Noninvasive Markers of Hepatic Fibrosis in Chronic Hepatitis B

Thierry Poynard · Yen Ngo · Mona Munteanu ·
Dominique Thabut · Vlad Ratziu

Abstract A serum biomarker (FibroTest; Biopredictive, Paris, France; FibroSure; LabCorp, Burlington, USA) and liver stiffness measurement (LSM) by Fibroscan (Echosens, Paris, France) have been extensively validated in chronic hepatitis C. This review updates the clinical validation of serum biomarkers and LSM in patients with chronic hepatitis B (CHB). One meta-analysis combined all published studies and another used a database combining FibroTest individual data. Sensitivity analysis assessed the impact of several factors, including authors' independence, length of biopsy, ethnicity, hepatitis B early antigen status, viral load, and alanine aminotransferase value. Only two biomarkers had several validations: FibroTest (8 studies, 1,842 patients), and Fibroscan (5 studies, 618 patients). For the diagnosis of advanced fibrosis, the standardized area under the receiver operating curve was 0.84 (0.79–0.86) for FibroTest and 0.89 (0.83–0.96) for LSM, without significant difference. No significant factors of variability were identified for FibroTest's performance. In conclusion, FibroTest and LSM were the most validated biomarkers of fibrosis in CHB. However, the reliability of Fibroscan must be better assessed.

Keywords Fibrosis biomarkers · Meta-analysis · Obuchowski measure · FibroTest · FibroSure · FibroScan · Biopsy · Elastography

T. Poynard (✉) · D. Thabut · V. Ratziu
Assistance Publique Hôpitaux de Paris-University Pierre et Marie
Curie, Liver Center, Groupe Hospitalier Pitié Salpêtrière,
75013 Paris, France
e-mail: tpoynard@teaser.fr

D. Thabut
e-mail: dthabut@gmail.com

V. Ratziu
e-mail: vratziu@teaser.fr

Y. Ngo · M. Munteanu
Biopredictive,
40 Rue du Bac,
75007 Paris, France

Y. Ngo
e-mail: yen@biopredictive.com

M. Munteanu
e-mail: mona@biopredictive.com

Introduction

The "gold standard" reference for determining hepatic histology in chronic hepatitis B (CHB) used to be liver biopsy. Because of the limitations of liver biopsy, however, several noninvasive biomarkers have been validated since 2001.

Between 2001 and 2008, two techniques—a serum biomarker (FibroTest; Biopredictive, Paris, France; Fibro-Sure; LabCorp, Burlington, USA) and liver stiffness measurement (LSM) (Fibroscan; Echosens, Paris, France)—were extensively validated using biopsy, and six other biomarkers also were investigated but with fewer studies: Enhanced Liver Fibrosis (ELF) test, Fibrometer, Fibrospect, FIB-4 (age, platelets, aspartate aminotransferase [AST] and alanine aminotransferase [ALT]), Hepascore, and APRI (AST-to-platelet ratio index) [1]. According to meta-analyses, most of these studies were initially published in patients with chronic hepatitis C (CHC). In eight previous meta-analyses published between 2002 and 2008 [2–9], only FibroTest had been validated in five studies including only patients with CHB [10–13, 14•]. No meta-analysis was possible for other biomarkers.

Unlike CHC [7], no guidelines recommend biomarkers for routine clinical use in patients with CHB, although "they might be helpful on a case-by-case basis" [15••].

The primary objective of this review is to update the last published meta-analyses on the clinical validation of serum biomarkers and LSM in the management of patients with

CHB. We used an integrated database of individual data and the new recommended Obuchowski measures to prevent spectrum effect and multiple testing [16••], and we searched for prognostic studies.

Methods

This study was conducted according to the principles expressed in the Declaration of Helsinki. Two meta-analyses were performed: one combined all identified published studies, and the other used an integrated database combining individual data provided by authors.

Design of Meta-Analysis of Published Studies

The primary outcome was the differentiation of nonadvanced fibrosis from advanced fibrosis, defined using the METAVIR scoring system (F0–F1 vs F2–F4) [17]. We chose this endpoint because it seems at least as rational as the viral load to assess the severity of CHB [15••, 18••]. The other endpoints were the performance for the diagnosis of cirrhosis and the prognostic values.

To select published studies, we used the Standards for Reporting of Diagnostic Accuracy (STARD) criteria and the Cochrane Database of Systematic Reviews (CDSR) methods [19]. Key STARD criteria include factors such as whether (1) the study population was relevant to the clinical question being addressed; (2) there was a careful description of the population from which the patients were drawn, as well as actual inclusions and exclusions; (3) recruitment and the mode of sampling were carefully described; (4) researchers interpreting the noninvasive test were blinded to the reference test result; and (5) sufficient data were provided to complete a 2×2 table of true and false positive and negative diagnoses. Studies published only with an abstract provided insufficient data and were excluded.

Search Strategy

We searched MEDLINE with the keywords "HBV, Fibrosis, FibroTest, Fibroscan, Elastography, ELF, Fibro-spect, and Hepascore." We hand-searched key journals (*Gastroenterology, Hepatology, Journal of Hepatology, Gut, Journal of Viral Hepatitis*, and *American Journal of Gastroenterology*) from February 2001 to December 2010 to validate the search, as well as the references of publications identified by previous searches.

Inclusion and Exclusion Criteria

We excluded all studies except those that included patients with CHB; stated that all patients had had the biomarker and liver biopsy; provided data for true positives and negatives, false positives and negatives, and area under the receiver operating curves (AUROCs) for advanced fibrosis; stated that the biomarker had been assessed blind to the biopsy; and stated the method used for defining the degree of fibrosis. We were careful to avoid including data from duplicate publications. We excluded the following biomarkers, which combined ALT or AST in their formula with a risk of confounding impact of necro-inflammatory histologic activity on the fibrosis stage estimates: Fibrometer, FIB-4, and APRI [20–22].

Statistical Methods Used in AUROCs Meta-Analysis

Meta-analysis was performed twice, once according to the absolute value of the not-standardized AUROCs (AUROC) and once according to the AUROCs standardized for the spectrum of fibrosis stages (StAUROC). We previously demonstrated that the AUROCs were highly related to the Difference between Advanced and Non Advanced mean fibrosis stages (DANA); the StAUROC is the AUROC adjusted for the difference of the observed DANA versus a standard DANA of 2.5 fibrosis METAVIR units (DANA = 2.5, if there was a uniform prevalence of 0.20 in each of the five stages). The regression between the AUROCs and the whole DANA spectrum of published studies in CHB enables the AUROC to be estimated from the DANA. The regression formula for standardizing AUROCs estimated from different stage prevalences, according to previously published method and to the identified biomarkers, was: $StAUROC = AUROC + 0.0482 \times (2.5\,DANA)$ [20, 23••].

The AUROC was estimated by the empirical (nonparametric) method of Delong et al. [24], equivalent to the Mann–Whitney statistic, and compared using the paired method of Zhou et al. [25]. The meta-analysis was stratified according to the biomarkers. The analysis used a random effect model, and the heterogeneity between effects according to publications and biomarkers was tested using Cochran's Q heterogeneity test. A sensitivity analysis was also performed according to the independence of authors for FibroTest.

Design of Individual Data Meta-Analysis

Patients

Patients with individual data included in the database were described in a previous overview of FibroTest in patients with chronic liver diseases [8]. For the present study, only patients with hepatitis B virus (HBV) were included. Patients' data were included with the agreement of authors in the present study if they met the following optimal conditions of analyses: they must have been included in a validation study including FibroTest, ALT, and biopsy with METAVIR scoring system for fibrosis score, according to

preanalytical and analytical recommendations. All patients have CHB with positive polymerase chain reaction test. All biomarkers have been assessed blinded to clinical, biochemical, or histologic patient characteristics.

Biochemical Analysis

The dosage methods for each of the FibroTest parameters followed the standardized conditions, as listed in the technical recommendations (http://www.biopredictive.com).

Liver Biopsies

Liver biopsies were processed using standard techniques. A pathologist who was unaware of the biomarkers evaluated the fibrosis stage according to the METAVIR scoring system [17]. Fibrosis was staged on a scale of 0 to 4: F0—no fibrosis; F1—portal fibrosis without septa; F2—few septa; F3—numerous septa without cirrhosis; and F4—cirrhosis. Biopsies were performed with a 16-gauge Hepafix Luer-Lock needle (Braun Melsungen, Melsungen, Germany) in the Paris center, and with various needles in the multicenter study from Marseille.

Statistical Analysis of Integrated Database

The use of the AUROC raises two methodologic issues. First, its use is based on the assumption that the gold standard is binary, whereas fibrosis staging uses an ordinal scale. This difference implies that fibrosis stages in the study sample have to be aggregated into two groups, a process that can lead to discordant conclusions, depending on how the groups are aggregated [16••, 20, 22]. Analysis based on the AUROC can also be biased by the way in which the proportion of each stage of fibrosis in the sample fits the distribution in the reference population to which the indices are applied. As a result, the comparison of different AUROCs based on samples with different stage distributions may be flawed (spectrum effect) [16••, 20, 22, 23••]. The spectrum effect reflects the inherent variation in test performance among population subgroups. Subgroup variation is not a bias, but is clinically relevant information to be identified and reported with appropriate analyses. If a spectrum effect is possible, heterogeneity should be assessed by subgroup analyses of test performance.

To overcome these methodologic issues, two methods were recently proposed and validated for fibrosis staging and activity grading in patients with CHC. We proposed a standardization of the AUROC for the distribution of fibrosis stages to deal with the spectrum effect [20, 23••]. Lambert et al. [16••] proposed using the Obuchowski measure to overcome both spectrum effect and ordinal scale. This measure can be used in situations in which the gold standard is not binary. Furthermore, the Obuchowski measure allows comparison of two biomarkers with a single test, avoiding appropriate correction for the type 1 error when comparing two biomarkers for different stages or grades [16••, 22].

Obuchowski Measure

The Obuchowski measure is a multinomial version of the AUROC. With N (=5) categories of the gold standard outcome (histologic activity grade) and AUROCst (the estimate of the AUROC of diagnostic tests for differentiating between categories s and t), the Obuchowski measure is a weighted average of the $N(N-1)/2(=10)$ different AUROCst corresponding to all the pairwise comparisons between two of the N categories.

Each pairwise comparison was weighted to take into account the distance between activity grades (ie, the number of units on the ordinal scale). A penalty function proportional to the difference in METAVIR units between grades was defined: the penalty function was 0.25 when the difference between stages was 1; 0.50 when the difference between stages was 2; 0.75 when the difference was 3; and 1 when the difference was 4. The Obuchowski measure can be interpreted as the probability that the noninvasive index will correctly rank two randomly chosen patient samples from different fibrosis stages according to the weighting scheme, with a penalty for misclassifying patients. Note that the overall Obuchowski measure is not equivalent to a usual AUROC curve, because the measurements are weighted according to the distance between stages. The FibroTest cutoffs were those recommended by the manufacturer since the first validation using biopsy: 0.27 for F1, 0.48 for F2, and 0.58 for F3 and 0.74 for F4 [20]. For ALT, the a priori simple cutoffs chosen were 25, 50, 75, and 100 IU/L, because we previously demonstrated that the expression of ALT activity using the upper limit of normal was hazardous [26].

Main Endpoint

The main endpoint was the accuracy estimated with Obuchowski measure. The FibroTest accuracy was compared with ALT, a standard marker of liver disease severity in CHB [15••].

Sensitivity Analyses

Sensitivity analyses were performed in the integrated database by comparing FibroTest performances according to the variability factors: gender, biopsy length [27], ethnicity, hepatitis B early antigen (HBeAg) status, HBV genotype, viral load, and ALT value. In one study, patients were included twice, because they had FibroTest and biopsy once before and once after the treatment; a sensitivity analysis was performed comparing patients

before and after treatment. When a difference was suspected (P<0.10) between nonstandardized AUROCs for advanced fibrosis, Obuchowski measures were assessed.

Each estimate was given with its 95% confidence interval (95% CI). Analyses were performed on NCSS software (Kaysville, Utah, USA) [28] and on R software [29].

Results

Databases

The search retrieved 42 references: 22 for FibroTest, 17 for LSM, two for Hepascore, and one for ELF score. From the 41, 17 were pre-included as original diagnostic studies including only CHB, and 15 studies fully satisfied the inclusion criteria (Table 1). Only one prognostic study was identified [18••].

For FibroTest, eight studies were pre-included, and all were included for not-standardized meta-analysis for advanced fibrosis. Since the previous meta-analysis, three new studies were included [30–32]. One study was not included in the standardized meta-analysis because it did not provide prevalence of each fibrosis stage [30]. One study did not provide data for cirrhosis [11].

For LSM measured by FibroScan, six studies were pre-included and five were included for meta-analysis [33–37]. One was excluded because it did not provide AUROC [37], and one was not included in the standardized meta-analysis because it did not provide prevalence of each fibrosis stage [35].

For Hepascore, two studies were pre-included and both included in the meta-analysis (Table 1) [31, 38].

For the integrated database, we excluded six patients with acute hepatitis suspected by security algorithms [8], explaining differences in the total number of subjects between integrated database (n=1,303) and the published studies (n=1,309) [10, 11].

Comparison Between Biomarker Performances According to Published Studies

Diagnosis of Advanced Fibrosis

Not-standardized AUROCs were all significantly higher than the random 0.50 value (P<0.001). Mean AUROC was 0.80 (95% CI, 0.77–0.82). There was no significant difference among performance of the three biomarkers: FibroTest AUROC = 0.79 (0.76–0.82), HepaScore AUROC = 0.75 (0.66–0.84), and LSM AUROC = 0.84 (0.78–0.89). There was no significant heterogeneity between studies (Cochran heterogeneity test = 9.5, P=0.80).

Standardized AUROCs, taking into account the spectrum of fibrosis stages, were all significantly higher than the random 0.50 value (P<0.001) (Fig. 1). Mean StAUROC

was 0.84 (95% CI, 0.80–0.87). There was no significant difference among the performance of the three biomarkers: FibroTest StAUROC = 0.84 (0.79–0.86), HepaScore StAUROC = 0.80 (0.71–0.89), and LSM StAUROC = 0.89 (0.83–0.96). There was no significant heterogeneity between studies (Cochran heterogeneity test = 6.3, P=0.90).

Diagnosis of Cirrhosis

Not-standardized AUROCs were all significantly higher than the random 0.50 value (P<0.001). Mean AUROC was 0.87 (95% CI, 0.82–0.91). There was no significant difference among the performance of the three biomarkers: FibroTest AUROC = 0.83 (0.78–0.87), HepaScore AUROC = 0.86 (0.74–0.98), and LSM AUROC = 0.93 (0.87–0.99). There was no significant heterogeneity between studies (Cochran heterogeneity test = 15.1, P=0.13).

Standardized AUROCs, taking into account the spectrum of fibrosis stages, were all significantly higher than the random 0.50 value (P<0.001). Mean StAUROC was 0.89 (95% CI, 0.85–0.92). There was no significant difference among the performance of the three biomarkers: FibroTest StAUROC = 0.85 (0.80–0.90), HepaScore StAUROC = 0.88 (0.76–1.00), and LSM StAUROC = 0.96 (0.90–1.00). There was no significant heterogeneity between studies (Cochran heterogeneity test = 16.6, P=0.09).

Sensitivity Analysis

There was no difference between the performances of FibroTest according to the independence of authors. The not-standardized AUROC for advanced fibrosis was 0.83 (0.78–0.87) for the four independent publications and 0.77 (0.73–0.81) for the four dependent publications. The other performances were also not significant (data not shown).

FibroTest's Performance According to Individual Data Meta-Analysis (Table 2)

Performance Between All Fibrosis Stages

The overall mean (95% CI) accuracy of FibroTest (Obuchowski measure) was 0.844 (0.832–0.856), greater than that of ALT, which was 0.803 (0.791–0.815), P<0.0001. FibroTest's performances were particularly greater than those of ALT for the diagnosis between F0 and F1 (0.61 [0.56–0.66] vs 0.52 [0.46–0.58]) and between F3 and F4 (0.58 [0.52–0.64] vs 0.50 [0.44–0.56]).

Sensitivity Analyses

Using not-standardized AUROC, FibroTest had a lower performance in patients without HBeAg (0.73 [0.68–0.77])

versus patients with HBeAg (0.78 [0.75–0.82]; *P*=0.046), and in Asian patients (0.72 [0.67–0.76]) versus non-Asian patients (0.78 [0.74–0.82]). These differences were no more significant using Obuchowski measures: 0.831 (0.814–0.848) versus 0.854 (0.838–0.870; *P*>0.10) for HBeAg status, and 0.836 (0.814–0.848) versus 0.853 (0.839–0.867; *P*>0.10) for ethnicity (Table 2). All the other characteristics were not associated with significant variability in the performances of FibroTest.

Prognostic Value of Fibrosis Biomarkers in CHB

Only one published study was identified concerning FibroTest [18••]. The first aim of this study was to compare the 4-year prognostic value of FibroTest and ActiTest to that of viral load and to that of biopsy (in a subpopulation with simultaneous biopsy). The second aim of the study was to compare the 4-year prognostic value of combining FibroTest, ActiTest, and viral load for a better definition of the inactive HBV carrier status.

The main endpoint was the absence of liver-related complications. Liver-related complications were defined as death, liver transplantation, or one of the following: decompensation (ascites, hepatic encephalopathy, or jaundice with total bilirubin >51 μmol/L), variceal bleeding, or hepatocellular carcinoma. The adjustment factors were age, sex, HBeAg, ethnic origin, alcohol consumption, HIV-Delta-hepatitis C virus coinfection, and treatment.

The study included 1,074 patients with baseline FibroTest, ActiTest, and viral load (41 years old, 69% male, 47% African, 27% Asian, 20% Caucasian, 15.5% coinfected, mean follow-up 7.7 years [2.5 years prospective and 5.2 years retrospective]). Manufacturers' definitions were used for normal FibroTest (≤0.27), and normal ActiTest (≤0.29). The prevalences of a three-class viral load in international units per milliliter (low = <log 3, intermediate = log 3–log 5, high = >log 5) were 55%, 28%, and 17%, respectively. The accuracy (AUROC [95% CI]) of FibroTest in 97 patients with simultaneous liver biopsies for the diagnosis of advanced fibrosis was similar to previous validations (AUROC=0.83 [0.71–0.91]), and higher than ALT (0.60 [0.47–0.71]) and viral load (0.53 [0.39–0.63]) (all *P*< 0.0001), and not different from liver biopsy (*P*=ns).

At 4-years' follow-up, 50 complications occurred (survival without complications was 93.4%) and 36 deaths (survival was 95.0%), including 27 related to HBV (survival was 96.1%). The prognostic value of FibroTest was higher than that of viral load or ALT when compared using AUROC curves (all *P*<0.0001), survival curves, and multivariate Cox model. Among the 336 patients (without coinfection with hepatitis C virus, Delta, or HIV) with the classic definition of inactive carrier, 74 (22%) had advanced

fibrosis presumed with FibroTest, and three died or had complications at 4 years' follow-up. A new definition of inactive carriers was proposed, with an algorithm combining "zero" scores for FibroTest and ActiTest (F0 and A0) and viral load classes. This new algorithm provided a 100% negative predictive value for the prediction of liver-related complications or death at 4 years.

Discussion

This overview largely confirmed in patients with CHB the diagnostic performance of FibroTest and LSM by Fibroscan, which was already extensively validated in patients with CHC [1–10]. These performances were highly significant both for the diagnosis of advanced fibrosis and for the diagnosis of cirrhosis. The other fibrosis biomarkers have been less studied, preventing any conclusions. Only two studies have been identified for Hepascore, but the performances were also similar to those in patients with CHC.

Strength of the Study

The methodology of the present overview permitted taking into account the spectrum effect in the assessment of performances and identification of factors associated with variability. The same biomarker can have an AUROC of 0.87 for discriminating stages from F0 to F4 and an AUROC of 0.51 for discriminating between stages F1 and F2 (Table 1). As in patients with CHC, there was a clear risk of misleading conclusions if biomarkers were indirectly compared without standardizations on fibrosis spectrum. The Obuchowski method permitted validation of the higher performance of FibroTest versus ALT for the diagnosis of fibrosis stages.

The methods permitted accounting for the spectrum variability between Asian and non-Asian patients as well as between HBeAg-positive and HBeAg-negative patients. There was also no impact of the authorship with same FibroTest performances among independent and not-independent publications.

Limits of the Study

*Reliability of Biomarkers*

The present overview was not designed to assess the failure rate and the reliability rate of the biomarkers, which are a main concern in the clinical performance. From recent large studies, LSM has a weakness in comparison to serum biomarkers, with a 3.1% failure rate and 15.8% unreliable results [39••,40•], compared to 0% failure for FibroTest and less than 5% unreliable

Table 1 Characteristics of the FibroTest diagnostic studies ($n=8$) for the staging of hepatic fibrosis in patients with chronic hepatitis B

Study	Number of patients	Methodology cause if mixed (n)	Age (y)	Stage/ prevalence	AUROC SE	F0 (%)	F1 (%)	F2 (%)	F3 (%)	F4 (%)	DANA	Size (mm)	Independent	Biopsy FT/ median days
FibroTest														
Myers et al. [10]	209	Prospective (42), Retrospective (167)	39	F2F3F4 0.29 / F4 0.09	0.78 0.04 / 0.87 0.04	76 (36)	72 (34)	32 (15)	10 (5)	19 (9)	2.31	18	No	1
Poynard et al. [11]	214	Prospective	40	F2F3F4 0.67 / F4 0.19	0.77 0.03	14 (8)	53 (25)	119 (24)	55 (26)	40 (19)	2.17	?	Mixed	150
Sebastiani et al. [12]	110	Prospective	43	F2F3F4 0.69 / F4	0.85 0.04	(14)	(18)	(36)	(12)	(20)	2.19	17	Yes	0
Zhao et al. [13]	123	Prospective	?	F2F3F4 / F4	0.81 0.04 / 0.72 0.07	25 (20)	27 (22)	31 (25)	29 (24)	11 (9)	2.20	?	Yes	?
Poynard et al. [14•]	924	Retrospective	39	F2F3F4 0.33 / F4 0.08	0.76 0.02 / 0.83 0.02	21 (2)	596 (65)	160 (17)	69 (7)	78 (8)	1.77	13	Mixed	90
Gui et al. [30]	100	Prospective	?	F2F3F4 / F4	0.84 0.05 / 0.86 0.07	?	?	?	?	?	?	?	Yes	0
Bottero et al. [31]	108	Prospective, HIV coinfected	42	F2F3F4 0.57 / F4 0.15	0.77 0.04 / 0.87 0.09	11 (9)	36 (33)	28 (26)	17 (16)	16 (15)	2.04	17	Yes	30
Bonnard et al. [32]	54	Prospective	35	F2F3F4 0.70 / F4 0.24	0.79 0.06 / 0.85 0.05	5 (8)	13 (22)	19 (32)	8 (14)	14 (24)	2.26	21	Mixed	120
FibroTest integrated database[a]	1303	Retrospective and prospective (1303)	39	F2F3F4 0.38 / F4 0.10	0.76 0.03 / 0.79 0.02	108 (8)	707 (54)	227 (17)	133 (10)	128 (10)	1.94	15[b]	Mixed	90

Hepascore														
Bottero et al. [31]	108	Prospective, HIV coinfected	42	F2F3F4 0.57 / F4 0.15	0.74 / 0.04 / 0.91 / 0.09	11 (10)	36 (33)	28 (26)	17 (16)	16 (15)	2.04	17	Yes	30
Zhou et al. [38]	146	Prospective	33	F2F3F4 0.47 / F4 0.10	0.77 / 0.06 / 0.78 / 0.05	15 (10)	63 (43)	26 (18)	27 (18)	15 (10)	2.26	>15	Yes	7
Elastography														
Fraquelli et al. [37]	104	Prospective	45	F2F3F4 0.52 / F4 0.16	? / ?	?	?	21 (20)	17 (16)	16 (16)	?	30	Yes	Same session
Miailhes et al. [36]	59	Prospective	43	F2F3F4 / F4	0.85 / 0.05 / 0.96 / 0.02	4 (7)	19 (32)	16 (27)	8 (14)	12 (20)	2.06	16	Yes	3
Marcellin et al. [33]	173	Prospective	40	F2F3F4 0.50 / F4 0.08	0.82 / 0.04 / 0.93 / 0.08	16 (9)	70 (41)	44 (25)	29 (17)	14 (8)	1.84	16	Yes	2
Wang et al. [35]	88	Prospective	?	F2F3F4 ? / F4 ?	0.86 / 0.05 / 0.89 / 0.04	?	?	?	?	?	?	15	Yes	?
Sporea et al. [34]	140	Retrospective	39	F2F3F4 0.77 / F4 0.05	0.66 / 0.05 / 0.97 / 0.08	1 (0.7)	32 (23)	67 (48)	33 (24)	7 (5)	1.47	>20	Yes	Same session
Bonnard et al. [32]	54	Prospective	35	F2F3F4 0.70 / F4 0.24	0.87 / 0.04 / 0.88 / 0.03	5 (8)	13 (22)	19 (32)	8 (14)	14 (24)	2.26	21	Mixed	120

[a] 54 patients have been excluded from the original studies as acute hepatitis suspected (ALT >622 IU/L; $n=36$) or coinfected with hepatitis delta virus ($n=18$)

[b] Available in 1,090 patients

? Not available in the article

ALT alanine aminotransferase, *AUROC* area under the receiver operating curve, *DANA* Difference between Advanced and Non Advanced mean fibrosis stages

Fig. 1 Meta-analysis of the area under the receiver operating curve (*AUROC*) observed in published studies of Fibrotest, HepaScore, and liver stiffness measurement (*LSM*) for the diagnosis of advanced fibrosis in patients with chronic hepatitis B. Standardized AUROCs (StAUROCs), taking into account the spectrum of fibrosis stages, were all significantly higher than the random 0.50 value (*P*<0.001). Mean StAUROC was 0.84 (95% CI, 0.80–0.87). There was no significant difference among performance of the three biomarkers: FibroTest StAUROC = 0.84 (0.79–0.86), HepaScore StAUROC = 0.80 (0.71–0.89), and LSM StAUROC = 0.89 (0.83–0.96). There was no significant heterogeneity between studies (Cochran heterogeneity test = 6.3, *P*=0.90)
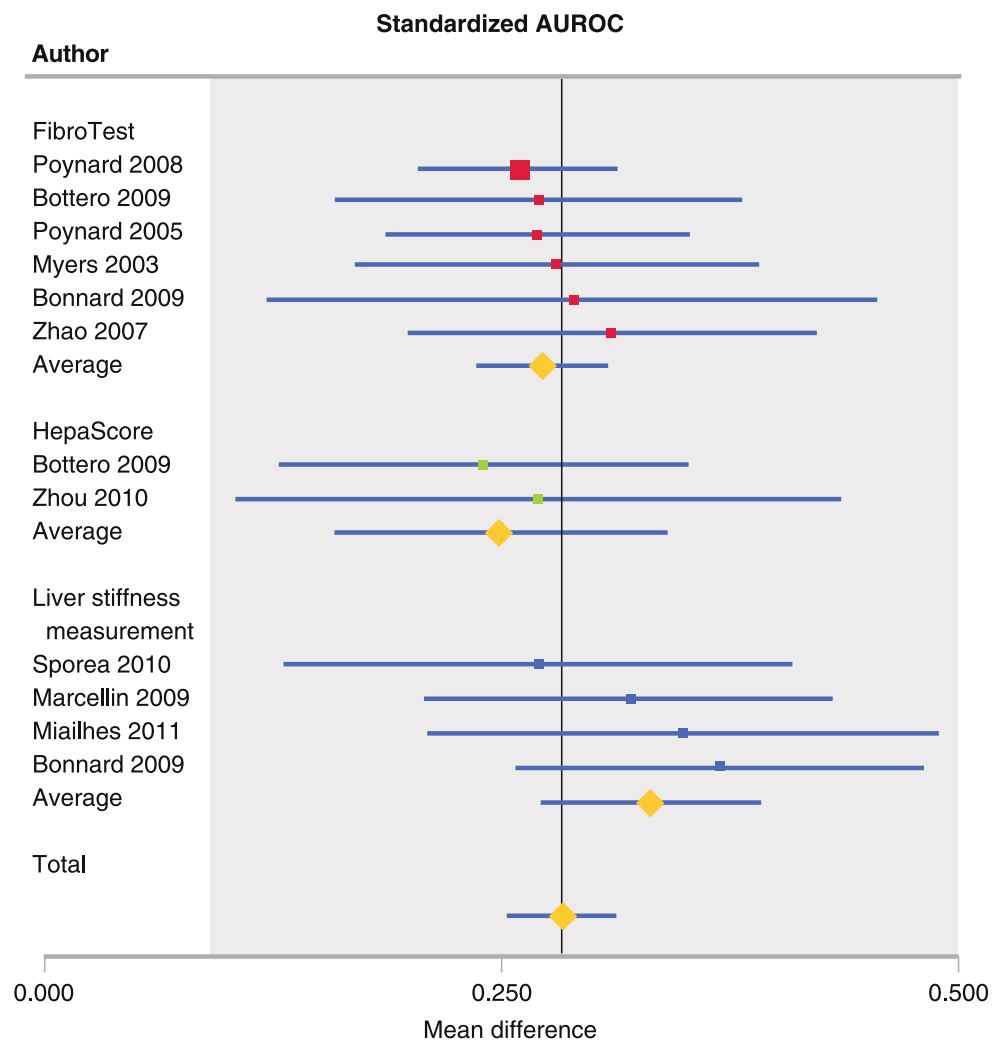
results [8]. The most frequent causes of unreliable results for Fibroscan are necro-inflammatory activity flares and metabolic factors, and for FibroTest, hemolysis and acute inflammation [8, 39••, 40•].

Individual Data

Another weakness was to have only individual data for FibroTest and not for LSM. Only the analysis of large databases can identify sources of variability in the diagnostic performances of biomarkers [20, 41••], including the variability of biopsy, the imperfect gold standard [27]. The mean length of biopsy was small in the present studies (usually <20 mm), including in the individual database (15 mm), and it was reassuring that the performances of Fibrotest were stable in the sensitivity analysis. Furthermore, standardization of AUROCs requires individual data, both for using DANA correction and Obuchowski measures [16••, 22, 23••].

Absence of a True Gold Standard

Because of the limitations of biopsy, even with 25-mm length, a need exists to improve the assessment of liver biomarker performances using methods without gold standards [41••, 42, 43•]. No latent class analyses have been performed in CHB, and only one study has demonstrated the prognostic value of FibroTest in CHB [18••], confirming those already observed in patients with CHC [43•] and alcoholic liver disease [44•].

Conclusions

This overview suggests that FibroTest and Fibroscan could be used as an alternative to liver biopsy in patients with CHB. However, for Fibroscan, a large analysis of individual data using Obuchowski measures seems necessary to assess the impact of necroinflammatory activity grades and steatosis,

Table 2 Sensitivity analyses of FibroTest diagnostic values according to patient characteristics

Characteristic (no. of patients)	Advanced fibrosis		Pairwise comparisons between all stages, Obuchowski measure AUROC (95% CI)
	Not standardized AUROC	Significance (P)	
All (1303)	0.76 (0.73–0.79)		0.844 (0.832–0.856)
Gender		>0.10	
Male (941)	0.75 (0.72–0.78)		
Female (362)	0.76 (0.70–0.81)		
Length of biopsy		>0.10	
<25 mm (942)	0.76 (0.71–0.85)		
≥25 mm (148)	0.79 (0.71–0.85)		
Not available (213)	0.76 (0.69–0.82)		
Ethnicity[a]		0.06	P>0.10
Asian (546)	0.72 (0.67–0.76)		0.836 (0.814–0.848)
Caucasian (601)	0.78 (0.74–0.82)		0.853 (0.839–0.867)
Sub-Saharan (156)[b]	0.80 (0.72–0.86)		
HBeAg[c,d]		0.046	P>0.10
Absent (548)	0.73 (0.68–0.77)[e]		0.854 (0.838–0.870)
Present (753)	0.78 (0.75–0.82)		0.831 (0.814–0.848)
After treatment		>0.10	
No (723)	0.75 (0.71–0.78)		
Yes (392)	0.78 (0.72–0.82)		
Not available (188)	0.78 (0.70–0.85)		
HBV genotype		>0.10	
Genotype A (61)	0.77 (0.57–0.88)		
Genotype non-A (137)	0.77 (0.73–0.79)		
Not available (1105)	0.76 (0.73–0.79)		
Viral load (IU/mL)		>0.10	
<3 Log (276)	0.79 (0.73–0.84)		
3–5 Log (115)	0.82 (0.72–0.88)		
5–7 Log (330)	0.72 (0.66–0.77)		
>7 Log (464)	0.77 (0.72–0.81)		
Not available (118)	0.76 (0.65–0.83)		
ALT <50 IU/L		>0.10	
No (825)	0.75 (0.72–0.79)		
Yes (478)	0.76 (0.70–0.80)		
ALT <25 IU/L		>0.10	
No (1151)	0.76 (0.73–0.79)		
Yes (152)	0.72 (0.57–0.82)		

[a] In Asian patients (n=546) the overall mean (95% CI) accuracy of FibroTest (standardized using Obuchowski measure) was 0.827 (0.807–0.846), no different than that of non-Asian (n=757), 0.853 (0.839–0.867), (NS)

[b] African American also included

[c] In 2 patients, status was unknown

[d] In patients with presence of HBeAg (n=753) the overall mean (95% CI) accuracy of FibroTest (standardized using Obuchowski measure) was 0.831 (0.814–0.848), no different than that of patients without HBeAg (n=548), 0.854 (0.838–0.870), (NS)

[e] P=0.046 vs "present"

ALT alanine aminotransferase, *AUROC* area under the receiver operating curve, *HBeAg* hepatitis B early antigen, *HBV* hepatitis B virus

particularly for the diagnosis of advanced fibrosis. However, neither biomarkers nor biopsy are sufficient, alone, to make a definitive decision in a given patient, and all the clinical and biological data must be taken into account.

This study reinforced our previous conclusion that, based on current evidence, a wise recommendation in patients with CHB would be a moratorium on liver biopsy as a first-line procedure, while awaiting studies demonstrating its cost utility versus that of biomarkers [45]. Biopsy as a second-line estimate of liver injury should still be indicated for intricate disease or clinicobiological discordances.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Castera L, Pinzani M. Non-invasive assessment of liver fibrosis: are we ready? Lancet. 2010;375:1419–20.
2. Gebo KA, Herlong HF, Torbenson MS, et al. Role of liver biopsy in management of chronic hepatitis C: a systematic review. Hepatology. 2002;36:S161–72.
3. 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.
4. Shaheen AA, Wan AF, Myers RP. FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy. Am J Gastroenterol. 2007;102:2589–600.
5. Poynard T, Morra R, Ingiliz P, et al. Biomarkers of liver fibrosis. Adv Clin Chem. 2008;46:131–52.
6. Poynard T, Morra R, Ingiliz P, et al. Assessment of liver fibrosis: noninvasive means. Saudi J Gastroenterol. 2008;14:163–73.
7. Fontaine H, Petitprez K, Roudot-Thoraval F, Trinchet JC. Guidelines for the diagnosis of uncomplicated cirrhosis. Gastroentérol Clin Biol. 2007;31:504–9.
8. Halfon P, Munteanu M, Poynard T. FibroTest-ActiTest as a non-invasive marker of liver fibrosis. Gastroentérol Clin Biol. 2008;32:22–39.
9. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology. 2008;134:960–74.
10. Myers RP, Tainturier MH, Ratziu V, et al. Prediction of liver histological lesions with biochemical markers in patients with chronic hepatitis B. J Hepatol. 2003;39:222–30.
11. Poynard T, Zoulim F, Ratziu V, et al. Longitudinal assessment of histology surrogate markers (FibroTest-ActiTest) during lamivudine therapy in patients with chronic hepatitis B infection. Am J Gastroenterol. 2005;100:1970–80.
12. Sebastiani G, Vario A, Guido M, Alberti A. Sequential algorithms combining non-invasive markers and biopsy for the assessment of liver fibrosis in chronic hepatitis B. World J Gastroenterol. 2007;13:525–31.
13. Zhao L, Xu D, Lu Z, et al. Validation of Fibrotest to diagnose liver fibrosis to patients with chronic hepatitis B. Chinese Journal of Practical Internal Medicine 2007;1274–1277.
14. • Poynard T, Ngo Y, Marcellin P, et al. Impact of adefovir dipivoxil on liver fibrosis and activity assessed with biochemical markers (FibroTest-ActiTest) in patients infected by Hepatitis B virus. J Viral Hepat. 2009;16:203–13. *This article describes the largest validation of serum biomarkers in treated and nontreated patients with CHB, with paired biopsies.*
15. •• Keeffe EB, Dieterich DT, Han SH, et al. A treatment algorithm for the management of chronic hepatitis B virus infection in the United States: 2008 update. Clin Gastroenterol Hepatol. 2008;6:1315–41. *This article is a complete and clear overview of the management of CHB. An updated version is needed, particularly for the recommendations concerning biomarkers.*
16. •• Lambert J, Halfon P, Penaranda G, et al. How to measure the diagnostic accuracy of non invasive liver fibrosis indices: the area under the ROC curve revisited. Clin Chem. 2008;54:1372–8. *This landmark study permitted assessment of the diagnostic value of a fibrosis biomarker without the risk of multiple testing between pairwise stage comparisons, taking into account the spectrum effect with a straightforward methodology. This paper confirmed the DANA method, but lacked individual data.*
17. 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.
18. •• Ngo Y, Benhamou Y, Thibault V, et al. An accurate definition of the status of inactive hepatitis B virus carrier by a combination of biomarkers (FibroTest-ActiTest) and viral load. PLoS ONE. 2008;3:e2573. *This article describes the first study showing that the prognostic value of serum biomarkers was equivalent to that of biopsy in patients with chronic hepatitis B. This study also suggested than combining biomarkers of Fibrosis and Activity are more predictive of clinical endpoints than viral load.*
19. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. The Standards for Reporting of Diagnostic Accuracy Group. Croat Med J. 2003;44:635–8.
20. Poynard T, Munteanu M, Morra R, et al. Methodological aspects for the interpretation of liver fibrosis non-invasive biomarkers: a 2008 update. Gastroentérol Clin Biol. 2008;32:8–21.
21. Poynard T, Ngo Y, Munteanu M, et al. Biomarkers of liver injury for hepatitis clinical trials: a meta-analysis of longitudinal studies. Antivir Ther. 2010;15:617–31.
22. Poynard T, Munteanu M, Ngo Y, et al. ActiTest accuracy for the assessment of histological activity grades in patients with chronic

hepatitis C, an overview using Obuchowski measure. Gastro-entérol Clin Biol. 2010;34:388–496.

23. •• Poynard T, Halfon P, Castera L, et al. Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. Clin Chem. 2007;53:1615–22. *This article describes the first study assessing the impact of spectrum effect on the not-standardized diagnostic performance of fibrosis biomarkers. This paper explains why there is no "specific" gray zone for serum fibrosis biomarkers between F1 and F2. A liver biopsy of 25-mm length has also a lower accuracy ("gray zone") between two adjacent stages. The same biomarker will have an AUROC of 0.80 for the diagnosis of F2F3F4 versus F0F1, 0.65 for the diagnosis of F2 versus F1, and 0.95 for the diagnosis of F4 versus F0.*

24. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating character-istic curves: a nonparametric approach. Biometrics. 1988;44:837–45.

25. Zhou X, Obuchowski N, McClish D. Statistical methods in diagnostic medicine. Wiley; 2002.

26. Piton A, Poynard T, Imbert-Bismut F, et al. Factors associated with serum alanine transaminase activity in healthy subjects: consequences for the definition of normal values, for selection of blood donors, and for patients with chronic hepatitis C. Multivirc Group. Hepatology. 1998;27:1213–9.

27. Poynard T, Halfon P, Castera L, et al. Variability of the area under the receiver operating characteristic curves in the diagnostic evaluation of liver fibrosis markers: impact of biopsy length and fragmentation. Aliment Pharmacol Ther. 2007;25:733–9.

28. Hintze J. NCSS 2007 user guide. 2007.

29. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Comput-ing; 2009. ISBN 3-900051-07-0, URL http://www.R-project.org.

30. Gui HL, Xie Q, Wang H, et al. FibroTest-ActiTest for predicting liver fibrosis and inflammatory activity in Chinese patients with chronic hepatitis B. Zhonghua Gan Zang Bing Za Zhi. 2008;16:897–901.

31. Bottero J, Lacombe K, Guéchot J, et al. Performance of 11 biomarkers for liver fibrosis assessment in HIV/HBV co-infected patients. J Hepatol. 2009;50:1074–83.

32. Bonnard P, Sombié R, Lescure FX, et al. Comparison of elastography, serum marker scores, and histology for the assessment of liver fibrosis in HBV-infected patients in Burkina Faso. Am J Trop Med Hyg. 2010;82:454–8.

33. Marcellin P, Ziol M, Bedossa P, et al. Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B. Liver Int. 2009;29:242–7.

34. Sporea I, Sirli R, Deleanu A, et al. Liver stiffness measurements in patients with HBV vs HCV chronic hepatitis: a comparative study. World J Gastroenterol. 2010;16:4832–7.

35. Wang JH, Changchien CS, Hung CH, et al. FibroScan and ultrasonography in the prediction of hepatic fibrosis in patients with chronic viral hepatitis. J Gastroenterol. 2009;44:439–46.

36. Miailhes P, Pradat P, Chevallier M, et al. Proficiency of transient elastography compared to liver biopsy for the assessment of fibrosis in HIV/HBV-coinfected patients. J Viral Hepat. 2011;18:61–9.

37. Fraquelli M, Rigamonti C, Casazza G, et al. Etiology-related determinants of liver stiffness values in chronic viral hepatitis B or C. J Hepatol. 2010. doi:10.1016/j.jhep.2010.07.017

38. Zhou K, Gao CF, Zhao YP, et al. Simpler score of routine laboratory tests predicts liver fibrosis in patients with chronic hepatitis B. J Gastroenterol Hepatol. 2010;25:1569–77.

39. •• 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. *This large, prospective study demonstrated the failure rate (2.8%) and the high rate of unreliable results (15.8%) of liver stiffness measurements using FibroScan. Interobserver variability was also observed, contrary to initial studies.*

40. • Coco B, Oliveri F, Maina AM, et al. Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. J Viral Hepat. 2007;14:360–9. *This study clearly demonstrated the risk of false-positive diagnosis of advanced fibrosis due to flares in patients with chronic hepatitis B.*

41. •• Poynard T, Ingiliz P, Elkrief L, et al. Concordance in a world without a gold standard: a new non-invasive methodology for improving accuracy of fibrosis markers. PloS ONE. 2008;3: e3857. *This article describes the first study demonstrating that the diagnostic value of fibrosis biomarkers can be assessed without the use of biopsy, an imperfect gold standard. It was also the first study using latent class method to conciliate the different false-positive and false-negative rates observed for FibroTest, Fibroscan, and biopsy.*

42. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology. 2003;38:1449–57.

43. • Ngo Y, Munteanu M, Messous D, et al. A prospective analysis of the prognostic value of biomarkers (FibroTest) in patients with chronic hepatitis C. Clin Chem. 2006;52:1887–96. *This article describes the first study showing that the prognostic value of serum biomarkers was equivalent to that of biopsy in patients with CHC.*

44. • Naveau S, Gaude G, Asnacios A, et al. Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. Hepatology. 2009;49:97–105. *This article describes the first study showing that the prognostic value of serum biomarkers was equivalent to that of biopsy in patients with alcoholic liver disease.*

45. Poynard T. First-line assessment of patients with chronic liver disease with non-invasive techniques and without recourse to liver biopsy. J Hepatol. 2011;54:586–587.