REVIEW

FibroTest/Fibrosure for Significant Liver Fibrosis and Cirrhosis in Chronic Hepatitis B: A Meta-Analysis

Nermin N. Salkic, MD, PhD[1], Predrag Jovanovic, MD, PhD[1], Goran Hauser, MD, PhD[2] and Majda Brcic, MD[1]

OBJECTIVES: **Extent of liver fibrosis is one of the most important factors in determining prognosis and the need for active treatment in chronic hepatitis B (CHB). Noninvasive alternatives such as FibroTest/Fibrosure (FT) have been developed in order to overcome the shortcomings of liver biopsy (LB). We aimed to systematically review studies describing the diagnostic accuracy of FT for predicting CHB-related fibrosis.**

METHODS: **MEDLINE and EMBASE searches and hand searching methods were performed to identify studies that assessed the diagnostic accuracy of FibroTest in HB patients using LB as a reference standard. We used a hierarchical summary receiver operating curves model and the bivariate model to produce summary receiver operating characteristic curves and pooled estimates of sensitivity and specificity.**

RESULTS: **We included 16 studies (*N*=2494) and 13 studies (*N*=1754) in the heterogenous meta-analysis for liver fibrosis and cirrhosis, respectively. The area under the hierarchical summary receiver operating curve for significant liver fibrosis and for all included studies was 0.84 (95% confidence interval (CI): 0.78–0.88). At the FT threshold of 0.48, the sensitivity, specificity, and diagnostic odds ratio (DOR) of FT for significant fibrosis were 61 (48–72%), 80 (72–86%), and 6.2% (3.3–11.9), respectively. The area under the hierarchical summary receiver operating curve for liver cirrhosis and for all included studies was 0.87 (95% CI: 0.85–0.90). At the FT threshold of 0.74, the sensitivity, specificity, and DOR of FT for cirrhosis were 62 (47–75%), 91 (88–93%), and 15.7% (8.6–28.8), respectively.**

CONCLUSIONS: **FibroTest is of value in exclusion of patients with CHB-related cirrhosis, but has suboptimal accuracy in the detection of significant fibrosis and cirrhosis. It is necessary to further improve the test or combine it with other noninvasive modalities in order to improve accuracy.**

Am J Gastroenterol 2014; 109:796–809; doi:10.1038/ajg.2014.21; published online 18 February 2014

INTRODUCTION

Chronic hepatitis B (CHB) is a major public health problem worldwide, with more than two billion people showing evidence of exposure and more than 240 million people showing evidence of chronic infection (1).

Substantial progress has been made in the treatment of chronically infected patients in the last couple of years. However, precise definition of the extent of the liver fibrosis in CHB remains one of the most important factors determining both the risk of further progression of the disease and the need for active treatment (2). The most recent guidelines by the European Association for Study of Liver recommend liver biopsy (LB) to determine the degree of necroinflammation and fibrosis, as hepatic histology can assist the decision to start treatment (3). In addition, there is a need to monitor treatment effects, with recovery of liver histology being one of the most important signs of success.

LB with subsequent histological analysis is considered as a gold standard that allows evaluation of presence and extent of the fibrotic process in liver tissue (4). However, LB is an invasive procedure associated with significant patient discomfort and a small but important risk of complications, with reported risk of hospitalization ranging from 1 to 5%, with a risk of severe complications of 0.57% and mortality rates varying from 0.009 to 0.12% (5–7). It is also prone to sample variability and accuracy seems to depend highly on the size of the sample (8,9). Another important limitation derives from the fact that the biopsy sample is evaluated by

[1]Department of Gastroenterology and Hepatology, University Clinical Center Tuzla, Tuzla, Bosnia and Herzegovina; [2]Department of Gastroenterology, Clinical Hospital Center Rijeka, Rijeka, Croatia. **Correspondence:** Nermin N. Salkic, MD, PhD, Department of Gastroenterology and Hepatology, University Clinical Center Tuzla, Trnovac bb, 75000 Tuzla, Bosnia and Herzegovina. E-mail: snermin@gmail.com
Received 18 October 2013; accepted 30 December 2013

histological scores, which have a certain amount of interobserver variability and are also dependent on the experience of the pathologist (10).

As a result of these issues, numerous investigators have attempted to devise noninvasive methods of assessing hepatic fibrosis, resulting in more than 20 different clinical scores or imaging modalities with variable diagnostic accuracy (11). Most attention has been focused on whether noninvasive methodologies can detect the presence or absence of minimal (i.e., F0–F1), significant (i.e., ≥F2), or advanced (i.e., ≥F3–F4) fibrosis according to the METAVIR histological score (12,13).

One of the most investigated and most frequently used tools is the FibroTest/Fibrosure (FT) (proprietary formula; Biopredictive, Paris, France)—a patented calculation of the combination of five serum biochemical parameters (α-2-macroglobulin, apolipoprotein A1, haptoglobin, L-glutamyltranspeptidase, and bilirubin)—which was developed by Poynard and colleagues (14,15). Clear advantages of FT include high applicability (>95%), widespread availability, and inter-laboratory reproducibility (16); however, there are also numerous drawbacks such as cost, failed external validation, lack of specificity for liver disease (results can be severely impaired by comorbidities, i.e., Gilbert's syndrome or hemolysis) (17), and difficulty in differentiating intermediate stages of fibrosis (18). Assessment of liver fibrosis without biopsy is very tempting, and despite the fact that recommendations suggest that noninvasive tests are still not ready to replace LB (2,18), FT has become widely present in clinical practice. Its accuracy for detection of fibrosis or even disease prognosis has been evaluated extensively in a variety of liver diseases and in several systematic reviews (19–24).

Despite its omnipresence, FT has not been as extensively studied in the CHB population as in chronic hepatitis C. There are several possible reasons for this: apart from the stage of fibrosis, other factors such as HBe antigen positivity, levels of alanine aminotransferase, and HB virus (HBV) DNA have important roles in deciding how and when to treat patients with CHB, and liver inflammation and HBV replication may confound interpretation of FT results (11). The most recent guidelines have indeed recognized the unresolved issue of the true place of FT and all noninvasive markers in the evaluation and follow-up of CHB patients, and recommend further development (3).

Currently, we are aware of only one meta-analysis focused on accuracy of FT and other noninvasive markers of hepatic fibrosis in CHB patients (22). However, it is somewhat limited by the relatively small number of included studies, along with the inclusion of predominantly single-center data on FT, which may impair its reproducibility and interpretability. In additon, there is an apparent diversity in the statistical methodology of meta-analysis in most of the reviews; most of them use various methods to analyze areas under the receiver operating characteristics curve (AUCs), yet none are prepared according to the methodology recommended by the Cochrane Collaboration Diagnostic Test Accuracy (DTA) Working Group (25). This may be an important issue; from a clinical standpoint, reported AUCs do give insights into the overall accuracy of a diagnostic test, but it is much more useful for a clinician to know how well tests perform at a certain threshold, measured by sensitivity, specificity, positive and negative predictive values, and likelihood ratio (LR).

Our primary objective was therefore to perform an independent meta-analysis of the diagnostic accuracy of FT for the prediction of significant liver fibrosis (F2–F4 vs. F0–F1) in CHB patients. Our secondary goal was to evaluate the diagnostic accuracy of FT for the prediction of liver cirrhosis (F4 vs. F0–F3) in CHB patients.

METHODS

Inclusion and exclusion criteria

We used inclusion criteria proposed by the Cochrane Collaboration DTA Working Group (26). We included all diagnostic cross-sectional studies, cohort studies, and randomized studies of test accuracy that compared FT accuracy with LB with fibrosis grade, assessed according to the METAVIR scale (12) (the reference standard) or any other scale.

We adopted wide initial inclusion criteria as many studies on this topic are inadequately reported and all data could rarely be extracted. Therefore, realistic inclusion criteria to initially include all studies reporting at least AUC were expected to allow the assessment and inclusion of a larger number of studies that could be managed with additional analyses, if sufficient data were found. No language restrictions were imposed.

Study participants comprised adult patients diagnosed with CHB. Studies including patients with other etiologies of liver disease were included if separate data for HBV-infected patients could be extracted; we also included human immunodeficiency virus (HIV) coinfected patients for a subsequent separate sensitivity analysis. We included all studies that

- reported that all patients had undergone LB and FT;
- allowed the possibility of obtaining the data necessary to create at least one 2×2 table of test performance (with numbers of true and false positives and negatives); and
- reported the method used for definition of the fibrosis grade.

We excluded studies including patients belonging to the pediatric population, hepatitis C/HBV coinfected patients, mixed chronic liver disease patients (but not CHB + non-alcoholic fatty liver disease), and liver/kidney transplant patients, as well as studies that were clearly extensions of previously published cohorts (where this was uncertain, authors were contacted for confirmation). We also excluded studies in which we were not able to obtain sufficient data for statistical analysis, as further described below.

Search methods and methodological assessment

We searched MEDLINE, EMBASE, and the Cochrane Library using the following search terms: FibroTest, Fibrosure, noninvasive/non invasive/non-invasive marker, liver fibrosis, liver fibrosis/cirrhosis, liver fibrosis assessment, liver fibrosis biomarkers, significant liver fibrosis, and advanced liver fibrosis. Recursive searches and cross-referencing were carried out using a "related citations" option in PubMed. No language or time limitations were used. Additional studies were identified by hand searching the reference lists of identified studies and review articles, and the

websites of the company marketing FT (www.biopredictive.com). We also hand searched tables of contents of key gastroenterology and hepatology journals (*Gastroenterology*, *Hepatology*, *Journal of Hepatology*, *Gut*, *Journal of Viral Hepatitis*, *American Journal of Gastroenterology*, *Liver International*, and *Alimentary Pharmacology & Therapeutics*) from January 2001 to September 2013.

Two review authors independently conducted the assessment of the titles and abstracts for eligibility and methodological quality. Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (27). This validated tool was designed to assess the internal and external validity of diagnostic accuracy studies included in systematic reviews, and is recommended by the Cochrane DTA Working Group (28). We used a standardized QUADAS-2 form, available from the developers website, for assessing the methodological quality of each included study (29). Disagreements between two authors were resolved with discussion between the two review authors, with a third author as final arbiter.

Data extraction and management

A standardized data extraction form was created to extract study design features and results data from each publication. For each study, two authors extracted data independently. We extracted year of publication, study design, sample size, presence of HIV coinfection, the QUADAS-2 methodological items, prevalence of each fibrosis stage on LB, along with total prevalence of significant fibrosis and cirrhosis, interval between biopsy and blood sampling for FT, size of LB sample, independence of study authors from FT developers, and type of scoring system used for histology (METAVIR vs. other).

We also recorded or derived the numbers of true positives, true negatives, false positives, and false negatives, or number of reported sensitivity, specificity, positive and negative predictive value with used cutoff value of FT and AUCs for both significant fibrosis and cirrhosis, if available. If data were not available in the publication, corresponding authors were contacted to provide supplementary data. In the case of inability to obtain data from the corresponding author, the respective study was excluded. Collected data were organized in a spreadsheet and a third author reviewed and resolved disagreements.

Poynard and colleagues proposed standardization of AUCs according to prevalence of fibrosis stages using difference between advanced and non-advanced fibrosis (DANA) in order to overcome the impact of spectrum bias on AUC estimates (30). Therefore, we calculated both DANA and standardized AUC according to formulas proposed by the authors—see **Table 1** (30).

Statistical analysis and data synthesis

We analyzed the included studies according to the methodology suggested by the Cochrane DTA Working Group (25). This methodology gives more usable results from a clinical point of view, as it is focused on two statistical measures of diagnostic accuracy: the sensitivity of the test (the proportion of those with the disease who have an abnormal test result) and the specificity of the test (the proportion of those without the disease who have a normal test result).

We included only studies in which we were able to obtain data to populate 2x2 tables. Initial analysis was performed using the Review Manager (RevMan 5.2, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration). After preparing and exporting data from RevMan, we used METADAS, an SAS macro (SAS 9.2, SAS Institute, Cary, NC), for meta-analysis of diagnostic accuracy studies, to compute the pooled sensitivity and specificity, and to plot the summary receiver operating characteristics curve with summary point and corresponding 95% confidence region (31). METADAS can fit two statistical models: the hierarchical summary receiver operating curves and the bivariate model (31,32). For the computation of the pooled absolute sensitivity and specificity, we fitted the bivariate model.

For calculation of the relative sensitivity and specificity, we used the bivariate model in METADAS by adding a covariate for the test, which estimates differences in logit sensitivity and logit specificity. When convergence failed for the bivariate model, accuracy parameters were estimated by omitting the correlation between the logit of the true positivity rate and the logit of the false positivity rate.

Large differences between studies are commonly noted in DTA meta-analyses, so heterogeneity is presumed to exist and random effects models are fitted by default.

Investigations of heterogeneity and sensitivity analyses. Factors that could have an impact on diagnostic accuracy included those involving methodological quality and study design, characteristics of the underlying population, and characteristics of the index and reference test. Multiple regressions were performed using the METADAS macro, with each time point providing another covariate to verify the influence of the chosen covariate on the accuracy estimates.

Factors such as study design, methodological quality, significant fibrosis/cirrhosis prevalence, applied cutoff (recommended by developers vs. others), size of biopsy sample, time from biopsy to blood sampling for FT, independence of authors, presence of HIV coinfection, and use of histological scores other than METAVIR were used to explore any heterogeneity discovered in the analysis and to assess the impact of heterogeneity on the relative accuracy.

Where differences were present across studies, we controlled for heterogeneity by conducting sensitivity analyses; in particular, we investigated diagnostic accuracy in studies that

(i) applied an FT cutoff that differed from that proposed by the developers,
(ii) had authors who were independent from the FT developers,
(iii) reported prevalence of significant fibrosis ≥54%, and cirrhosis >18%,
(iv) included patients who were HIV coinfected, and
(v) graded LB samples with histological scores other than METAVIR.

It was reported that in a large group of patients (N=2235) with chronic hepatitis C, naturally observed prevalence of significant fibrosis and cirrhosis was 54 and 18%, respectively (30,33). As we were unable to find similar data for hepatitis B patients, we used the abovementioned rates for sensitivity analysis regarding the prevalence.

Table 1. Characteristics of the included studies

Study	Type	HIV/HBV	METAVIR	Independent	Biopsy size (mm)	Biopsy to FT time (days)	Sample	Prevalence		FT cutoff	AUC		StAUC[b]
								F2–F4	F4	F2–F4/F4	F2–F4/F4	DANA[a]	F2–F4/F4
Myers *et al.* (46)	Mixed	No	Yes	No	18	1	209	0.29	0.09	0.4/0.8	0.78/NA	2.30	0.80/NA
Zhao *et al.* (47)	Prospective	No	No	Yes	>15	NA	123	0.58	0.09	0.31/0.72	0.81/0.72	2.20	0.84/0.75
Sebastiani *et al.* (48)	Prospective	No	Yes	Yes	17	0	110	0.69	0.20	0.48/0.74	0.85/0.76	2.20	0.88/0.79
Gui *et al.* (49)	Prospective	No	No	Yes	15	NA	100	0.39	0.12	0.4/0.55	0.84/0.86	NA	NA
Poynard *et al.* (50)	Retrospective	No	Yes	No	13	150	462	0.33	0.08	0.48/0.74	0.76/NA	1.77	0.81/NA
Bottero *et al.* (51)	Prospective	Yes	Yes	Yes	17	30	108	0.56	0.15	0.43/0.74	0.77/0.87	2.04	0.82/0.92
Bonnard *et al.* (52)	Prospective	No	Yes	No	21	120	54	0.76	0.26	0.37/0.5	0.79/0.85	2.16	0.83/0.89
Castera *et al.* (53)	Prospective	No	Yes	Yes	22.6	NA	60	0.73	0.25	0.48/0.74	0.71/0.74	2.48	0.71/0.74
Mbaye *et al.* (54)	Prospective	No	Yes	Yes	30	60	69	0.28	0	0.48/NA	0.55/NA	1.73	0.64/NA
Miailhes *et al.* (55)	Prospective	Yes	Yes	No	16	3	59	0.61	0.20	0.38/0.58	0.86/0.92	2.32	0.88/0.94
Stibbe *et al.* (56)	Prospective	No	Yes	Yes	20	0	48	0.46	0.10	0.31/0.75	0.8/0.81	2.00	0.85/0.86
Raftopoulos *et al.* (57)	Prospective	No	Yes	Yes	21	24	145	0.42	0.08	0.48/0.73	0.72/0.92	1.99	0.77/0.97
Sebastiani *et al.* (58)	Retrospective	No	Yes	Yes	17.7	NA	253	0.58	0.18	0.48/0.74	0.69/0.68	2.04	0.74/0.73
Kim *et al.* (59)	Prospective	No	No	Yes	21.3	0	170	0.71	0.28	0.31/0.67	0.90/0.88	2.29	0.92/0.90
Kim *et al.* (60)	Prospective	No	No	Yes	21	0	194	0.85	0.39	0.32/0.68	0.90/0.87	2.15	0.94/0.91
Park *et al.* (61)	Prospective	No	No	Yes	22	0	330	0.80	0.24	0.32/0.68	NA	1.91	NA

AUC, area under the receiver operating curve; DANA, difference between advanced and non-advanced fibrosis stages; stAUC, standardized AUC; FT, FibroTest/Fibrosure; HIV/HBV, hepatitis B and HIV-coinfected patients. METAVIR, liver biopsy assessed according to METAVIR or not; Independent, independence from FT developers; FT cutoff, FT cutoff used to predict; NA, data not available.

[a]DANA=[(prevalence F2 · 2+prevalence F3 · 3+prevalence F4 · 4)(prevalence F2+prevalence F3+prevalence F4)]−[prevalence F1(prevalence F0+prevalence F1)] (30).

[b]stAUC=AUC+0.0482×(2.5 DANA) (30).

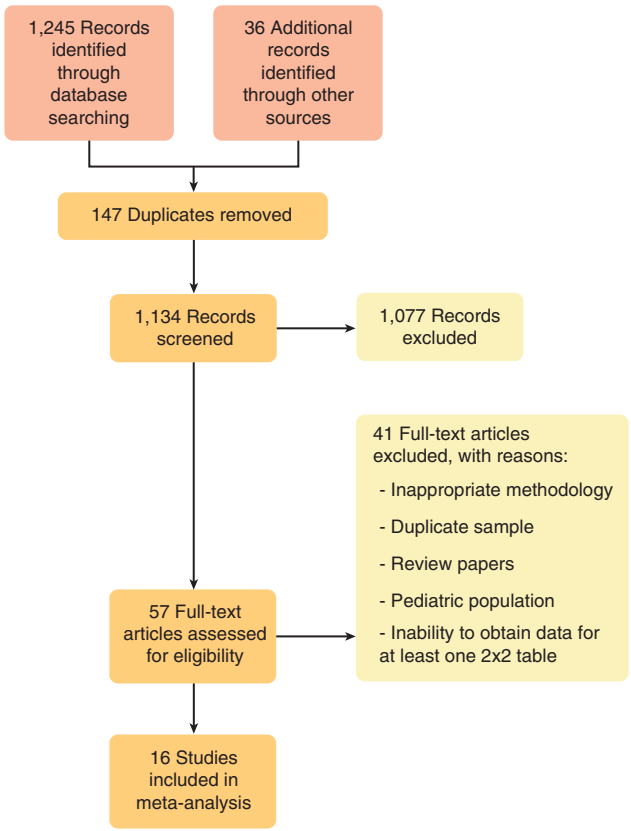

Figure 1. Flow diagram for electronic search and selection of studies.

We did not undertake any formal assessment of reporting bias in our review due to the current uncertainty about how to assess reporting bias in DTA reviews (34).

RESULTS

Search results

The results of the electronic database searching and hand searching are outlined in **Figure 1**. After eliminating duplicates, the initial electronic search identified 1,134 titles and abstracts for potential inclusion in the review. We obtained full-text copies of 57 studies. Of these, 31 studies were excluded for being systematic reviews or meta-analyses, belonging to the pediatric population, or dealing with prognostic potential of FT (i.e., prognosis of liver fibrosis progression, liver-disease-related death, and complications) without reporting data on its diagnostic accuracy. Additional 10 studies were excluded for having inappropriate statistical methodology, or presenting duplicated or insufficient data (35–45). Finally, we were able to include a total of 16 studies in the meta-analysis. The details of all studies included in the meta-analysis are reported in **Table 1**.

Characteristics of included studies

We included 16 studies with a total of 2,494 patients. The overall prevalence of significant fibrosis (F2–4) and cirrhosis (F4) was 55.33% (range: 28–85%) and 16.92% (range: 0–39%), respectively. Reported AUCs for diagnosis of significant fibrosis ranged from 0.55 to 0.90, whereas standardized AUCs ranged from 0.64 to 0.94. Calculated values of DANA ranged from 1.73 to 2.48. With the exclusion of one study with insufficient data on prevalence of fibrosis stages (49), the value of DANA calculated for 2,394 patients from 15 studies was 2.105.

Two studies ($N=167$) included HIV coinfected patients (51,55). In four studies ($N=784$), at least one of the authors belonged to the team of FT developers (46,50,52,55). In five studies ($N=917$), LB was assessed with a histological score other than METAVIR (47,49,59–61). In eight studies ($N=1070$), mean length of biopsy sample was ≥20 mm (52–54,56,57,59–61).

In four studies ($N=536$), we were not able to obtain data on time period between biopsy and serum sampling for FT calculation (47,49,53,58), and five studies ($N=852$) reported blood sampling on the day of biopsy (48,56,59–61). Six studies ($N=1,099$) used the recommended FT cutoff for fibrosis (48,50,53,54,57,58) and seven ($N=1,186$) used the recommended FT cutoff for cirrhosis (48,50,51,53,56–58). In one study, there was no reported case of cirrhosis (F4 according to METAVIR) (54).

A study by Poynard *et al.* (50) reported data on the same group of patients before and after the treatment, so we extracted data for FT-LB comparison before treatment, therefore eliminating repeated measurements.

The results of methodological quality assessment according to the QUADAS-2 scale are depicted for all of the 16 included studies (**Figure 2**). Most of the methodological concern lies within the reference standard, as five studies used a histological score other than METAVIR (47,49,59–61). Of these, three studies used the Batts and Ludwig scoring system (59–61), one study used the Ishak system (49), and one used the Scheuer system (47).

Another possible issue regarding methodological quality was raised in the study by Mbaye *et al.* (54), in which there was a significant possibility of selection bias as the authors selected for LB only those patients with liver stiffness measurement values between 7 and 13 kPa, therefore limiting the spectrum of fibrosis grades within the sample. Both of these concerns were addressed in heterogeneity and sensitivity analyses.

Diagnosis of significant fibrosis

As already noted, we included all 16 studies in the analysis for significant liver fibrosis (METAVIR F2-F4). Summary representation of the overall analysis is presented in **Figures 3** and **4**. Sensitivity ranged from 37 to 90%, whereas specificity ranged from 48 to 98%.

The area under the hierarchical summary receiver operating curves for significant liver fibrosis and for all studies was 0.84 (95% confidence interval (CI) = 0.78–0.88). The meta-analytical summary estimate, irrespective of the used FT threshold, corresponded with pooled sensitivity of 71.2% (95% CI = 64.6 to 77.1%), specificity of 81.4% (95% CI = 74.8 to 86.6%), positive LR (LR+) of 3.83 (2.77–5.31), and negative LR (LR−) of 0.35 (0.28–0.44). The diagnostic odds ratio (DOR) was therefore 10.85 (6.70–17.57). However, these measurements must be carefully considered as

Figure 2. Summary of methodological quality of studies according to Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool concerning risk of bias and applicability in review authors' judgments about each domain for each included study and review authors' judgments about each domain, presented as percentages across included studies.

Study	TP	FP	FN	TN	Cut-off	HIV/HBV	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Myers *et al.*(46)	33	30	28	118	0.4	No	≤0.54	0.54 (0.41, 0.67)	0.80 (0.72, 0.86)
Zhao *et al.*(47)	64	27	7	25	0.31	No	>0.54	0.90 (0.81, 0.96)	0.48 (0.34, 0.62)
Sebastiani *et al.*(48)	61	3	14	32	0.48	No	>0.54	0.81 (0.71, 0.89)	0.91 (0.77, 0.98)
Gui *et al.*(49)	29	5	10	56	0.4	No	≤0.54	0.74 (0.58, 0.87)	0.92 (0.82, 0.97)
Poynard *et al.*(50)	113	90	58	201	0.48	No	≤0.54	0.66 (0.58, 0.73)	0.69 (0.63, 0.74)
Bottero *et al.*(51)	43	13	18	34	0.43	Yes	>0.54	0.70 (0.57, 0.81)	0.72 (0.57, 0.84)
Bonnard *et al.*(52)	32	4	9	17	0.37	No	>0.54	0.78 (0.62, 0.89)	0.81 (0.58, 0.95)
Miailhes *et al.*(55)	24	3	7	18	0.38	Yes	>0.54	0.77 (0.59, 0.90)	0.86 (0.64, 0.97)
Mbaye *et al.*(54)	7	13	12	37	0.48	No	≤0.54	0.37 (0.16, 0.62)	0.74 (0.60, 0.85)
Castera *et al.*(53)	27	3	17	13	0.48	No	>0.54	0.61 (0.45, 0.76)	0.81 (0.54, 0.96)
Sebastiani *et al.*(58)	79	18	67	89	0.48	No	>0.54	0.54 (0.46, 0.62)	0.83 (0.75, 0.90)
Stibbe *et al.*(56)	19	8	3	18	0.31	No	≤0.54	0.86 (0.65, 0.97)	0.69 (0.48, 0.86)
Raftopoulos *et al.*(57)	26	17	22	80	0.48	No	≤0.54	0.54 (0.39, 0.69)	0.82 (0.73, 0.89)
Kim *et al.*(60)	130	2	34	28	0.32	No	>0.54	0.79 (0.72, 0.85)	0.93 (0.78, 0.99)
Kim *et al.*(59)	91	1	30	48	0.31	No	>0.54	0.75 (0.67, 0.83)	0.98 (0.89, 1.00)
Park *et al.*(61)	200	15	65	50	0.32	No	>0.54	0.75 (0.70, 0.81)	0.77 (0.65, 0.86)

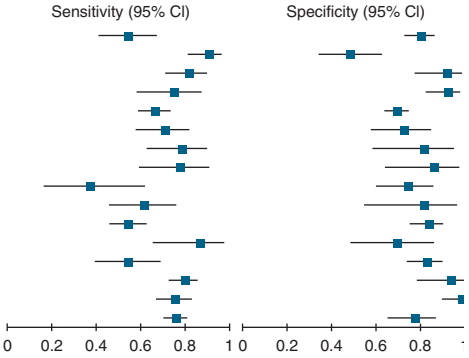

Figure 3. Forest plot of FibroTest/Fibrosure for detection of significant liver fibrosis (METAVIR F2–F4).

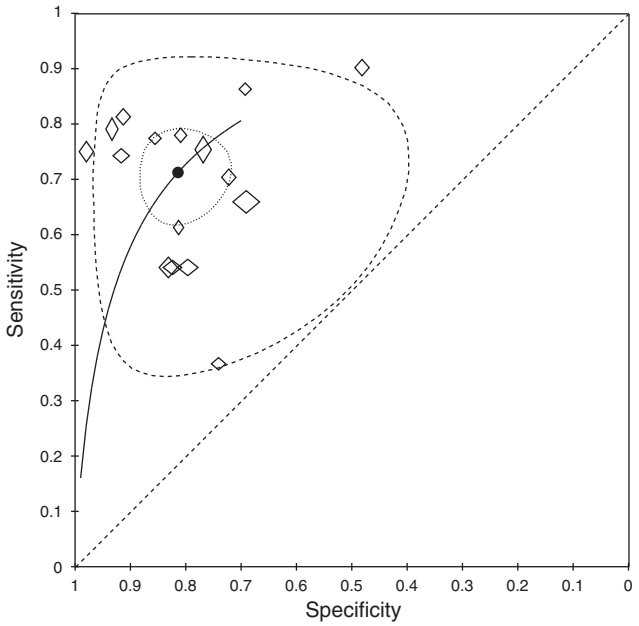

Figure 4. Summary receiver operating characteristic (ROC) plot of FibroTest/Fibrosure for detection of significant liver fibrosis (METAVIR F2–F4). The marked point on the curve represents the summary estimate of test performance, and the area delimited by dots represents the 95% confidence region of the summary estimate. The area delimited by the dashed line represents the 95% prediction region, within which there is a 95% confidence that the true sensitivity and specificity of a future study should lie.

they were not pooled from studies with identical FT threshold. Performance of tests in a cohort of 1,000 patients with a 50% prevalence of significant fibrosis is presented in **Table 2**. Overall, there was strong evidence of heterogeneity among the 16 included studies, as graphically illustrated on the forest plot and summary receiver operating characteristic plot in **Figures 3** and **4**.

Notably, there were also three distinct outliers positioned at the very edge of the prediction area (47,54,59). Comparison of models with and without outliers as covariates did not show any statistically significant change in a −2log likelihood ($\chi^2 = 0.83$; df = 3; $P = 0.842$); hence, there was no association of test performance with the presence of outliers. Comparison of pooled sensitivity and specificity for models with and without outliers is presented in **Table 2**.

We also evaluated the diagnostic performance of FibroTest according to various thresholds used. We summarized all thresholds into three categories according to the recommendations of the developers: cutoff of 0.31 (corresponding with F1), cutoff ranging from 0.32 to 0.48 (F1–F2), and cutoff of 0.48 (F2) (14). The cutoff according to these three categories was added as a covariate in the bivariate model. Comparison of models with and without cutoff class as covariate did not show any significant change in a −2log likelihood ($\chi^2 = 4.605$; df = 6; $P = 0.595$). There were no differences between pooled sensitivities and specificities for all three thresholds (**Table 2**).

However, there was a significant difference in pooled sensitivities (relative sensitivity 0.79; $P = 0.036$) between studies that used the recommended cutoff of 0.48 and those that used other threshold values for diagnosis of significant fibrosis, with the latter having better pooled sensitivity (**Table 2** and **Figure 5**). In addition, studies that used FT threshold <0.48 also exhibited a better pooled DOR (relative DOR: 2.39; $P = 0.05$). However, when studies with recommended FT cutoff were excluded, a reduction in variation of sensitivity was observed but with a greater degree of variation of specificity—see **Figure 5**.

The presence of HIV infection included as a covariate did not result in a significantly different model ($\chi^2 = 0.219$; df = 3; $P = 0.975$), nor were there differences in pooled sensitivities and specificities between studies that included HIV coinfected patients and those who did not.

The inclusion of prevalence of significant fibrosis ≥0.54 as a covariate did not result in a significantly different model according to a change in a −2log likelihood ($\chi^2 = 5.602$; df = 3; $P = 0.133$). Nevertheless, there was a tendency for better overall accuracy in studies with prevalence of significant fibrosis ≥0.54, according to relative difference in DOR (2.40; $P = 0.06$)—see **Table 2**.

There were no differences in pooled sensitivities, specificities, and DORs between studies reported by independent investigators and those reported by FT developers—see **Table 2**. The same was true for studies that reported the average length of biopsy sample to be ≥0.20 mm; they did not differ in pooled sensitivity, specificity,

Table 2. Summary of findings for diagnosis of significant fibrosis with an assessment of heterogeneity and clinical repercussions of findings when applied in a cohort of 1,000 people with a 50% prevalence of significant liver fibrosis

Subgroup	Sens	Spec	FN	FP	PPV	NPV	DOR	AUC
Overall	71.2%[1] (64.6–77.1%)	81.4%[2] (74.8–86.6%)	144	93	79.3%	73.9%	10.9* (6.7 to 17.6)	0.84
Presence of outliers: [1]P=*0.62;* [2]P=*0.66*								*P=0.94
Without outliers	70.7%[1] (63.0–77.4%)	82.0%[2] (75.4–87.1%)	147	90	79.7%	73.7%	10.5* (3.3–33.4)	0.83
FibroTest cutoff used: [3]P=*0.63;* [4]P=*0.14;* [5]P=*0.53;* [6]P=*0.95*								*P=0.78; $P=0.29
0.31 (F1)	76.5%[3,4] (65.1–85.0%)	78.8%[5,6] (62.7–89.2%)	117	106	78.3%	77.0%	12.1*,$ (4.8–30.2)	0.85
0.31–0.48 (F1–F2)	73.3%[3] (64.6–80.6%)	83.8%[5] (73.9–90.4%)	133	81	81.9%	75.8%	14.2* (7.1–28.5)	0.86
0.48 (F2)	62.3%[4] (46.8–75.6%)	79.4%[6] (69.0–86.9%)	188	103	75.2%	67.8%	6.3$ (2.9–14.1)	0.78
Recommended FibroTest cutoff used: [7]P=*0.036;* [8]P=*0.68*								*P=0.05
0.48	60.9%[7] (48.3–72.2%)	79.9%[8] (71.7–86.2%)	195	100	75.2%	57.1%	6.2* (3.3–11.9)	0.77
0.31–0.48	76.3%[7] (71.0–81.0%)	82.2%[8] (72.5–89.0%)	118	89	81.1%	77.6%	14.9* (8.2–27.0)	0.86
HIV/HBV coinfection: [9]P=*0.72;* [10]P=*0.84*								*P=0.98
Absent	70.1%[9] (63.7–77.1%)	81.7%[10] (74.7–87.1%)	92	150	79.3%	73.2%	10.8* (6.5–18.1)	0.83
Present	73.9%[9] (55.8–86.4%)	79.5%[10] (52.1–93.3%)	131	102	78.3%	75.3%	11.0* (2.5–47.6)	0.84
Prevalence of significant fibrosis (F2–F4): [11]P=*0.13;* [12]P=*0.39*								*P=0.06
<0.54	63.0%[11] (48.7–75.3%)	75.0%[12] (70.0–85.3%)	185	125	71.6%	67.0%	6.3* (3.2–12.4)	0.78
≥0.54	78.6%[11] (68.8–80.4%)	83.4%[12] (74.1–89.7%)	107	83	82.6%	79.6%	15.1* (8.3–27.4)	0.86
Author independence: [13]P=*0.55;* [14]P=*0.22*								*P=0.14
No	67.5%[13] (51.1–80.5%)	75.2%[14] (62.8–84.5%)	162	124	73.1%	69.8%	6.3* (2.7–14.5)	0.78
Yes	72.6%[13] (65.3–78.9%)	83.2%[14] (76.0–88.6%)	137	84	81.2%	75.2%	13.1* (7.8–22.3)	0.85
Average size of liver biopsy (length): [15]P=*0.84;* [16]P=*0.42*								*P=0.62
<20 mm	71.9%[15] (62.5–79.7%)	79.1%[16] (69.3–86.4%)	141	104	77.5%	73.8%	9.7* (5.0–18.6)	0.82
≥20 mm	70.6%[15] (60.5–78.9%)	83.7%[16] (74.7–90.0%)	147	82	81.2%	74.0%	12.3* (6.2–24.6)	0.85
Liver biopsy and FibroTest sampling on same day: [17]P=*0.23;* [18]P=*0.04;* [19]P=*0.17;* [20]P=*0.04*								*P=0.007; $P=0.06
NA	70.5%[17] (58.6–80.1%)	78.3%[19] (66.4–86.8%)	148	108	76.5%	72.6%	8.6$ (4.9–15.1)	0.81
No	63,5%[18] (51.1–74.3%)	76.0%[20] (69.1–81.8%)	182	120	72.6%	67.6%	5.5* (3.5–8.6)	0.76
Yes	77.9%[17,18] (73.2–81.9%)	89.2%[19,20] (71.0–96.6%)	110	54	87.8%	80.1%	29.1$,* (9.4–90.7)	0.91
METAVIR used for grading liver biopsy: [21]P=*0.01;* [22]P=*0.16*								*P=0.036
No	78.3%[21] (74.2–81.9%)	87.1%[22] (68.9–95.4%)	108	64	85.9%	80.1%	24.4* (8.1–73.3)	0.90
Yes	66.3%[21] (57.8–73.9%)	78.1%[22] (73.0–82.4%)	168	109	75.2%	69.9%	7.0* (4.7–10.5)	0.79

Sens, sensitivity; Spec, specificity; FN, -false negative; FP, false positive; PPV, positive predictive value; NPV, negative predictive value; DOR, diagnostic odds ratio; AUC, area under the ROC curve.
95% confidence intervals of selected measures are presented in parentheses.

or DOR from studies that reported the average length of biopsy sample to be <0.20 mm—see **Table 2**.

In the five studies that reported blood sampling for FT to have taken place on the same day as LB, the pooled sensitivity and specificity were significantly higher (relative sensitivity 1.23; $P=0.04$ and relative specificity=1.17; $P=0.04$) compared with studies reporting blood sampling after or before the day of LB—see **Table 2**. DOR was also significantly better in studies with FT blood sampling performed on the day of LB (relative DOR: 5.29; $P=0.007$). Studies in which data on interval of blood sampling were not available did not differ significantly from those with FT sampling on the day of LB, but with borderline statistical (in) significance ($P=0.06$).

In the five studies that did not use the METAVIR scoring system, the pooled sensitivity was significantly better than in those that used METAVIR (relative sensitivity=1.18; $P=0.01$). Although

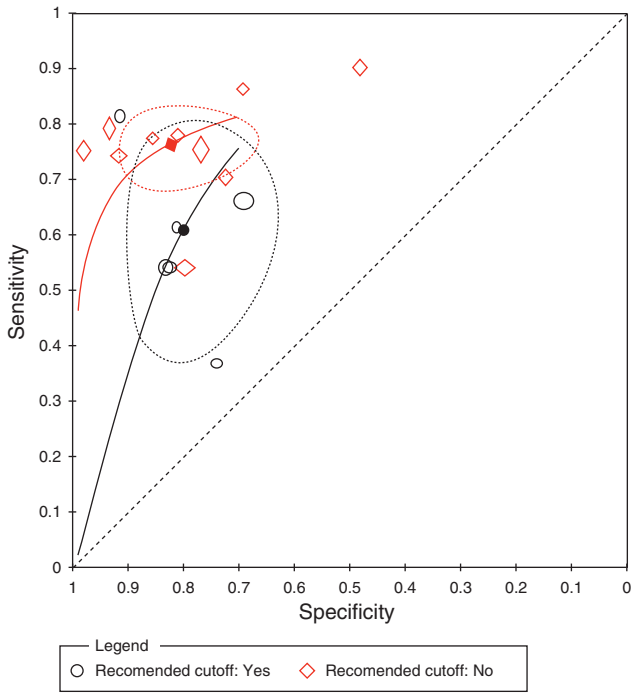

Figure 5. Summary receiver operating characteristic (ROC) plot of FibroTest/Fibrosure for detection of significant liver fibrosis (METAVIR F2–F4). The curves represent the summary ROC curves for studies that used the recommended FibroTest/Fibrosure (FT) cutoff of 0.48 and those that used other cutoffs. The marked point on the ROC curves represents the summary estimate of test performance and the area outline surrounding it represents the 95% confidence region of summary estimate.

pooled specificities did not differ (**Table 2**), pooled DOR was again better in studies that did not use the METAVIR (relative DOR: 3.48; $P=0.04$). However, there is one important caveat that needs to be considered as another possible explanation—all five studies originate from Asia (China and Korea).

We also evaluated seven studies with methodological concerns according to QUADAS-2 score—see **Figure 2** (46,47,49,54,59–61). Models with and without studies with methodological concerns did not significantly differ ($\chi^2=1.137$; df = 3; $P=0.768$). Pooled sensitivity and specificity were 70.3% (95% CI = 59.8–78.9%) and 79.9% (95% CI = 71.7–86.1%), respectively, in studies without methodological concerns and were 72.9% (95% CI = 64.4–80.1%) and 83.2% (95% CI = 72.2–90.4%), respectively, in studies with methodological concerns. There were no significant differences between pooled sensitivities and specificities ($P=0.67$ and 0.59, respectively).

Diagnosis of cirrhosis

We included 13 studies for the meta-analysis, with a total of 1,754 patients, as one study did not have any cases of liver cirrhosis (METAVIR F4) (54) and in two studies we were unable to derive data for the population of the 2×2 table (46,50). The overall prevalence of METAVIR F4 in the included studies was 20.76% (range: 9–39%). Reported AUCs for diagnosis of cirrhosis ranged from 0.68 to 0.92, whereas standardized AUCs ranged from 0.73 to 0.97.

Summary representation of the overall analysis is presented in **Figures 6** and **7**. The area under the hierarchical summary receiver operating curves for liver cirrhosis and for all included studies was 0.87 (95% CI = 0.85–0.90). Sensitivity ranged from 42 to 100%, much more widely than specificity, which ranged from 78 to 98% (**Figure 6**). The meta-analytical summary estimate, irrespective of the used FT threshold, corresponded with the pooled sensitivity of 71.5% (95% CI = 62.1–79.3%), specificity of 87.0% (95% CI = 83.8–89.6%), LR+ of 5.49 (4.62–6.54), and LR− of 0.33 (0.25–0.44). DOR was therefore 16.77 (12.07–23.30). Again, these measures must be carefully considered as they were not pooled from studies with identical FT thresholds. The test performance in a cohort of 1,000 patients with a 20% prevalence of cirrhosis is presented in **Table 3**.

There was evidence of heterogeneity with a strong presence of threshold effect, as depicted in **Figure 6**, based on the distribution of studies around the summary receiver operating characteristic curve and the shape of confidence area, as well as by a strong negative correlation of sensitivities and specificities ($\rho=-0.78$; $P=0.005$). After exclusion of two outliers, with a reported sensitivity of 100% (55, 56), the model without outliers differed significantly from the model with outliers ($\chi^2=11.54$; df = 3; $P=0.009$); however, the values of pooled sensitivity and specificity changed marginally (69.4 and 87.0%, respectively).

Studies that used the recommended FT threshold of 0.74 had better pooled specificity than those that used lower FT thresholds (**Table 3**). Although there was a tendency toward lower pooled sensitivity in studies that used FT threshold 0.74 (as expected), this was not statistically significant ($P=0.06$). There was no difference in pooled DORs, either. Nevertheless, in studies that used FT cutoff for cirrhosis < 0.74, there was clearly reduced heterogeneity (**Figure 8**).

Subgroup analysis according to presence of HIV coinfection, prevalence of cirrhosis > 0.18, independence of authors, length of biopsy sample ≥20 mm, and use of a histology scoring system other than METAVIR did not produce any significant differences in pooled sensitivities, specificities, or DORs (**Table 3**).

Pooled sensitivities and specificities in studies that reported blood sampling for FT to have been performed on the same day as LB did not differ significantly from those that reported sampling after or before the day of LB, or from those that did not report the time interval between biopsy and blood sampling at all. However, pooled DOR in studies that did not report the time interval between biopsy and FT was significantly lower in comparison with the other two subgroups (**Table 3**).

We also evaluated seven studies with lower methodological quality according to QUADAS-2 score—see **Figure 2** (46,47,49,54,59–61). Models with and without studies with lower methodological quality did not significantly differ ($\chi^2=3.385$; df = 3; $P=0.336$). Pooled sensitivity and specificity were 70.4% (95% CI = 54.9–82.3%) and 89.0% (95% CI = 85.6–91.6%), respectively, in studies with better methodological quality and were 75.6% (95% CI = 65.2–83.9%) and 84.3% (95% CI = 78.9–88.6%), respectively,

Study	TP	FP	FN	TN	Cut-off	HIV/HBV	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Zhao *et al.*(47)	6	18	5	94	0.72	No	≤0.18	0.55 (0.23, 0.83)	0.84 (0.76, 0.90)
Sebastiani *et al.*(47)	12	4	10	84	0.74	No	>0.18	0.55 (0.32, 0.76)	0.95 (0.89, 0.99)
Gui *et al.*(49)	10	14	2	74	0.55	No	≤0.18	0.83 (0.52, 0.98)	0.84 (0.75, 0.91)
Bottero *et al.*(51)	12	14	4	78	0.74	Yes	≤0.18	0.75 (0.48, 0.93)	0.85 (0.76, 0.91)
Bonnard *et al.*(52)	11	9	2	32	0.5	No	>0.18	0.85 (0.55, 0.98)	0.78 (0.62, 0.89)
Raftopoulos *et al.*(57)	17	14	5	109	0.73	No	≤0.18	0.77 (0.55, 0.92)	0.89 (0.82, 0.94)
Miailhes *et al.*(55)	10	8	0	34	0.58	Yes	>0.18	1.00 (0.69, 1.00)	0.81 (0.66, 0.91)
Sebastiani *et al.*(58)	19	18	26	190	0.74	No	≤0.18	0.42 (0.28, 0.58)	0.91 (0.87, 0.95)
Stibbe *et al.*(56)	5	1	0	42	0.75	No	≤0.18	1.00 (0.48, 1.00)	0.98 (0.88, 1.00)
Castera *et al.*(53)	7	3	8	35	0.74	No	>0.18	0.47 (0.21, 0.73)	0.92 (0.79, 0.98)
Kim *et al.*(59)	37	26	10	97	0.67	No	>0.18	0.79 (0.64, 0.89)	0.79 (0.71, 0.86)
Kim *et al.*(60)	60	19	15	100	0.68	No	>0.18	0.80 (0.69, 0.88)	0.84 (0.76, 0.90)
Park *et al.*(61)	54	29	24	223	0.68	No	>0.18	0.69 (0.58, 0.79)	0.88 (0.84, 0.92)

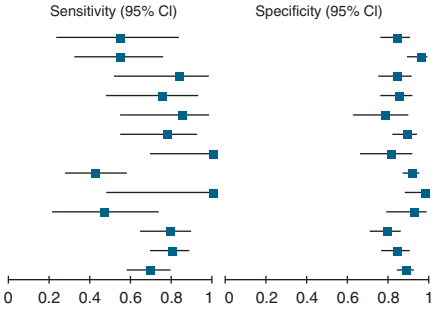

Figure 6. Forest plot of FibroTest/Fibrosure for detection of liver cirrhosis (METAVIR F4).

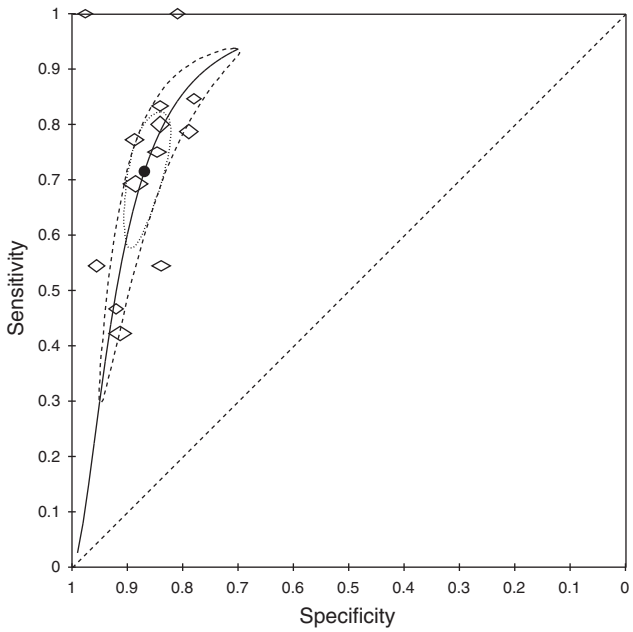

Figure 7. Summary receiver operating characteristic (ROC) plot of FibroTest/Fibrosure for detection of liver cirrhosis (METAVIR F4). The marked point on the curve represents the summary estimate of test performance, and the area delimited by dots represents the 95% confidence region of the summary estimate. The area delimited by the dashed line represents 95% prediction region, within which there is a 95% confidence that the true sensitivity and specificity of a future study should lie.

in studies with lower methodological quality. There were no significant differences between pooled sensitivities and specificities ($P=0.54$ and 0.11, respectively).

DISCUSSION

In this meta-analysis we have summarized the diagnostic accuracy of FT for CHB-related significant fibrosis and liver cirrhosis. This is the largest review of the diagnostic accuracy of FT in CHB patients, and the largest review by independent investigators. In addition, as statistical aspects of a systematic review of DTA still present challenges, it is of utmost importance to use validated and robust methodology that will provide usable results from a clinical standpoint; this is the main reason why we have chosen the methodological approach recommended by the Cochrane DTA Group. Additional strengths of our review include a meticulous search of published studies, formal assessment of methodological quality, assessment of heterogeneity with sensitivity analysis, and assessment of clinical repercussions of our findings.

One of the main limitations is the significant heterogeneity of included studies. A considerable variation between the results of diagnostic studies is a common occurrence, possibly to a greater extent than is seen for therapeutic interventions (62). This is perhaps one of the main sources of heterogeneity and a direct consequence of the fact that the importance of rigorous design has been less well appreciated for diagnostic studies than for therapeutic interventions, resulting in a poorer adherence to methodological constraints. This is noticeable in many studies that we included, and can be considered as a general problem in many studies dealing with the diagnostic accuracy of liver fibrosis markers, as already noted by others (63). We have also chosen to evaluate only FibroTest, despite the availability of other tests, mainly because of the limited number of publications describing the accuracy of other such tests in CHB patients. Some potentially eligible studies were excluded because we could not derive data to construct 2×2 tables. We also did not take into consideration the prognostic value of FT in this review—there is clear potential for the utility of FT in this area too (64).

Overall, our results in terms of pooled AUCs for significant fibrosis and cirrhosis are in agreement with findings reported from a previous meta-analysis conducted by FT developers (AUC: 0.79 and 0.84 for significant fibrosis and cirrhosis, respectively) (22). Interestingly, these measurements were similar to the findings reported in an independent meta-analysis of FT in HCV patients (AUC: 0.81 and 0.90 for significant fibrosis and cirrhosis, respectively) (24). It appears that with the current (imperfect) gold standards such as LB, these are the best numbers for FT we can achieve in terms of overall accuracy. This is confirmed by the fact that all direct, indirect, and combined serum markers of liver fibrosis have AUCs clustering around the value of 0.85 (13). Therefore, when discussing the accuracy of any marker in the case of discordant results between biopsy and biomarker such as FT, the cause of discordance can be either failure of fibrosis marker or failure of biopsy to detect true stage (17).

Table 3. Summary of findings for diagnosis of liver cirrhosis with an assessment of heterogeneity and clinical repercussions of findings when applied in a cohort of 1,000 people with 20% prevalence of liver cirrhosis

Subgroup	Sens	Spec	FN	FP	PPV	NPV	DOR	AUC
Overall	71.5% (62.1–79.3%)	87.0% (83.8–89.6%)	57	104	57.9%	92.4%	16.8 (12.1–23.3)	0.87
Recommended FibroTest cutoff used: [1]P=0.06; [2]P=0.001								*P=0.72
0.74	61.5%[1] (46.6–74.5%)	90.8%[2] (88.0–93.0%)	77	74	62.6%	90.4%	15.7* (8.6–28.8)	0.87
<0.74	79.9%[1] (71.7–86.2%)	83.5%[2] (79.6–86.7%)	40	132	54.5%	94.3%	17.9* (12.1–26.5)	0.88
HIV/HBV coinfection: [3]P=0.06; [4]P=0.26								*P=0.40
Absent	69.0%[3] (59.4–77.2%)	87.7%[4] (84.2–90.4%)	62	98	58.4%	91.9%	10.8* (6.5–18.1)	0.83
Present	90.6%[3] (41.6–99.2%)	83.1%[4] (74.6–89.1%)	19	135	57.3%	97.3%	11.0* (2.5–47.6)	0.84
Prevalence of cirrhosis (F4): [5]P=0.69; [6]P=0.65								*P=0.85
≤0.18	69.7%[5] (51.0–83.6%)	87.7%[6] (83.9–90.8%)	61	98	58.6%	92.1%	16.5* (3.2–12.4)	0.87
>0.18	73.7%[5] (63.8–81.6%)	86.4%[6] (81.1–90.4%)	53	109	57.5%	92.9%	17.8* (8.3–27.4)	0.88
Author independence: [7]P=0.30; [8]P=0.29								*P=0.79
No	69.8%[7] (59.9–78.1%)	87.7%[8] (84.4–90.3%)	60	98	58.7%	92.1%	16.4* (11.7–23.0)	0.87
Yes	80.5%[7] (54.6–93.4%)	82.3%[8] (71.0–89.8%)	39	142	53.2%	94.4%	19.1* (6.4–57.3)	0.88
Average size of liver biopsy (length): [9]P=0.57; [10]P=0.98								*P=0.49
<20 mm	68.5%[9] (50.8–82.1%)	87.2%[10] (82.9–90.5%)	63	102	57.2%	91.7%	14.8* (7.8–28.1)	0.86
≥20 mm	73.9%[9] (64.5–81.5%)	87.1%[10] (81.5–91.3%)	52	103	58.9%	93.0%	19.2* (13.3–27.7)	0.88
Liver biopsy and FibroTest sampling on same day: [11]P=0.13; [12]P=0.10; [13]P=0.73; [14]P=0.22								*P=0.04; $P=0.56
NA	55.7%[11] (38.4–71.7%)	89.1%[13] (82.3–93.5%)	89	87	56.1%	88.9%	9.1* (4.9–16.9)	0.82
No	84.1%[12] (68.2–92.9%)	84.2%[14] (77.9–89.0%)	32	126	57.1%	95.5%	28.3$ (12.2–65.8)	0.91
Yes	72.2%[11,12] (63.5–79.5%)	89.2%[13,14] (71.0–96.6%)	56	86	62.6%	92.8%	21.2*,$ (12.9–34.8)	0.89
METAVIR used for grading liver biopsy: [15]P=0.53; [16]P=0.11								*P=0.72
No	75.7%[15] (65.2–83.9%)	84.3%[16] (78.9–88.6%)	49	126	54.7%	93.3%	16.8* (11.3–25.0)	0.87
Yes	70.4%[15] (54.9–82.3%)	89.0%[16] (85.6–91.6%)	59	88	61.5%	92.3%	19.2* (10.3–35.7)	0.88

Sens, sensitivity; Spec, specificity; FN, false negative; FP, false positive; PPV, positive predictive value; NPV, negative predictive value; DOR, diagnostic odds ratio; AUC, area under the ROC curve.
95% confidence intervals of selected measures are presented in parentheses.

Even in the best possible conditions, in which LB accuracy is highest (sensitivity and specificity of biopsy are 90%) and the prevalence of significant disease is 40%, the AUC for a perfect liver fibrosis marker would be 0.90, which makes it hard to distinguish a perfect surrogate from a bad one (65). Owing to the markedly poor risk-to-benefit ratio of LB, it has been argued that it should not be recommended as a first-line procedure but rather as a second-line estimate of liver injury, reserved for cases involving complex disorders and discordance between clinical and noninvasive findings (19). LB is prone to sampling errors and intra- and interobserver variability, with the additional impact of the level of experience of the pathologist on overall accuracy and reliability (10,63,66). The size of liver sample may also be of importance (17,67), but even a 25-mm-long biopsy sample may have up to 25% rate of discordance for fibrosis staging (68).

In our analysis of liver-biopsy-related factors that may have an impact on accuracy, studies with an average length of LB ≥20 mm did not have better measurements of diagnostic accuracy for either significant fibrosis or cirrhosis, in comparison with those reporting an average length <20 mm. However, studies using histological score other than METAVIR had significantly better sensitivity and DOR for significant fibrosis, although there were no differences in measurements of diagnostic accuracy for liver cirrhosis (**Tables 2** and **3**). It is important to emphasize that all five studies that used a non-METAVIR score were conducted on patients of Asian origin; therefore, this may be another confounder. What may make this of even greater significance is the fact that three of the five studies used the Batts and Ludwig scoring system, which is similar to the METAVIR score.

One common problem in the studies included in the present review was the use of FT thresholds not recommended by developers, as more than half the included studies used nonrecommended FT thresholds for both fibrosis and cirrhosis. This is another possible source of bias and a strong source of heterogeneity, as selective reporting of the thresholds identified to optimize test accuracy can introduce bias if they are selected in a data-driven manner (69). This

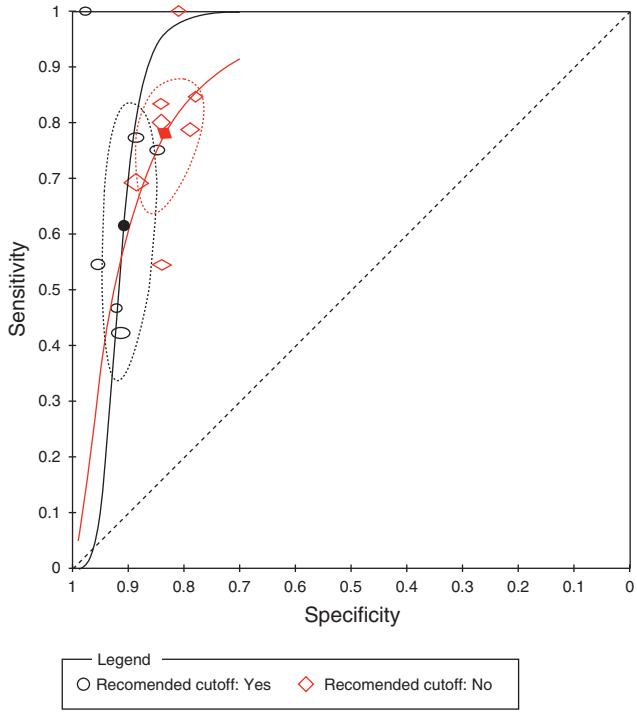

Figure 8. Summary receiver operating characteristic (ROC) plot of FibroTest/Fibrosure (FT) for detection of liver cirrhosis (METAVIR F4). The curves represent the summary ROC curves for studies that used the recommended FT cutoff of 0.74 and those that used other cutoffs. The marked point on the ROC curves represents the summary estimate of test performance and the area outline surrounding it represents the 95% confidence region of summary estimate.

was confirmed in a sensitivity analysis, in which we demonstrated a significantly higher pooled sensitivity and DOR for significant fibrosis in studies that used FT cutoff <0.48 (**Table 2**), suggesting better diagnostic accuracy in studies using FT cutoff in the range of 0.31–0.48. However, studies that used the recommended cutoff for cirrhosis had better pooled specificity in comparison with those that used FT threshold <0.74 (**Table 3**). FibroTest at the proposed cutoff of 0.74 had excellent performance for exclusion of liver cirrhosis.

The developers of the FT recommend the application of FT as a continuous rather than binary variable in order to maximize its effectiveness (15). However, in clinical practice, knowing the exact fibrosis stage is not as important as knowing whether the patient has mild or advanced liver disease (F0–F1 vs. F2–F4 and F0–F3 vs. F4) (11). This is of potentially crucial clinical interest in cases where the patient is to be evaluated for antiviral treatment, as the decision to start treatment and the choice of medications may be influenced by fibrosis stage. For example, differentiation between the F1 and F2 stage may be of utmost importance when deciding not to initiate long-term treatment with analogs in patients with the risk of developing resistance (62). Therefore, it may be a better strategy to concentrate on finding the best-performing FT thresholds for detection and exclusion of both significant fibrosis and cirrhosis.

Obviously, at least for a CHB population, there is a clear need for further refinement of proposed FT thresholds, especially for significant fibrosis, and their adaptation to clinical needs. This, however, may prove to be a hard task; Parkes and colleagues, in a review of 14 studies with 10 different tests, showed that cutoff levels with clinically relevant predictive values for the presence or absence of significant fibrosis were applicable to only 35% of the population of patients with chronic hepatitis C (63). Others have argued, however, that sensitivities and specificities above 85% can be considered as adequate for identifying patients with significant fibrosis, as there are no relevant clinical consequences of false positives or false negatives (32).

The prevalence of advanced fibrosis and cirrhosis may have an impact on overall accuracy because of a spectrum effect, and it has been observed that in cases where spectrum bias is present, either sensitivity or specificity would be expected to change (62). Therefore, one could expect higher sensitivity and specificity in populations in which extreme stages of fibrosis (F0 or F4) are present than in populations with higher prevalence of intermediate stages (F1–F3). This was recognized as an issue in studies dealing with accuracy of fibrosis markers; as mentioned above, Poynard *et al.* (30) proposed standardization of the reported AUCs according to the prevalence of fibrosis stages. We also investigated whether the prevalence of significant fibrosis and cirrhosis may have an impact on accuracy by evaluating studies in which reported prevalence of significant fibrosis and cirrhosis was above 54 and 18%, respectively. We have chosen these two thresholds as, again, Poynard *et al.* (30) reported natural prevalence of advanced fibrosis and cirrhosis in a cohort of 2,235 CHC patients to be 54 and 18%, respectively. As expected, both pooled sensitivity and specificity, and especially DOR, were better in studies with prevalence of significant fibrosis ≥54%, but without statistical significance.

Calculated values of prevalence of significant fibrosis and cirrhosis in a total of 2,494 patients with CHB included in this meta-analysis were 55.33 and 16.92% respectively—similar to the above-mentioned values. In addition, in the 2,394 patients from 15 studies with sufficient data, the value of DANA was 2.105, again similar to the DANA reported in the CHC patients (30). Therefore, the natural distribution of stages of fibrosis in CHB and CHC patients appears to be similar.

There was no difference between the performances of FibroTest on the basis of the independence of authors; however, although there was no statistical difference, pooled DOR and AUC were higher in studies conducted by independent authors. Although there were four studies in which we were not able to obtain data on time interval between LB and FT, studies with blood sampling for FT and LB performed on the same day had better pooled sensitivity, specificity, and DOR for significant fibrosis than those that reported a time difference between FT and LB. However, such difference was not observed in the analysis for liver cirrhosis. Inclusion of a special population of HIV-coinfected patients also did not have any impact on diagnostic accuracy, although there was a tendency for better sensitivity in studies with HIV-coinfected patients. However, owing to the small number of included studies,

this result should be considered with caution and leaves space for additional exploration in future studies.

There are several implications for future research. More studies on diagnostic accuracy for liver fibrosis are needed in populations of CHB patients, regardless of the noninvasive methods being explored. Future authors of studies exploring the performance of FT in CHB patients should be encouraged to insist on a rigorous design and methodology. Although reporting AUCs is usual, it is also important to report data for diagnostic accuracy at a recommended threshold(s) in order to allow for easier extraction of data for future meta-analyses (2×2 table). There is a need for studies exploring best-performing FT thresholds for both dichotomous outcomes, as well as studies in which the central research question explores the diagnostic accuracy of FT in Asian populations, but with the use of the MEATVIR scoring system. More studies in HIV–HBV-coinfected patients are also warranted. A future review in which diagnostic performance of FT is compared with other noninvasive markers could also be of tremendous use, as could a review of studies using a combination of two or more noninvasive markers.

Implications for practice deriving from our results suggest that FibroTest is of excellent utility for excluding cirrhosis in patients with CHB, but has suboptimal performance in detection of significant fibrosis and cirrhosis and in exclusion of significant fibrosis. It is of importance for clinicians to adhere to the recommended thresholds for dichotomous outcomes until better ones are derived and to sample blood for FT on the day of LB, as this obviously may have an impact on accuracy.

CONFLICT OF INTEREST

Guarantor of the article: Nermin N. Salkic, MD, PhD.

Specific author contributions: Nermin N. Salkic designed the research, performed initial searches, hand searched full-text papers, checked extracted data consistency, and performed statistical analysis; Predrag Jovanovic and Majda Brcic independently screened initial search results, selected papers of interest, performed methodological assessment, and extracted data; Goran Hauser evaluated the methodology of the search, screened statistical analysis for errors, and obtained full-text articles. All authors took part in writing the paper.

Financial support: None.

Potential competing interests: None.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Diagnostic accuracy of FibroTest/Fibrosure (FT) for liver fibrosis and cirrhosis in chronic hepatitis B (CHB) has been insufficiently evaluated.

WHAT IS NEW HERE

- ✓ FT is of excellent utility for excluding cirrhosis in patients with CHB.
- ✓ The diagnostic performance of FibroTest in detection of significant fibrosis and cirrhosis and in exclusion of significant fibrosis is suboptimal.
- ✓ It is important for clinicians to adhere to the recommended thresholds of FibroTest until better ones are derived.

REFERENCES

1. WHO fact sheet No 204: Hepatitis B. World Health Organization; July 2013. Available from http://www.who.int/mediacentre/factsheets/fs204/en/index.html.
2. Sebastiani G, Alberti A. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. World J Gastroenterol 2006;12:3682–94.
3. European Association for the Study of the Liver. EASL clinical practice guidelines: Management of chronic hepatitis B virus infection. J Hepatol 2012;57:167–85.
4. Lok AS, McMahon BJ. Chronic hepatitis B. Hepatology 2001;34:1225–41.
5. 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.
6. Terjung B, Lemnitzer I, Dumoulin FL *et al.* Bleeding complications after percutaneous liver biopsy. An analysis of risk factors. Digestion 2003;67:138–45.
7. West J, Card TR. Reduced mortality rates following elective percutaneous liver biopsies. Gastroenterology 2010;139:1230–7.
8. Brunetti E, Silini E, Pistorio A *et al.* Coarse vs. fine needle aspiration biopsy for the assessment of diffuse liver disease from hepatitis C virus-related chronic hepatitis. J Hepatol 2004;40:501–6.
9. Colloredo G, Guido M, Sonzogni A *et al.* Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. J Hepatol 2003;39:239–44.
10. Rousselet MC, Michalak S, Dupre F *et al.* Sources of variability in histological scoring of chronic viral hepatitis. Hepatology 2005;41:257–64.
11. Castéra L. Noninvasive methods to assess liver disease in patients with hepatitis B or C. YGAST 2012;142:1293–1302.e4.
12. The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. Hepatology 1994;20:15–20.
13. Pinzani M, Vizzutti F, Arena U *et al.* Technology Insight: noninvasive assessment of liver fibrosis by biochemical scores and elastography. Nat Clin Pract Gastroenterol Hepatol 2008;5:95–106.
14. 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.
15. 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.
16. Cales P, Veillon P, Konate A *et al.* Reproducibility of blood tests of liver fibrosis in clinical practice. Clin Biochem 2008;41:10–8.
17. Poynard T, Munteanu M, Imbert-Bismut F *et al.* Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. Clin Chem 2004;50:1344–55.
18. Rockey DC, Bissell DM. Noninvasive measures of liver fibrosis. Hepatology 2006;43:S113–20.
19. Halfon P, Munteanu M, Poynard T. FibroTest-ActiTest as a non-invasive marker of liver fibrosis. Gastroentérol Clin Biol 2008;32:22–39.
20. Poynard T, Lassailly G, Diaz E *et al.* Performance of biomarkers FibroTest, ActiTest, SteatoTest, and NashTest in patients with severe obesity: meta analysis of individual patient data. PloS One 2012;7:e30325.
21. Poynard T, Morra R, Halfon P *et al.* Meta-analyses of FibroTest diagnostic value in chronic liver disease. BMC Gastroenterol 2007;7:40.
22. Poynard T, Ngo Y, Munteanu M *et al.* Noninvasive markers of hepatic fibrosis in chronic hepatitis B. Curr Hepat Rep 2011;10:87–97.
23. Poynard T, Ngo Y, Perazzo H *et al.* Prognostic value of liver fibrosis biomarkers: a meta-analysis. Gastroenterol Hepatol 2011;7:445–54.
24. Shaheen AAM, Wan AF, Myers RP. FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy. Am J Gastroenterols 2007;102:2589–600.
25. Macaskill P, Gatsonis C, Deeks JJ *et al.* Chapter 10: Analysing and Presenting Results. In Deeks JJ, Bossuyt PM, Gatsonis C (eds). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0 The Cochrane Collaboration 2010.
26. Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: Deeks JJ, Bossuyt PM, Gatsonis C (eds). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]: The Cochrane Collaboration, 2008.

27. Whiting PF, Rutjes AW, Westwood ME; *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
28. Reitsma JB, Rutjes AWS, Whiting P *et al.* Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (eds). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0: The Cochrane Collaboration, 2009.
29. Whiting P. The development of QUADAS-2: A revised tool for the quality assessment of diagnostic test accuracy studies 2007. Available from http://www.bris.ac.uk/quadas/resources/.
30. Poynard T, Halfon P, Castéra 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.
31. Takwoingi Y, Deeks JJ. MetaDAS: a SAS macro for meta-analysis of diagnostic accuracy studies. User Guide Version 1.3 2010 Available from: http://srdta.cochrane.org/software-development.
32. Martinez SM, Crespo G, Navasa M *et al.* Noninvasive assessment of liver fibrosis. Hepatology 2011;53:325–35.
33. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C* The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;349:825–32.
34. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.
35. Anastasiou J, Alisa A, Virtue S *et al.* Noninvasive markers of fibrosis and inflammation in clinical practice: prospective comparison with liver biopsy. Eur J Gastroenterol Hepatol 2010;22:474–80.
36. Boursier J, Vergniol J, Sawadogo A *et al.* The combination of a blood test and Fibroscan improves the non-invasive diagnosis of liver fibrosis. Liver Int 2009;29:1507–15.
37. Chen J, Liu C, Chen H *et al.* Study on noninvasive laboratory tests for fibrosis in chronic HBV infection and their evaluation. J Clin Lab Anal 2013;27:5–11.
38. de Lédinghen V, Vergniol J, Barthe C *et al.* Non-invasive tests for fibrosis and liver stiffness predict 5-year survival of patients chronically infected with hepatitis B virus. Alim Pharmacol Ther 2013;37:979–88.
39. Degos F, Perez P, Roche B *et al.* Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study). J Hepatol 2010;53:1013–21.
40. Dolmazashvili E, Zhamutashvili M, Svanidze M *et al.* Fibroscan and FibroTest/FibroMax to assess liver fibrosis/cirrhosis in patients with chronic HBV and HCV infection in Georgia. Georgian Med News 2008;165:83–7.
41. Friedrich-Rust M, Rosenberg W, Parkes J *et al.* Comparison of ELF, FibroTest and FibroScan for the non-invasive assessment of liver fibrosis. BMC Gastroenterol 2010;10:103.
42. Friedrich-Rust M, Schwarz A, Ong M *et al.* Real-time tissue elastography versus FibroScan for noninvasive assessment of liver fibrosis in chronic liver disease. Ultraschall Med 2009;30:478–84.
43. Mallet V, Dhalluin-Venier V, Roussin C *et al.* The accuracy of the FIB-4 index for the diagnosis of mild fibrosis in chronic hepatitis B. Alim Pharmacol Ther 2009;29:409–15.
44. Uyar C, Akcam FZ, Ciris M *et al.* Comparison of FibroTest-ActiTest with histopathology in demonstrating fibrosis and necroinflammatory activity in chronic hepatitis B and C. Indian J Pathol Microbiol 2010;53:470–5.
45. Cassinotto C, Lapuyade B, Aït-Ali A *et al.* Liver fibrosis: noninvasive assessment with acoustic radiation force impulse elastography–comparison with FibroScan M and XL probes and FibroTest in patients with chronic liver disease. Radiology 2013;269:283–92.
46. Myers RP, Tainturier M-H, Ratziu V *et al.* Prediction of liver histological lesions with biochemical markers in patients with chronic hepatitis B. J Hepatol 2003;39:222–30.
47. Zhao L, Xu D, Lu Z *et al.* Validation of Fibrotest to diagnose liver fibrosis to patients with chronic hepatitis B. Chinese J Pract Int Med 2007;27:1274–7.
48. Sebastiani G, Vario A, Guido M *et al.* 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.
49. Gui H-L, Xie Q, Wang H. 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.
50. 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.
51. 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.
52. Bonnard P, Sombié R, Lescure F-X *et al.* Comparison of elastography, serum marker scores, and histology for the assessment of liver fibrosis in hepatitis B virus (HBV)-infected patients in Burkina Faso. Am J Trop Med Hyg 2010;82:454–8.
53. Castéra L, Bernard P-H, Le Bail B *et al.* Transient elastography and biomarkers for liver fibrosis assessment and follow-up of inactive hepatitis B carriers. Alim Pharmacol Ther 2011;33:455–65.
54. Mbaye PS, Sarr A, Sire J-M *et al.* Liver stiffness measurement and biochemical markers in senegalese chronic Hepatitis B patients with normal ALT and high viral load. PloS one 2011;6:e22291.
55. 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.
56. Stibbe KJM, Verveer C, Francke J *et al.* Comparison of non-invasive assessment to diagnose liver fibrosis in chronic hepatitis B and C patients. Scand J Gastroenterol 2011;46:962–72.
57. Raftopoulos SC, George J, Bourliere M *et al.* Comparison of noninvasive models of fibrosis in chronic hepatitis B. Hepatol Int 2011;6:457–67.
58. Sebastiani G, Castéra L, Halfon P *et al.* The impact of liver disease aetiology and the stages of hepatic fibrosis on the performance of non-invasive fibrosis biomarkers: an international study of 2411 cases. Alim Pharmacol Ther 2011;34:1202–16.
59. Kim BK, Kim HS, Park JY *et al.* Prospective validation of ELF test in comparison with Fibroscan and FibroTest to predict liver fibrosis in Asian subjects with chronic hepatitis B. PloS One 2012;7:e41964.
60. Kim BK, Kim SU, Kim HS *et al.* Prospective validation of FibroTest in comparison with liver stiffness for predicting liver fibrosis in Asian subjects with chronic hepatitis B. PloS One 2012;7:e35825.
61. Park MS, Kim BK, Cheong JY *et al.* Discordance between liver biopsy and FibroTest in assessing liver fibrosis in chronic hepatitis B. PloS One 2013;8:e55759.
62. Dinnes J, Deeks J, Kirby J *et al.* A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess 2005;9:1–113, iii.
63. Regev A, Berho M, Jeffers LJ *et al.* Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol 2002;97:2614–8.
64. Poynard T, Munteanu M, Deckmyn O *et al.* Validation of liver fibrosis biomarker (FibroTest) for assessing liver fibrosis progression: proof of concept and first application in a large population. J Hepatol 2012;100:541–8.
65. Mehta SH, Lau B, Afdhal NH *et al.* Exceeding the limits of liver histology markers. J Hepatol 2009;50:36–41.
66. Maharaj B, Maharaj RJ, Leary WP *et al.* Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. Lancet 1986;1:523–5.
67. Halfon P, Bacq Y, De Muret A *et al.* Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C. J Hepatol 2007;46:395–402.
68. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;38:1449–57.
69. Leeflang MM, Moons KG, Reitsma JB *et al.* Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 2008;54:729–37.