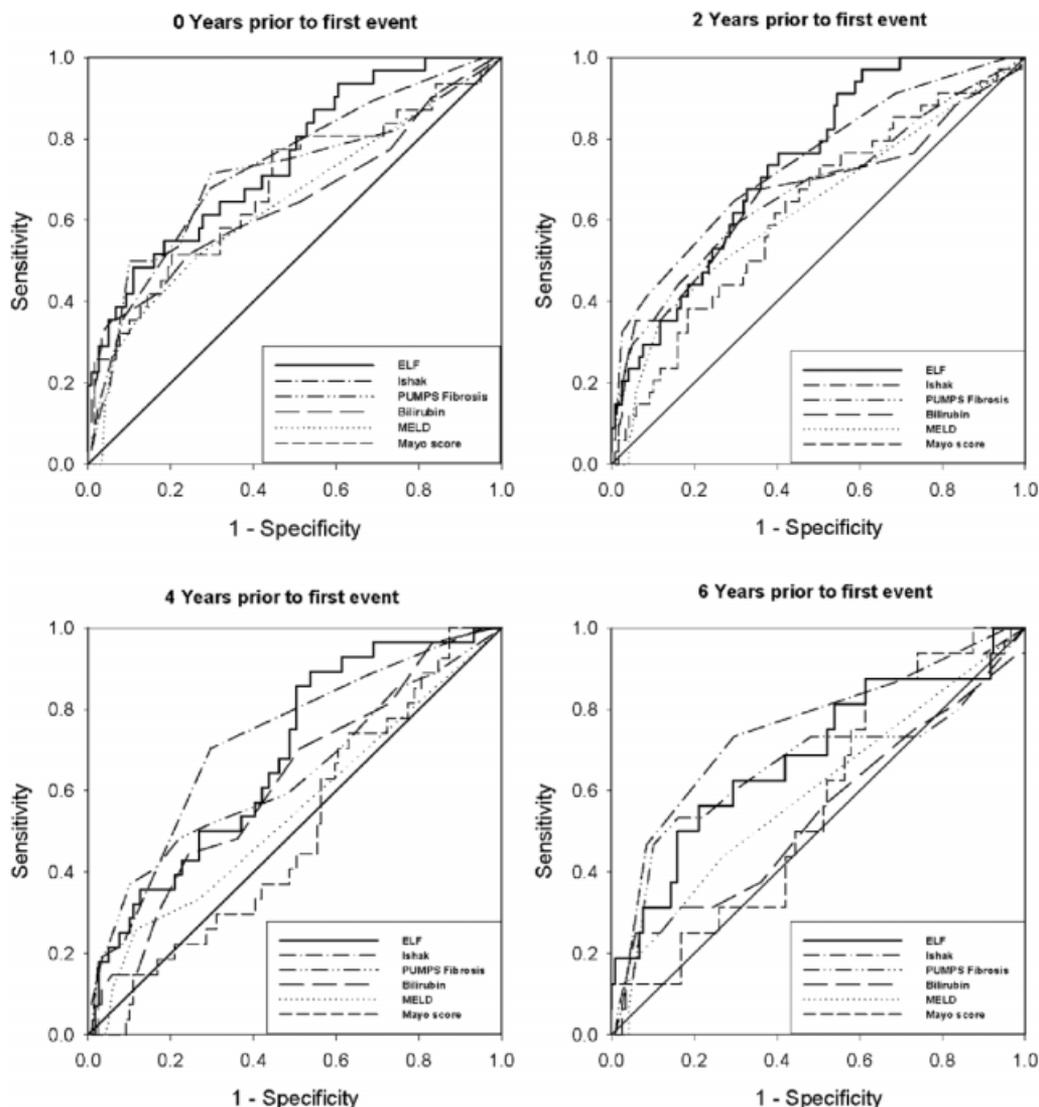


Fig. 2. Performance of prognostic tests at specified time points in the disease process of PBC. The ability of each prognostic test to predict whether an individual would develop a clinical complication was examined at different time points before the first event. Near the time of the first event (year 0), the ROC curves overlapped considerably. However, at earlier times (4 and 6 years before the first clinical event), ELF score and histological fibrosis performed better as prognostic tests. (analysis of variance $P = 0.003$ at 4 years and $P = 0.007$ at 6 years).

ELF performed as well as either Ishak score or PBC stage.

The event-free survival according to baseline ELF score is depicted in Fig. 1. The log rank test comparing the three subgroups showed a highly significant difference ($P < 0.0001$) in event-free survival depending on the baseline ELF score.

Rear-Looking Prediction of Clinical Progression.

The ability of each test to predict whether a subject would later develop a clinical event was also retrospectively determined at increasing time intervals (0, 2, 4, 6, and 8 years) before the first event. For each time point, ROC curves were generated for each diagnostic test (Fig. 2), and the AUROC was calculated (Table 4)

as a measure of the prognostic value of each test. The AUROCs of the various tests were not significantly different from each other at years 0 and 2 (analysis of variance, $P = 0.24$ at 0 years and $P = 0.25$ at 2 years). However, at earlier time points in the disease process (4 and 6 years before the first complication), there were significant differences (analysis of variance $P = 0.003$ at 4 years and $P = 0.007$ at 6 years). The differences in the prognostic tests performed 8 years before the first clinical event were also significant ($P < 0.001$); however, the confidence interval of the AUROCs were quite wide at this time because of the small number of patients in the event group at that time, which made comparisons between tests difficult to interpret.

Table 4. Performance of Prognostic Tests in PBC: Area Under the Receiver Operating Characteristic Curves with 95% Confidence Intervals

Test	Area Under ROC (CI)				
	Years Prior to First Complication				
	0	2	4	6	8
European Liver Fibrosis Panel	0.748 (0.651-0.845)	0.739 (0.653-0.825)	0.682 (0.578-0.786)	0.684 (0.528-0.841)	0.784 (0.600-0.968)
Ishak score (0-6)	0.732 (0.627-0.836)	0.743 (0.647-0.839)	0.721 (0.617-0.824)	0.751 (0.606-0.896)	0.584 (0.423-0.745)
PBC stage (1-4)	0.706 (0.583-0.829)	0.675 (0.562-0.788)	0.634 (0.505-0.764)	0.657 (0.471-0.843)	0.305 (0.077-0.534)
Bilirubin (log _e)	0.650 (0.526-0.774)	0.659 (0.541-0.777)	0.617 (0.501-0.732)	0.539 (0.368-0.711)	0.559 (0.098-1.0)
MELD	0.642 (0.538-0.746)	0.630 (0.531-0.729)	0.544 (0.439-0.649)	0.592 (0.455-0.730)	0.612 (0.327-0.898)
Mayo PBC risk score	0.680 (0.564-0.796)	0.625 (0.518-0.733)	0.501 (0.388-0.614)	0.431 (0.294-0.568)	0.441 (0.119-0.763)

At each time point in the disease process, Ishak staging system and ELF score had highest predictive values (AUROC). In fact, there was no statistically significant difference in the AUROC of ELF and histological staging at any time point. The AUROC of histological staging by the seven-point Ishak system was consistently slightly greater than the AUROC of the four-point PBC stage. The prognostic performance of ELF (AUROC) was also higher than the other noninvasive tests (bilirubin, MELD, Mayo risk score) at each time point, although this difference was not statistically significant at later points in the disease process (0 and 2 years before the first event). The most strongly significant differences were between ELF and Mayo R score AUROCs at earlier time points in the disease process (4 and 6 years before the first event; $P = 0.004$ and $P = 0.009$, respectively).

Correlation of ELF with Histology. The initial ELF score in each subject correlated well with the extent of disease on liver biopsy, using either the Ishak scoring system ($\rho = 0.52$, $P < 0.0001$) or the PBC staging system ($\rho = 0.60$, $P < 0.0001$). However, as expected, there was much overlap in individual ELF scores, particularly in persons with intermediate stages of histological fibrosis (Fig. 3). ELF score was an accurate noninvasive marker of histological fibrosis. The AUROC of ELF to predict cirrhosis (Ishak 5-6) score was 0.76 (95% CI 0.63-0.89) and to predict significant fibrosis (Ishak 3-6) was 0.75 (CI 0.67-0.82) (Fig. 4)

Serial Determinations of ELF. One hundred forty-nine subjects had more than one frozen serum sample and thus could be used to evaluate change in serial ELF score determinations over the 8-year follow-up period. The ELF score increased by 0.032 per year (95% CI: 0.008-0.005, $P = 0.009$), and the rate of increase was no different comparing the event (slope 0.046 per year, CI:

-0.002 to 0.094, $n = 38$) and no event (slope 0.027 per year, 95%CI: 0 to 0.054, $n = 111$) groups ($P = 0.50$).

Discussion

This is the first study to examine the utility of a serum fibrosis marker panel in primary biliary cirrhosis, and in particular, to predict clinical outcomes. The relationship between the baseline ELF score and the likelihood of developing a clinical complication over the next 6 years was highly significant, as demonstrated by the event-free survival Kaplan-Meier analysis.

Other noninvasive prognostic models, such as the Mayo R score and MELD, are based on bilirubin, albumin, prothrombin time, edema, and creatinine. These biochemical and clinical markers become abnormal in the later stages of cirrhosis and hepatic dysfunction, which is why they are excellent predictors of death. However, these markers are not sensitive to disease progression earlier in the course of PBC. This is most likely the explanation for the poor prognostic performance of Mayo R score, MELD, and bilirubin 6 years before the first clinical event, similar to the chance of flipping a coin (AUROC = 0.420-0.564). Even though the parent trial was not originally designed to compare prognostic tests, the availability of concurrent data and long-term follow-up enabled us to make comparisons and determine retrospectively the most accurate prognostic tests.

Clinical jaundice is known to be a late but poor prognostic sign in PBC. The PUMPS trial excluded patients with a history of bilirubin greater than 2.0, so only two patients in this cohort had jaundice at study entry. Although both of these patients did develop a clinical event, a large percentage of patients with normal baseline bilirubin also progressed to have a clinical event. In the retro-

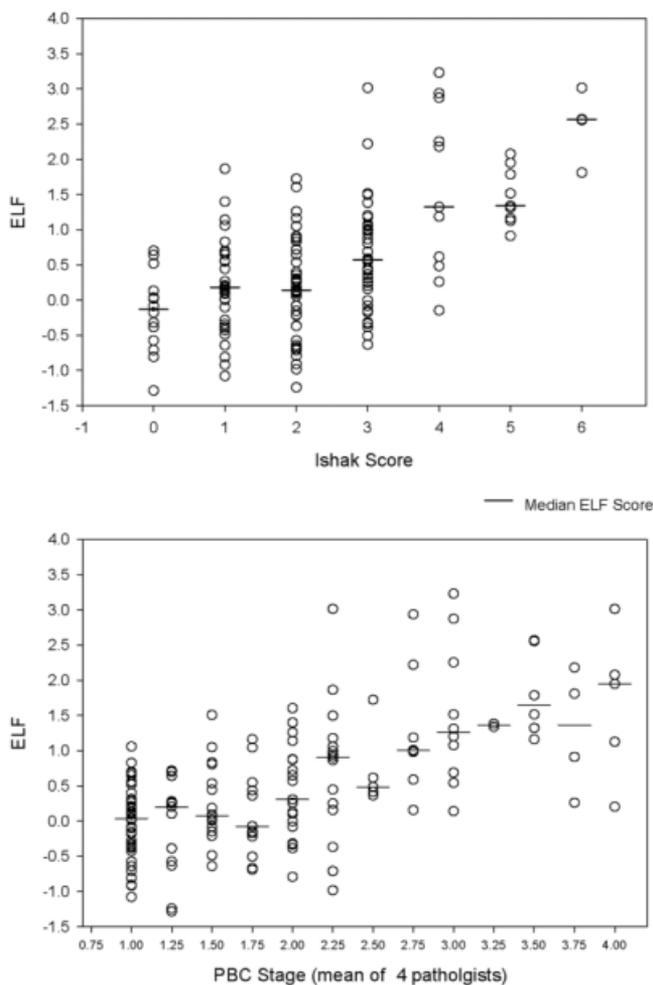


Fig. 3. Distribution of first ELF score according to biopsy score. The first serum ELF score correlated well with histology. Because the mean PBC stage was derived by averaging the scores of four pathologists, the calculated values ranged from 1.0 to 4.0, in increments of 0.25.

spective analysis, taking the \log_e of bilirubin improved its prognostic ability, although the AUROC of \log_e bilirubin (reported in this paper) was still lower than that of ELF or histology.

The key components of the ELF algorithm are hyaluronic acid, tissue inhibitor of metalloproteinase 1, and procollagen type III N-terminal propeptide, which are expressed during early stages of collagen deposition in the liver, and this is most likely the reason that ELF retains its prognostic ability even in early stages of the disease process (AUROC, 0.737-0.863 at all time points). None of the prognostic tests, however, including biopsy and ELF, had AUROCs of over 0.9. Thus, no test is error-free when trying to predict clinical progression, and this must be considered when counseling individual patients in the clinical setting.

Liver biopsy has been the reference standard for staging PBC and other chronic liver diseases for over 50 years, and

the prognostic information gained from liver biopsy is useful. The three most popular staging systems for PBC are the Ludwig, Popper, and Scheuer systems.¹¹⁻¹³ This study used the the Ludwig system, which has become favored over the others because it is increasingly recognized that florid duct lesions or ductular proliferation need not be present to define early stages. ELF algorithm was found to be superior to PBC stage at all time points for predicting clinical progression. Of course, a liver biopsy provides additional diagnostic information, but this is not always necessary. A typical middle-aged woman with cholestatic liver tests, positive AMA, and no other liver disease risk factors does not require a confirmatory liver biopsy.¹⁴ Although not originally developed for PBC, the Ishak system was also used to score each of the biopsy specimens in this study. The performance of the Ishak score was not statistically different from the performance of the ELF score. Thus, we propose that if a liver biopsy is not needed for diagnostic purposes, the ELF score can be used for prognostic staging and monitoring during follow-up. In the situation in which a liver biopsy is appropriate for diagnostic purposes, the Ishak system should be used for more accurate staging. The Ishak system is based entirely on the extent of fibrosis and provides better prognostic information but cannot establish the diagnosis of PBC. The diagnosis must still be made by the presence of typical histological features such as nonsuppurative destructive cholangitis.

The possibility of replacing staging liver biopsies with simple noninvasive marker algorithms is gaining momentum as evidence is accumulating in other chronic liver diseases that these panels are reasonable surrogate markers of histological liver fibrosis. Several cross-sectional studies

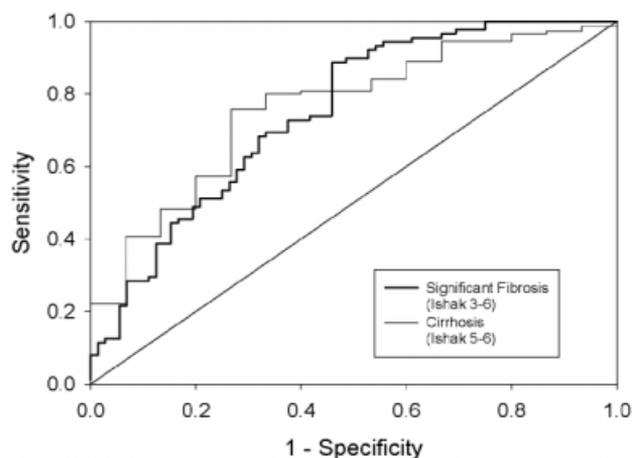


Fig. 4. ELF as a predictor of histological fibrosis and cirrhosis. ELF score was an accurate noninvasive marker of histological fibrosis. The AUROC of ELF to predict cirrhosis was 0.76 (95% CI 0.63-0.89) and to predict significant fibrosis was 0.75 (CI 0.67-0.82)

have evaluated the ability of marker panels to quantify fibrosis in hepatitis C and nonalcoholic fatty liver disease.¹⁵⁻¹⁹ Consistently, these marker panels are reasonably accurate at distinguishing the absence of fibrosis from advanced cirrhosis, but the accuracy of the test in a single individual, particularly an individual with an intermediate degree of fibrosis, is lower.^{15,20-23} These studies have been criticized because the "gold standard" of histology is inaccurate because of the high rate of sampling error known to occur in patients with diffuse parenchymal liver diseases.²⁴ One study of patients with cirrhosis from different causes demonstrated that three samples taken by the same route in the same patients were concordant for the degree of fibrosis in only 50% of cases.²⁵ Although increasing the length of the biopsy improves diagnostic accuracy, the rate of discordance in biopsy specimens larger than 25 mm is still 25% in hepatitis C.²⁶ Thus, although "bigger is better," needle biopsies are by their nature somewhat inaccurate. PBC, in particular, is characterized by patchy involvement of portal areas. Whole section scanning of 50 PBC liver explants demonstrated that only 20% of PBC livers have a consistent histological stage of fibrosis throughout the liver, even when the tissue was obtained at clinically defined end-stage disease.²⁷ The correlation between ELF score and histological fibrosis in these patients was strong. However, as expected, there was also significant overlap of individual ELF scores, particularly in persons with intermediate stages of histological fibrosis. The ELF score, therefore, was not a perfect reflection of the histological fibrosis score; however, its excellent performance as a prognostic marker implies that biopsy error may actually account for part of the discord between ELF and histological fibrosis score.

One advantage of the current study was that serial serum samples were available in a very well-defined cohort with meticulously determined clinical outcomes over a long period. Thus, prognostic tests could be evaluated at different time points and also evaluated for change over time. The average ELF score increased over time (by 0.032 per year). This average rate of increase was not different between subjects who experienced an event and those who did not, which indicates that the rate of PBC fibrosis progression may be relatively steady, regardless of whether the starting point is early or advanced disease. It also suggests that the patients who developed the clinical events started the study at a later time point in their disease process. These data argue against the theory that there may be a rapidly fibrotic and slowly fibrotic form of PBC, at least in the cohort studied here.

In conclusion, the ELF algorithm is a highly accurate noninvasive measure of PBC disease severity that provides

accurate prognostic information. The ability of this algorithm to predict clinical outcomes is as good as or better than traditional prognostic tests for PBC, particularly in early stages of the disease. It is a simple, noninvasive technique with long-range predictive ability that should be very useful in the clinical assessment of PBC.

Acknowledgment: Reagents were provided by Bayer Diagnostics/Siemens Molecular Solutions and i-Quir Limited.

References

1. Corpechot C, Carrat F, Bahr A, Chretien Y, Poupon RE, Poupon R. The effect of ursodeoxycholic acid therapy on the natural course of primary biliary cirrhosis. *Gastroenterology* 2005;128:297-303.
2. McHutchison J, Poynard T, Afdhal N, International Fibrosis Group Meeting Participants. Fibrosis as an end point for clinical trials in liver disease: a report of the international fibrosis group. *Clin Gastroenterol Hepatol* 2006;4:1214-1220.
3. 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.
4. 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 Gastroenterol* 2006;6:6.
5. Naveau S, Raynard B, Ratziu V, Abella A, Imbert-Bismut F, Messous D, et al. Biomarkers for the prediction of liver fibrosis in patients with chronic alcoholic liver disease. *Clin Gastroenterol Hepatol* 2005;3:167-174.
6. Nojgaard C, Johansen JS, Christensen E, Skovgaard LT, Price PA, Becker U, et al. Serum levels of YKL-40 and PIIINP as prognostic markers in patients with alcoholic liver disease. *J Hepatol* 2003;39:179-186.
7. Ngo Y, Munteanu M, Messous D, Charlotte F, Imbert-Bismut F, Thabut D, et al. A prospective analysis of the prognostic value of biomarkers (FibroTest) in patients with chronic hepatitis C. *Clin Chem* 2006;52:1887-1896.
8. Combes B, Emerson SS, Flye NL, Munoz SJ, Luketic VA, Mayo MJ, et al. Methotrexate (MTZ) plus ursodeoxycholic acid (UDCA) in the treatment of primary biliary cirrhosis. *HEPATOLOGY* 2005;42:1184-1193.
9. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995;22:696-699.
10. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics* 1988;44:837-845.
11. Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch A Pathol Anat Histol* 1978;379:103.
12. Popper H, Schaffner F. Nonsuppurative destructive chronic cholangitis and chronic hepatitis. In: Popper H, Schaffner F, eds. *Progress in Liver Diseases*, vol 3. New York: Grune & Stratton, Inc.; 1970:336-354.
13. Scheuer PJ. Primary biliary cirrhosis. *Proc R Soc Med* 1967;60:1257.
14. Zein CO, Angulo P, Lindor KD. When is liver biopsy needed in the diagnosis of primary biliary cirrhosis? *Clin Gastroenterol Hepatol* 2003;1:89-95.
15. Imbert-Bismut F, Ratzui V, Pironi L, Charlotte F, Benhamou Y, Poynard T, for the MULTIVIRC group. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001;357:1069-1075.
16. Guha IN, Parkes J, Roderick PR, Harris S, Rosenberg WM. Non-invasive markers associated with liver fibrosis in non-alcoholic fatty liver disease. *Gut* 2006;55:1650-1660.

17. Zaman A, Rosen HR, Ingram K, Corless CL, Oh E, Smith K. Assessment of FIBROSpectII to detect hepatic fibrosis in chronic hepatitis C patients. *Am J Med* 2007;120:2809-2814.
18. Guha IN, Parkes J, Roderick P, Chattopadhyay D, Cross R, Harris S, et al. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: validating the European Liver Fibrosis Panel and exploring simple markers *HEPATOLOGY* 2008;47:455-460.
19. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *HEPATOLOGY* 2007;45:846-854.
20. Poynard R, Imbert-Bismut I, Ratzui V, Chevret S, Jardel C, Moussalli J, et al. Biochemical markers of liver fibrosis in patients infected by hepatitis C virus: longitudinal validation in a randomized trial. *J Virol Hepatitis* 2002;9:128-133.
21. Poynard T, McHutchison J, Manns M, Myers RP, Albrecht J. Biochemical surrogate markers of liver fibrosis and activity in a randomized trial of peginterferon alfa-2b and ribavirin. *HEPATOLOGY* 2003;38:481-492.
22. Myers RP, Ratzui V, Imbert-Bismut F, Charlotte F, Poynard T, for the MULTIVIRC group. Biochemical markers of liver fibrosis: a comparison with historical features in patients with chronic hepatitis c. *Am J Gastroenterol* 2002;97:2419-2425.
23. Myers RP, de Torres M, Imbert-Bismut F, Ratzui V, Charlotte F, Poynard T, et al. Biochemical markers of fibrosis in patient with chronic hepatitis C: a comparison with prothrombin time, platelet count, and age-platelet index. *Dig Dis Sci* 2003;48:146-153.
24. Scheuer PJ. Liver biopsy size matters in chronic hepatitis: bigger is better. *HEPATOLOGY* 2003;38:1356-1357.
25. Maharaj B, Maharaj RJ, Leary WP, Coppan RM, Naran AD, Pirie D, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. *Lancet* 1986;8480:523-525.
26. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *HEPATOLOGY* 2003;38:1449-1457.
27. Garrido MC, Hubscher SG. Accuracy of staging in primary biliary cirrhosis. *J Clin Pathol* 1996;49:556-559.