Hepat Mon. 2016 February; 16(2): e31983. doi: 10.5812/hepatmon.31983

Published online 2016 February 16. **Research Article**

Diagnostic Estimation of Noninvasive Tests for Hepatic Fibrosis in Chronic Hepatitis B Patients Without a Gold Standard

Yi Xiang Zheng,[1] Shu Juan Ma,[2] and Meng Hou Lu[1,*]

[1]Viral Hepatitis Key Laboratory of Hunan Province, Department of Infectious Disease, Xiangya Hospital, Central South University, Changsha, China
[2]Department of Epidemiology and Health Statistics, School of Public Health, Central South University, Changsha, China

*Corresponding Author: Meng Hou Lu, Viral Hepatitis Key Laboratory of Hunan Province, Department of Infectious Disease, Xiangya Hospital, Central South University, Changsha, China. E-mail: lumenghou@126.com

Received 2015 July 30; **Revised** 2015 October 19; **Accepted** 2016 January 6.

Abstract

Background: Assessment of hepatic fibrosis stage in patients with chronic hepatitis B (CHB) is indispensable for prognosis evaluation and therapeutic regime. Noninvasive tests are fast, safe and cheap and need low technical requirements for diagnosing hepatic fibrosis in CHB patients.

Objectives: Using the latent class model with a random-factor to estimate relative accuracy of noninvasive tests for the diagnosis of hepatic fibrosis without a gold standard in a large population with CHB.

Patients and Methods: A total of 544 patients with CHB were assessed for fibrosis stage by four noninvasive tests containing liver stiffness measurement (LSM), aspartate aminotransferase-to-platelet ratio index (APRI), fibrosis index based on 4 factors (FIB-4) and globulin and platelet (GP). The diagnostic evaluation was made by the latent class method with random effect which analyzed the clinical data above to assess the accuracy of four ways of noninvasive diagnosis.

Results: The latent class model with random effect permitted to conciliate the observed data and estimates of test performances. For significant fibrosis, the specificity/sensitivity were 83.24%/91.59% (APRI), 90.05%/95.57% (FIB-4), 75.11%/66.01% (LSM) and 71.13%/98.33% (GP), respectively. For cirrhosis, the specificity/sensitivity were 84.04%/17.91% (APRI), 89.86%/17.09 (FIB-4), 78.64%/37.07% (LSM) and 82.28%/37.07% (GP), respectively.

Conclusions: FIB-4 confirmed the best value for diagnosis of significant fibrosis. APRI had a sub-optimal diagnosis accuracy for significant fibrosis. LSM showed the most balance diagnosis value for cirrhosis with the highest sensitivity and moderate specificity.

Keywords: Hepatitis B, Liver Cirrhosis, Diagnosis, Latent Class Model

1. Background

Hepatitis B virus (HBV) infection is the most common chronic viral infection in the world (1) and roughly 30% of the world's population show serological evidence of current or past infection. About 350 million people are chronic carriers of HBV and 30% - 40% of them develop end-stage liver diseases or liver cancer (2). Hepatitis B is the prevailing cause of cirrhosis in developing countries, especially most areas of Asia and sub-Saharan Africa (3). Most chronic hepatitis B (CHB) cases are notoriously asymptomatic until occurrence of decompensated cirrhosis. Cirrhosis might lead to mortality without transplantation in as high as 85% patients over 5 years (3). Cirrhosis in CHB would aggravate the prognosis and treatment burden. Indeed, the incidence of hepatocellular carcinoma (HCC) per 100 person years ranges from 2 to 3.7 in patients with cirrhosis compared with only 0.3 to 0.6 in patients with active hepatitis B without cirrhosis and 0.02 to 0.2 in asymptomatic carriers (4). Cirrhosis usually leads to various complications which influence the quality of CHB patients' life severely, including portal hypertension, ascites, hepatic, encephalopathy and active esophageal variceal hemorrhage. The increasing burden of liver diseases and problems of late presentation with decompensation emphasize the need for mass screening to identify patients with chronic liver diseases (5). Therefore, early diagnosis of fibrosis and cirrhosis is important, especially for CHB patients who are in asymptomatic phase. Furthermore, CHB patients with significant fibrosis and cirrhosis are recommended to receive antiviral therapy in current guidelines (6). Surveillance of hepatic fibrosis is necessary for long-term treatment.

Liver biopsy is a traditional way to evaluate hepatic fibrosis, which is considered as the reference method. However, there is a laundry list of limitations of biopsy as follows; high cost, invasiveness, risk of complications, need for expert histological interpretation, risk of occupational exposure to handlers and so on. Moreover, liver biopsy cannot provide a dynamic monitor of progression of hepatic fibrogenesis. All these limitations directly affect extensive applications of biopsy in liver fibrosis screening for CHB patients.

More substitutions are applied to overcome limitations, like noninvasive tests, which are widely used in the current clinical practice. The latest world health organization (WHO) HBV guideline strongly recommends noninvasive assessment of liver disease stage at baseline and during follow-up (7). Aspartate aminotransferase-to-platelet ratio index (APRI), fibrosis index based on the 4 factors (FIB-4) and liver stiffness measurement (LSM) are the most common and convenient noninvasive tests, which are prospective for extensive application in resource-limited settings. To standardize noninvasive tests and ascertain their utility, evaluation of these diagnostic methods is essential. Assessments in previous researches were based on biopsy analysis as the reference standard. However, liver biopsy analysis is an imperfect reference standard; errors in liver biopsy results make it impossible to distinguish a perfect surrogate from ones that are now judged as clinically unacceptable (8).

With imperfect side most time, especially in the underdeveloped areas, these estimations of accuracy (sensitivity, specificity and AUROC) of the diagnostic test would be false and unreliable. Liver biopsy would not be a gold standard. An alternative method is needed to assess the accuracy of noninvasive tests. When the result of a reference (or gold standard) test is missing or not error-free, the accuracy of diagnostic test is often assessed through latent class models with two latent classes, representing diseased or non-diseased status (9). In the situation without a gold standard, latent class model (LCM) analysis has been recommended for diagnostic evaluation of several tests (10).

Outcome of interest cannot be measured directly in many research situations. These unobservable outcomes of APRI, FIB-4 and LSM for diagnosis of fibrosis/cirrhosis can be measured indirectly by eliciting responses that are related to the construct of interest, which named latent variables.

2. Objectives

The aim of this study was to apply this methodology to estimate relative accuracy of APRI, FIB-4 and LSM for the diagnosis of fibrosis, without a gold standard in a large group of CHB patients. We evaluated noninvasive tests recommended in the WHO HBV guideline, which would be beneficial and feasible to their application on a large scale, especially in resource limited settings.

3. Patients and Methods

3.1. Patients

Patients were recruited into trails from the clinic of Xiangya hospital affiliated to Central South University (Changsha,. China) from March 1, 2013 to April 1, 2014. The inclusion criteria were 1) CHB defined by HBsAg positivity for more than 6 months; 2) ALT (glutamic-pyruvic transaminase) and AST (glutamic oxalacetic transaminase) values within the normal range (our laboratory reference value was 40 U/L) or < 2 × ULN (upper limit of normal); 3) absence of liver comorbidity including hepatitis delta superinfection, HCV co-infection, chronic ethanol consumption (210 g/week in men and 140 g/week in women), Wilson's disease, HIV co-infection or auto-immune hepatitis; 5) no pregnancy. Patients received comprehensive clinical and laboratory examinations within one day after FibroScan test. The basic information included age and gender, while clinical data included ALT, AST, albumin, globulin, total bilirubin (TBIL) and platelet count (PLT). The characteristics of the study population are listed in Table 1.

3.2. Liver Stiffness Measurement

All patients underwent FibroScan (Echosens^{TM}, Paris, France) for LSM within 24 hours after the blood biochemical test. Trained operators blinded to study design performed LSM. For LSM reliability, recommended criteria were at least ten valid LSM values acquired from each patient, with a success rate greater than 60%, and the interquartile range/median LSM no more than 30% of the corresponding LSM value (11). The following cut-offs recommended by WHO HBV guideline were used to estimate the presumed fibrosis stages: LSM > 12.5 kPa for cirrhosis (METAVIR F4); LSM > 7.75 kPa for significant fibrosis (METAVIR ≥ F2) (7).

3.3. Calculation of APRI, FIB-4 and GP

APRI, FIB-4 and GP are biochemical models for diagnosis of hepatic fibrosis calculated in formulas as follows: APRI = (AST/upper limit of normal) × 100/PLT count (10^9/L) (12). The following cut-offs recommended by WHO HBV guideline for fibrosis stages: APRI cut-off points were chosen to predict significant fibrosis and cirrhosis: APRI > 1 for cirrhosis (METAVIR F4), APRI > 0.5 for significant fibrosis (METAVIR ≥ F2) (7). FIB-4 = [age (year) × AST (IU/L)]/[PLT count (10^9/L) × ALT (IU/L)$^{1/2}$] (13). According to the WHO HBV guideline, FIB-4 has been developed and validated for detection of fibrosis stages ≥ F3 and not for cirrhosis. However, the guideline has suggested establishing and validating FIB-4 cut-offs for the diagnosis of cirrhosis and advanced fibrosis (7). We summarized the meta-analysis and clinical researches to assume two cut-offs; FIB-4 > 1.45 for significant fibrosis (METAVIR ≥ F2), FIB-4 > 3.6 for cirrhosis (METAVIR F4) (14).

We estimated noninvasive tests through LCM-R; however, the number of tests for evaluation should be more than four (9). Therefore, we had to introduce one more test method. Globulin and PLT have shown association with liver fibrosis and cirrhosis. Xu-Dong Liu et al. used two markers (globulin and PLT) to develop a new fibrosis model named GP model in HBV infection for predicting cirrhosis and fibrosis. GP model = GLOB (g/dL) × 100/PLT (× 10^9/L). The cut-offs used to predict significant fibrosis and cirrhosis were listed below; GP > 1.68 for significant fibrosis (METAVIR ≥ F2) and GP > 2.53 for cirrhosis (METAVIR F4) (15).

3.4. Design and Modeling

We used models to estimate performances of four noninvasive tests without a gold standard. The latent class model using the standard maximum likelihood method to combine the test results from each patient constructed the reference standard (9, 10, 16). Two authors (YX Z and SJ M) independently rechecked all tests (APRI, FIB-4, GP and LSM) used in every patient, with each test producing a dichotomous test result (e.g. test was either positive or negative). Usually, we used test parameters of prevalence, sensitivity and specificity to assess their functions, which were unknown in this study. In every evaluation, 16 distributions of subjects according to the four test results would be generated through modeling. Hence, each patient had a result combining the above tests, then we calculated the likelihood of each observed combination and the numbers of subjects for combination. Standard maximum likelihood methods could be used to obtain a unique solution (10, 17).

3.5. Latent Class Analysis

The precondition of this method acknowledged that no gold standard existed and all the available tests were related to unknown true status; liver fibrosis/cirrhosis present or absent. These uncertain outcomes were defined as latent class. This study contained many variations, but all tests had in common the use of a statistical model to combine different pieces of information (test results) from each patient, to construct a reference standard.

The four noninvasive tests in this study had a common dependency that they were initially validated by biopsy. The traditional two-class model might not fit data either seen as an artifact of measurement instrument or as a result of within-class heterogeneity. The latent class model with a random-factor, the LCM-R model was used to set the analysis, which allowed for local dependencies and within-class heterogeneity (9, 10, 16, 17). In this study, we assumed that the result of each test was governed by two mechanisms or factors; the true status of liver fibrosis and individual biological process or the diagnostic test technological characteristics to fit the LCM-R model.

3.6. Sources of Impairment to Fit Model

Dependency or heterogeneity of test might significantly impair the fit of standard LCM without random effect. We used bivariate residuals of baseline latent class analysis to find sources of fit impairment. The pair of tests was excluded step by step until a model fitting the observed results was obtained. When the likelihood-ratio goodness-of-fit value L^2 (likelihood squared) significance was more than 0.05, it was identified as fitting the observed results (10, 16-18).

3.7. Sensitivity Analysis

To assess possible variability due to the cut-off of tests and stability of fit model, we made a sensitivity analysis with higher cut-off of APRI for significant fibrosis (1.5) and cirrhosis (2.0) (7). The sensitivity of the fit model was judged by comparing the variations of results in the fit model analyses with different cut-offs of APRI.

3.8. Statistical Analysis

Characteristics of included patients were present as means, 95% CIs (95% confidence intervals), counts, median and IQR. Data input, calculation and basic analyses were made using IBM SPSS 22.0 (Chicago, IL, USA), while LatentGold-4.5 software (Statistical Innovation, Belmont, MA, USA) was used to estimate the model parameters. The random effect analysis used 2 clusters option and "continuous factor" (CFactor) option (One CFactor) of the LatentGold-4.5 software. To confirm the best model for analysis, P value of the likelihood squared (L^2) had to be greater than 0.05 and the Bayesian information criterion (BIC), defined as $L^2 - \log(N) \times df$ (degrees of freedom of the data), had to be the smallest among all competing models (18).

4. Results

4.1. Participant Information

584 CHB patients enrolled in the study at beginning. Nine patients were HCV antibody positive and ALT or AST values of 21 patients were more than 2 ULN. Thirty patients were excluded because of unsatisfying criteria at last. A total of 554 patients fulfilled the inclusive criteria. Clinical and demographic characteristics of the study population are shown in Table 1.

4.2. Assessment of Test Performances by LCM-R Model

Performances of tests were assessed using LCM-R model without a gold standard. Sixteen possible combinations of the four test results based on LCM-R model for presuming significant fibrosis and cirrhosis are shown in Tables 2 and 3. A total of 292 (52.71%) subjects (148 all negatives and 144 all positives) were complete concordance within the four tests for diagnosis of significant fibrosis and 401 (72.38%) subjects (358 all negatives and 43 all positives) for diagnosis of cirrhosis. Fit of the models using LCM-R for observed distribution of test results are interpreted in Table 4. For significant fibrosis, specificity values were respectively FIB-4 (90.05%), APRI (83.23%), LSM (75.11%) and GP (71.13%). Sensitivity values were respectively GP (98.33%), FIB-4 (95.57%), APRI (91.59%) and LSM (66.01%). For cirrhosis, specificity values were FIB-4 (87%), APRI (82.71%), LSM (75.63%) and GP (71.07%). All the sensitivities were low with the following ranking; GP (23.9%), APRI (12.87%), FIB-4 (10.62%) and LSM (4.82%). The overall assessments by the Youden index were FIB-4 (0.8562), APRI (0.7483), GP (0.6946) and LSM (0.4112) for significant fibrosis diagnoses, respectively. All of which were unsatisfied for cirrhosis diagnoses as much small as the Youden index (Table 4).

4.3. Assessment of Significant Sources of Impairment in Modeling

LCM model for significant fibrosis became fit (P > 0.05), until the bivariate residuals of LSM-APRI and LSM-FIB-4 were excluded, while that was LSM-GP in LCM model for cirrhosis. Bivariate residual of LSM-APRI and LSM-FIB-4 significantly impaired the model fit for significant fibrosis, which identified as the main sources of impairment in LCM models. The main source of impairment in LCM model for cirrhosis came from LSM-GP, whose bivariate residual was excluded to reach model fit. The bivariate residuals of other pairs were lower and the direct effect of them would not impair the fit of model for significant fibrosis and cirrhosis (Table 5).

4.4. Assessment of Sensitivity of the Fit Model

Generally, there were subtle differences among test cut-offs in different studies and populations. APRI also had another higher cut-off; APRI > 2 for cirrhosis (METAVIR F4) and APRI > 1.5 for significant fibrosis (METAVIR ≥ F2). The LCM-R model still fitted for distributions with higher cut-off for APRI, which increased the specificity of APRI for significant fibrosis and cirrhosis from 83.23% to 99.98% and 84.04% to 98.28%, respectively. Nevertheless, sensitivity decreased from 91.59% to 25.84% and 17.91% to 12.41%, respectively. The changes to other tests were little or no.

Table 1. Clinical, Biochemical and Basic Data of Included Patients

Features	Mean (95% CI)	Median (IQR)
Gender	554	
Male	419	
Female	135	
Age, y	39.6 (38.8 - 40.5)	40.0 (13.0)
ALB, g/L	47.4 (47.1 - 47.7)	47.0 (4.1)
GLOB, g/L	28.4 (28.1 - 28.7)	28.0 (5.1)
ALT, U/L	34.4 (33.5 - 36.4)	31.4 (22.0)
AST, U/L	34.9 (33.4 - 36.35)	28.8 (14.8)
T.Bil, umol/L	15.6 (15.0 - 16.3)	13.8 (8.6)
PLT $\times 10^9$/L	154.5 (149.1 - 160.0)	152.0 (86.0)
LSM, kPa	9.5 (8.9 - 10.1)	6.9 (5.1)

Abbreviations: ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; GLOB, globulin; IQR, inter quartile range; LSM, liver stiffness; PLT, platelet count; T.Bil, total bilirubin.

Table 2. Distribution of 544 Subjects According to the 16 Possible Combinations of the Four Test Results[a]

16 Possible Combinations	Type of Combination				Number of Subjects Observed	Expected by the Latent Class Model
	APRI	FIB-4	GP	LSM		
1	0	0	0	0	148	148.1
2	0	0	0	1	53	53.7
3	0	0	1	0	33	32.7
4	0	0	1	1	17	16.0
5	0	1	0	0	7	9.2
6	0	1	0	1	17	13.4
7	0	1	1	0	8	5.7
8	0	1	1	1	7	10.8
9	1	0	0	0	20	18.7
10	1	0	0	1	13	12.8
11	1	0	1	0	10	10.9
12	1	0	1	1	9	9.8
13	1	1	0	0	5	3.8
14	1	1	0	1	59	62.2
15	1	1	1	0	4	5.8
16	1	1	1	1	144	140.3

Abbreviations: APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis index based on the 4 factors; GP, globulin and platelet; LSM, liver stiffness measurement.
[a]Presumed significant fibrosis (present = 1) or not (absent = 0).

Table 3. Distribution of 544 Subjects According to the 16 Possible Combinations of the Four Test Results[a]

16 Possible Combinations	Type of Combination				Number of Subjects Observed	Expected by the Latent Class Model
	APRI	FIB-4	GP	LSM		
1	0	0	0	0	358	356.5
2	0	0	0	1	50	50.2
3	0	0	1	0	33	32.4
4	0	0	1	1	14	14.7
5	0	1	0	0	1	2.0
6	0	1	0	1	4	3.2
7	0	1	1	0	1	0.6
8	0	1	1	1	1	1.8
9	1	0	0	0	4	4.0
10	1	0	0	1	11	11.2
11	1	0	1	0	1	1.4
12	1	0	1	1	12	12.3
13	1	1	0	0	1	0.5
14	1	1	0	1	20	21.0
15	1	1	1	0	0	0
16	1	1	1	1	43	41.4

APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis index based on the 4 factors; GP, globulin and platelet; LSM, liver stiffness measurement.
[a]Presumed cirrhosis (present = 1) or not (absent = 0).

Table 4. Best Latent Class Model With Random Effect of Fibrosis Estimate Performances (n = 544)

Performance of Test	Specificity[a], %	Sensitivity[a], %	Youden index[b]	+LR[c]	−LR[d]
Best model for significant fibrosis[e]					
APRI	83.24	91.59	0.7483	5.46	0.1
FIB-4	90.05	95.57	0.8562	9.61	0.05
GP	71.13	98.33	0.6946	3.41	0.02
LSM	75.11	66.01	0.4112	2.65	0.45
Best model for cirrhosis[f]					
APRI	84.04	17.91	0.0195	1.12	0.98
FIB-4	89.86	17.09	0.0695	1.69	0.92
GP	82.28	21.03	0.0331	1.19	0.96
LSM	78.64	37.07	0.1571	1.74	0.8

Abbreviations: APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis index based on the 4 factors; GP, globulin and platelet; LSM, liver stiffness measurement; +LR, positive likelihood ratio; −LR, negative likelihood ratio.
[a]No confidence interval for the LCM-derived sensitivity and specificity estimates because these estimates were calculated from combinations of conditional probabilities, which had individual maximum-likelihood estimated standard errors.
[b]Youden index = sensitivity + specificity - 1.
[c]+LR = sensitivity/(1 - specificity).
[d]−LR = (1 - sensitivity)/specificity.
[e]L-Squared (standard error calculated using bootstrap): 5.35 (0.016); goodness of fit likelihood ratio test statistics: P value = 0.069, model fit when P > 0.05; bayesian information criterion: -7.28.
[f]L-Squared (standard error calculated using bootstrap): 3.27 (0.022); goodness of fit likelihood ratio test statistics: P value = 0.19, model fit when P > 0.05; bayesian information criterion: -9.36.

Table 5. Direct Effects of Pairs of Variables That Impaired the Fit of the Baseline Latent Class Model[a]

	Bivariate Residuals			Model Improvement After Excluding Residuals		
	FIB-4	LSM	GP	Pair Excluded	Fit (Cumulative)	P value[b] After Pair Exclusion
Significant fibrosis				None	17.4	0.0079 (no fit)
APRI	0.0122	1.1464	0.0082	LSM-APRI	12.43	0.029 (no fit)
FIB-4		0.3159	0.8298	LSM-FIB-4	7.26	0.12
LSM	NA	NA	0.0002	NA	NA	NA
Cirrhosis	NA	NA	NA	None	NA	NA
APRI	0.4453	0.1421	0.0763	LSM-GP	15.31	0.018 (no fit)
FIB-4	NA	0.2703	0.0717	NA	8.53	0.13
LSM	NA	NA	1.737	NA	NA	NA

Abbreviations: APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis index based on the 4 factors; GP, globulin and platelet; LSM, liver stiffness measurement; NA, not available.
[a]Effects were estimated by bivariate residuals of the baseline latent class analysis without random effects. The effect of the most significant pair was excluded to achieve non-significance.
[b]Model fit when $P > 0.05$.

5. Discussion

In this study, we used a novel and appropriate method to estimate sensitivity and specificity of noninvasive fibrosis tests. When a gold standard is absent, latent class models are often used where the unknown gold standard test is treated as a latent variable (19). LCM-R model showed well compatible application value for diagnosis. Since the reference standard for liver fibrosis stages was deficient and the limitation of liver biopsy fibrosis restricted its use, the model without using reference for estimating was compatible with distribution of above four noninvasive fibrosis tests. This work for the first time evaluated diagnostic value of the four common noninvasive fibrosis tests for HBV-relative liver fibrosis stages without a gold standard.

Many studies have been published on predicting significant fibrosis and cirrhosis among CHB patients in the past few years (20). The biochemistry markers model and physical detection methods are the two main ways of noninvasive tests. Salkic et al. study suggested that algorithm based on routine laboratory tests was an usable, applicable and accurate tool for diagnosis of CHB related fibrosis and cirrhosis, which was suitable for resource-limited settings where more expensive modalities were unavailable (21). Noninvasive fibrosis tests have been considered as the appropriate substitution to overcome limitations of liver biopsy to assess liver-fibrosis. All noninvasive fibrosis tests were firstly confirmed by the liver biopsy as the reference standard. However, the limitations of sample error and subjective bias in pathological diagnosis made liver biopsy not perfect enough as a gold standard to assess other tests. The sample of liver biopsy is only a small part of the whole liver, which might not be representative for the severity of hepatic fibrosis and lead to underdiagnosis of cirrhosis with sampling error (22). Although increasing the length of liver biopsy and using 16 gauge needle to ensure enough caliber of biopsy specimens could reduce the risk of sampling error, sampling variability still cannot be completely avoided (22, 23). Liver biopsy could not serve as the gold standard without strict conditions. The latent class model with random effects took a full consideration to the random variability factor in the model. All the tests (APRI, FIB-4, GP and LSM) were initially validated using biopsy and therefore it was rational to use a method that considered this non-independence among tests. Moreover, the latent class without random effects could not fit tests results distribution after estimating, which further testified the fit of LCM-R from the other side.

In estimation of LCM-R for noninvasive fibrosis tests, FIB-4 showed the best performance for diagnosis of significant fibrosis with high specificity and sensitivity (> 90%). The comprehensive performance was assessed by the Youden index, with higher value representing higher quality. Therefore, FIB-4 showed the best value for diagnosis of significant fibrosis. Although FIB-4 was initially applied to predict significant fibrosis in patients with HIV/HCV coinfection, its usage has been expanded to CHB patients (24). In previous studies, FIB-4 was only recommended for diagnosis of mild liver fibrosis, but the cut-off of FIB-4 for cirrhosis was still controversial (25, 26). In this study, we found that FIB-4 indeed had wonderful diagnostic value for significant fibrosis in CHB patients, while its ability to detect cirrhosis was deficient. Considering past studies and meta-analysis, we supposed the cut-off for cirrhosis as 3.6 and corresponding specificity was nearly 90% for cirrhosis; however, the sensitivity was less than 20%. Beyond that, hepatic fibrosis might be a risk factor for HCC. A research from Korea supported that high FIB-4 was a highly predictive risk factor of HCC incidence in CHB carriers (27), which was consistent with its high value for assessing hepatic fibrosis in our study.

APRI and GP both had well performances for diagnosis of significant fibrosis (sensitivity > 90%, specificity > 70%). GP was a new biochemistry marker model for HBV-relative liver fibrosis test, which had the sensitivity and specificity of 72.4% and 69.6% for minimal fibrosis, 72.7% and 84.5% for cirrhosis in the first report (15). However, more other rigorous clinic studies of GP were unavailable for further confirmation. One side, we selected this innovative method for revaluation; for another, it was regarded as a matched group for other tests estimation. Similar to FIB-4, APRI was a widely used test and the calculating parameters were the most common. Accuracy of the two tests for significant fibrosis in CHB patients had been compared in a meta-analysis; sensitivity and specificity values of FIB-4 were 65.4% and 73.6%, while those of APRI were 70.0% and 60.0%, respectively (28). Despite comparison in our evaluation was a little different from that meta-analysis, FIB-4 and APRI were both recommended for diagnosis of significant fibrosis with moderate accuracies in CHB patients.

LSM showed an unsatisfactory performance with lower specificity (75.11%) and sensitivity (66.01%) for diagnosis of significant fibrosis, compared with above tests. For diagnosis of cirrhosis, performance of all tests weakened, especially sensitivity (< 40%), while specificity was relatively high (< 75%). The Youden index of all tests was too small to indicate their suitable value for cirrhosis diagnoses. Even so, LSM had the most balanced diagnosis for cirrhosis with the highest sensitivity (37.03%), well specificity (78.64%) and biggest Youden index. LSM was also first suggested for predicting hepatic fibrosis in patients with HCV (29). Subsequent studies had confirmed it to be reliable for detection of significant fibrosis or cirrhosis in HBV patients and cut-off values were only slightly different from those observed in HCV patients (30, 31). Consistent with existing research conclusions, our estimation also suggested that diagnostic accuracy of LSM was relatively high for cirrhosis, but relatively poor for significant fibrosis (32).

Compared with performances of tests in previous studies and meta-analyses with biopsy as the gold standard, the APRI, FIB-4, GP and LSM showed better performances for diagnosis of significant fibrosis, while less value for diagnosis of cirrhosis by LCM-R (latent class model with a random-factor) model (14, 15, 20, 24, 26, 33). Models using LCM without random effects for significant fibrosis and cirrhosis did not fit the observed distribution (P value of L^2 was less than 0.05, Table 5), which suggested a random effect due to dependency among tests (as expected due to previous validation of APRI, FIB-4, GP and LSM by biopsy). In the LCM-R model assessment, relative performances of APRI, FIB-4, GP and LSM would be helpful in the absence of a gold standard.

5.1. Impaired Sources of Major Variability Among Tests

To identify the strength of LCM for estimation, we considered random effect of initial dependency among noninvasive fibrosis tests and discovered their paired residual. As previously estimation mentioned above, bivariate residuals of LSM-APRI and LSM-FIB-4 were the impaired source in modeling for significant fibrosis. The rational explanation might be that necrosis and inflammation increase LSM independent of fibrosis stages (34) and ALT increases LSM linearly in chronic hepatitis B patients at any fibrosis stage (35). In spite of exclusion of patients with obviously increased aminotransferase in this study, we still needed to think over recessive necrosis, inflammation and steatosis in liver. The LSM-GP pair was the most important residual for diagnosis of cirrhosis. PLT was usually significantly influenced in the later period of cirrhosis with hypersplenism, which would make GP disable to distinguish earlier cirrhosis.

5.2. Limitations

As a diagnosis-evaluation, this study could not give the AUROC of each test, because LCM-R could only give an estimation of test performance, which was the main limitation of this method. There was no confidence interval for the LCM-derived sensitivity and specificity estimates, because these estimates were calculated from combinations of conditional probabilities, which had individual maximum-likelihood and estimated standard errors. During the analysis of the tests, bivariate residuals of LSM-APRI, LSM-FIB-4 and LSM-GP impaired the fit of models; therefore, more studies should be performed to identify the causes of high discordances rates between these pairs including their intra- and inter-observers variability. The cut-off of all tests had been controversial in different studies, which were still not unanimous. Much more clinic researches are needed to get accurate cut-off and raise the diagnostic efficacy. In this study, we quoted the WHO HBV guideline and strict meta-analysis to define the cut-offs for the tests. Liver biopsy was not performed in this study, so we could not compare and verify assessment of tests between traditional analysis according to liver biopsy as a gold standard and model-estimation using LCM-R. In spite of these limitations, this study was estimation and verification of the tests and the performances cannot represent the true status, but we believe it is approximate to the truth.

5.3. Conclusions

In this model without gold standard, high specificity and sensitivity (> 90%) of FIB-4 were confirmed for diagnosis of significant fibrosis. APRI also had sub-optimal diagnosis accuracy (sensitivity > 90%, specificity > 70%) for significant fibrosis. LSM showed the best diagnosis value for cirrhosis with the highest sensitivity (37.03%) and well specificity (78.64%).

Through the estimation of above four noninvasive fibrosis tests by LCM-R model, we could get their diagnostic performances and relative dominance of each test in diagnosis and select the best or combined test to achieve the best accuracy for clinical application depending on their diagnostic values.

Acknowledgments

The authors thank Pro Xuegong Fan, Doctor Xing Li and Doctor Yin Zong Xiao for supporting the study and revising.

Footnote

Authors' Contribution: Yi Xiang Zheng developed the original idea and protocol, abstracted and analyzed data, and wrote the manuscript. Shu Juan Ma and Meng Hou Lu contributed to development of the protocol, abstracted data, and prepared the manuscript.

References

1. Trepo C, Chan HL, Lok A. Hepatitis B virus infection. *Lancet.* 2014;**384**(9959):2053–63. doi: 10.1016/s0140-6736(14)60220-8. [PubMed: 24954675]
2. Yuen MF, Lai CL. Hepatitis B in 2014: HBV research moves forward–receptors and reactivation. *Nat Rev Gastroenterol Hepatol.* 2015;**12**(2):70–2. doi: 10.1038/nrgastro.2014.226. [PubMed: 25511530]
3. Schuppan D, Afdhal NH. Liver cirrhosis. *Lancet.* 2008;**371**(9615):838–51. doi: 10.1016/s0140-6736(08)60383-9. [PubMed: 18328931]
4. van Bommel F, Berg T. Treatment of HBV related cirrhosis. *Liver Int.* 2013;**33 Suppl 1**:176–81. doi: 10.1111/liv.12074. [PubMed: 23286862]
5. Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. *Lancet.* 2014;**383**(9930):1749–61. doi: 10.1016/s0140-6736(14)60121-5. [PubMed: 24480518]
6. European Association For The Study Of The L. EASL clinical practice guidelines: Management of chronic hepatitis B virus infection. *J Hepatol.* 2012;**57**(1):167–85. doi: 10.1016/j.jhep.2012.02.010. [PubMed: 22436845]
7. WHO. *Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection.* World Health Organization; 2015. Available from: http://www.who.int/hiv/pub/hepatitis/hepatitis-b-guidelines/en/.
8. Mehta SH, Lau B, Afdhal NH, Thomas DL. Exceeding the limits of liver histology markers. *J Hepatol.* 2009;**50**(1):36–41. doi: 10.1016/j.jhep.2008.07.039. [PubMed: 19012989]
9. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics.* 1996;**52**(3):797–810. [PubMed: 8805757]
10. Poynard T, de Ledinghen V, Zarski JP, Stanciu C, Munteanu M, Vergniol J, et al. Relative performances of FibroTest, Fibroscan, and biopsy for the assessment of the stage of liver fibrosis in patients with chronic hepatitis C: a step toward the truth in the absence of a gold standard. *J Hepatol.* 2012;**56**(3):541–8. doi: 10.1016/j.jhep.2011.08.007. [PubMed: 21889468]
11. Castera L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology.* 2010;**51**(3):828–35. doi: 10.1002/hep.23425. [PubMed: 20063276]
12. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology.* 2003;**38**(2):518–26. doi: 10.1053/jhep.2003.50346. [PubMed: 12883497]
13. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology.* 2006;**43**(6):1317–25. doi: 10.1002/hep.21178. [PubMed: 16729309]
14. Li Y, Chen Y, Zhao Y. The diagnostic value of the FIB-4 index for staging hepatitis B-related fibrosis: a meta-analysis. *PLoS One.* 2014;**9**(8):e105728. doi: 10.1371/journal.pone.0105728. [PubMed: 25165830]
15. Liu XD, Wu JL, Liang J, Zhang T, Sheng QS. Globulin-platelet model predicts minimal fibrosis and cirrhosis in chronic hepatitis B virus infected patients. *World J Gastroenterol.* 2012;**18**(22):2784–92. doi: 10.3748/wjg.v18.i22.2784. [PubMed: 22719186]
16. Hadgu A, Qu Y. A biomedical application of latent class models with random effects. *J R Stat Soc Ser C-Appl Stat.* 1999;**48**(3) doi: 10.1111/1467-9876.00162.
17. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess.* 2007;**11**(50):iii. [PubMed: 18021577]
18. Brito DA, Strauss J, Magidson V, Tikhonenko I, Khodjakov A, Koonce MP. Pushing forces drive the comet-like motility of microtubule arrays in Dictyostelium. *Mol Biol Cell.* 2005;**16**(7):3334–40. doi: 10.1091/mbc.E05-01-0057. [PubMed: 15857957]
19. Zhang B, Chen Z, Albert PS. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics.* 2012;**68**(4):1294–302. doi: 10.1111/j.1541-0420.2012.01789.x. [PubMed: 23006010]
20. Xu XY, Kong H, Song RX, Zhai YH, Wu XF, Ai WS, et al. The effectiveness of noninvasive biomarkers to predict hepatitis B-related significant fibrosis and cirrhosis: a systematic review and meta-analysis of diagnostic test accuracy. *PLoS One.* 2014;**9**(6):e100182. doi: 10.1371/journal.pone.0100182. [PubMed: 24964038]
21. Salkic NN, Cickusic E, Jovanovic P, Denjagic MB, Iljazovic-Topcic S, Bevanda M, et al. Online combination algorithm for non-invasive assessment of chronic hepatitis B related liver fibrosis and cirrhosis in resource-limited settings. *Eur J Intern Med.* 2015;**26**(8):628–34. doi: 10.1016/j.ejim.2015.07.005. [PubMed: 26194460]
22. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol.* 2002;**97**(10):2614–8. doi: 10.1111/j.1572-0241.2002.06038.x. [PubMed: 12385448]
23. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology.* 2003;**38**(6):1449–57. doi: 10.1016/j.hep.2003.09.022. [PubMed: 14647056]
24. Mallet V, Dhalluin-Venier V, Roussin C, Bourliere M, Pettinelli ME, Giry C, et al. The accuracy of the FIB-4 index for the diagnosis of mild fibrosis in chronic hepatitis B. *Aliment Pharmacol Ther.* 2009;**29**(4):409–15. doi: 10.1111/j.1365-2036.2008.03895.x. [PubMed: 19035983]
25. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology.* 2007;**46**(1):32–6. doi: 10.1002/hep.21669. [PubMed: 17567829]
26. Wang H, Xue L, Yan R, Zhou Y, Wang MS, Cheng MJ, et al. Comparison of FIB-4 and APRI in Chinese HBV-infected patients with persistently normal ALT and mildly elevated ALT. *J Viral Hepat.* 2013;**20**(4):e3–10. doi: 10.1111/jvh.12010. [PubMed: 23490387]
27. Suh B, Park S, Shin DW, Yun JM, Yang HK, Yu SJ, et al. High liver fibrosis index FIB-4 is highly predictive of hepatocellular carcinoma in chronic hepatitis B carriers. *Hepatology.* 2015;**61**(4):1261–8. doi: 10.1002/hep.27654. [PubMed: 25502481]
28. Xiao G, Yang J, Yan L. Comparison of diagnostic accuracy of aspartate aminotransferase to platelet ratio index and fibrosis-4 index for detecting liver fibrosis in adult patients with chronic hepatitis B virus infection: a systemic review and meta-analysis. *Hepatology.* 2015;**61**(1):292–302. doi: 10.1002/hep.27382. [PubMed: 25132233]
29. Ogawa E, Furusyo N, Toyoda K, Takeoka H, Otaguro S, Hamada M, et al. Transient elastography for patients with chronic hepatitis B and C virus infection: Non-invasive, quantitative assessment of liver fibrosis. *Hepatol Res.* 2007;**37**(12):1002–10. doi: 10.1111/j.1872-034X.2007.00160.x. [PubMed: 17608672]
30. Marcellin P, Ziol M, Bedossa P, Douvin C, Poupon R, de Ledinghen V, et al. Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B. *Liver Int.* 2009;**29**(2):242–7. doi: 10.1111/j.1478-3231.2008.01802.x. [PubMed: 18637064]
31. Sporea I, Sirli R, Deleanu A, Tudora A, Popescu A, Curescu M, et al. Liver stiffness measurements in patients with HBV vs HCV chronic hepatitis: a comparative study. *World J Gastroenterol.* 2010;**16**(38):4832–7. [PubMed: 20939112]

32. Degos F, Perez P, Roche B, Mahmoudi A, Asselineau J, Voitot H, 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**(6):1013–21. doi: 10.1016/j.jhep.2010.05.035. [PubMed: 20850886]
33. Zhu X, Wang LC, Chen EQ , Chen XB, Chen LY, Liu L, et al. Prospective evaluation of FibroScan for the diagnosis of hepatic fibrosis compared with liver biopsy/AST platelet ratio index and FIB-4 in patients with chronic HBV infection. *Dig Dis Sci.* 2011;**56**(9):2742–9. doi: 10.1007/s10620-011-1659-1. [PubMed: 21399926]
34. Coco B, Oliveri F, Maina AM, Ciccorossi P, Sacco R, Colombatto P, et al. Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat.* 2007;**14**(5):360–9. doi: 10.1111/j.1365-2893.2006.00811.x. [PubMed: 17439526]
35. Wong VW, Lampertico P, de Ledinghen V, Chang PE, Kim SU, Chen Y, et al. Probability-based interpretation of liver stiffness measurement in untreated chronic hepatitis B patients. *Dig Dis Sci.* 2015;**60**(5):1448–56. doi: 10.1007/s10620-014-3488-5. [PubMed: 25563720]