

RESEARCH ARTICLE

Identification and external validation of the optimal FIB-4 and APRI thresholds for ruling in chronic hepatitis B related liver fibrosis in tertiary care settings

Kecheng Liu¹  | Mengbin Qin¹ | Kunlin Tao¹ | Zhihai Liang² | Fuqing Cai¹ | Liebin Zhao¹ | Peng Peng¹ | Shiquan Liu¹ | Jun Zou¹ | Jiean Huang¹

¹Department of Gastroenterology, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China

²Department of Gastroenterology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

Correspondence

Jiean Huang, Department of Gastroenterology, The Second Affiliated Hospital of Guangxi Medical University, Nanning 530007, China.
Email: hjagxmu@163.com

Abstract

Background: With the initially defined thresholds, the most widely used serum biomarkers for staging liver fibrosis (ie, APRI and FIB-4 scores) proved to be ineffective among patients with chronic hepatitis B virus infection (CHB). Whether optimizing the FIB-4 and APRI thresholds could improve their diagnostic accuracy requires further research.

Methods: Using data of treat-naïve CHB patients from three tertiary hospitals, we explored the optimal FIB-4 and APRI thresholds to rule in liver fibrosis accurately. Subsequently, we validated the applicability of the newly defined thresholds to the CHB patients from another two tertiary hospitals.

Results: The fibrosis stages between discovery cohort ($n = 433$) and the external validation cohort ($n = 568$) were statistically different ($P < .001$). When ruling in significant fibrosis and advanced fibrosis by the newly defined FIB-4 thresholds (2.25 and 3.00, respectively), 24.0% and 14.3% of patients, respectively, could be classified with excellent accuracy (PPVs of 91.3% and 80.6%, respectively; misdiagnosis rates of 6.0% and 5.4%, respectively), supported by the internal and external validation tests. Regrettably, the more accurate and robust thresholds of APRI score for ruling in significant fibrosis and advanced fibrosis could not be found. Besides, the FIB-4 and APRI scores should not be recommended for ruling in cirrhosis because of poor clinical diagnostic performance.

Conclusion: The newly defined FIB-4 thresholds for ruling in significant fibrosis and advanced fibrosis showed superior and reproducible clinical diagnostic accuracy. The well-validated threshold (≥ 2.25) of FIB-4 score could aid in antiviral treatment decisions for treat-naïve adult CHB patients by accurately ruling in significant fibrosis in tertiary care settings.

KEYWORDS

biomarkers, clinical decision rules, diagnostic errors, hepatitis B, liver cirrhosis

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Journal of Clinical Laboratory Analysis* Published by Wiley Periodicals LLC

1 | INTRODUCTION

Chronic hepatitis B virus infection (CHB) has become a severe public health burden all over the world.¹ Because CHB patients with significant hepatic fibrosis have a high risk of complications, staging of hepatic fibrosis is essential for managing CHB patients in clinical practice. Liver biopsy, the imperfect gold standard for detecting liver fibrosis, is not favored by physicians or patients because of its invasiveness, sampling errors, and limitations.² In this situation, non-invasive alternative methods, such as transient elastography technology and serum test formula, have attracted considerable attention. However, transient elastography technology performed with FibroScan (Echosens, Paris, France) has not been widely available currently due to the high cost of equipment, especially for resource-limited areas. Serum test formulas based on routine hematological and biochemical tests are inexpensive and more accessible, appearing to be ideal alternatives to liver biopsy. Among them, the fibrosis index based on the four factors (FIB-4) and the aspartate transaminase to platelet ratio index (APRI) recommended by various clinical guidelines and expert consensus²⁻⁷ have been extensively used to stage CHB related hepatic fibrosis.

Unfortunately, the performance of FIB-4 and APRI scores with initially defined thresholds for ruling in mild fibrosis and significant fibrosis proved to be unsatisfactory in CHB patients because of very high misdiagnosis rates.⁸ Two recent meta-analyses^{9,10} concerning the diagnostic value of FIB-4 score for hepatic fibrosis in CHB patients show that the diagnostic performance is affected by the range of thresholds, implying that optimizing the diagnostic thresholds may improve the diagnostic performance. Therefore, numerous studies^{8,10-12} have tried to re-determine the diagnostic thresholds of FIB-4 and APRI scores based on discriminative measurements (ie, sensitivity and specificity). However, sensitivity and specificity can only guide decisions on which diagnostic test to select, while negative predictive value (NPV) and positive predictive value (PPV) aim to rule out or rule in disease.² In clinical practice, what really matters is the number of patients accurately ruled out or ruled in by the serum test formulas.³ Consequently, for the serum test formulas to be useful in practice, identification of the optimal diagnostic thresholds should go beyond these discriminative parameters and take PPV and NPV into account.

Inspired by the recent success of identifying the optimal FIB-4 threshold for ruling out cirrhosis,¹³ we propose a similar strategy to explore the optimal thresholds of FIB-4 and APRI scores for ruling in hepatic fibrosis among adult CHB patients. The clinical diagnostic accuracy of newly defined optimal thresholds was then assessed by internal and external validation tests.

2 | MATERIALS AND METHODS

2.1 | Patients

A total of 257 consecutive treat-naïve adult CHB patients undergoing liver biopsies for assessing hepatic fibrosis stage at the Second

Affiliated Hospital of Guangxi Medical University and the First Affiliated Hospital of Guangxi Medical University (two tertiary hospitals in the city of Nanning) between May 2012 and April 2019 were screened for eligibility. CHB patients were defined as those patients with the persistent presence of serum HBsAg for more than half a year. Patients satisfied any of the following conditions were excluded: (a) co-infection with human immunodeficiency virus, hepatitis E virus, hepatitis D virus, or hepatitis C virus, (b) coexistence of alcoholic liver disease, autoimmune liver disease, drug hepatitis, biliary cirrhosis, or Wilson's disease; (c) status of liver transplantation or decompensated cirrhosis; (d) diagnoses of diabetes, hematological diseases; (e) coexistence of severe cardiovascular, pulmonary, renal, gastrointestinal system diseases; (f) antibiotic usage in the past 6 weeks; (g) glucocorticoid or immunosuppressive agents usage in the past 6 months. One hundred and seventy-seven eligible patients without missing data of key variables (ie, age, platelet count, ALT, AST, fibrosis stage) are included as a part of discovery dataset (named Nanning dataset, shown in Figure S1). Besides, an anonymous dataset containing 256 CHB patients from Huai'an No. 4 People's Hospital between 2008 and 2015 was also added to the discovery dataset (named Huai'an dataset). The details of the Huai'an dataset have been reported elsewhere.¹⁴ Finally, a total of 433 treatment-naïve CHB patients undergoing biopsies for assessing hepatic fibrosis stage formed the discovery dataset. The research protocol was permitted, and the requirement of patient informed consent was waived by the Ethics Committee of the Second Affiliated Hospital of Guangxi Medical University.

An external validation dataset from an existing study¹⁵ included 568 treatment-naïve or treatment-experienced CHB patients of another two cities (Shanghai and Xiamen). Detailed information about these participants was available in the original literature.

2.2 | FIB-4 and APRI scores

FIB-4 and APRI scores (named fibrosis scores) were calculated based on the following formulas: $FIB-4 = (age \times AST) / (\text{platelet count} \times (ALT)^{1/2})$; $APRI = ([AST/ULN \text{ of } AST] / \text{platelet count}) \times 100$, the upper limit of normal (ULN) of AST was set at 40IU/L. To compare the performance for avoiding misdiagnosis, the pre-defined thresholds based on the specificity of 90% ($APRI \geq 1.74$ and $FIB-4 \geq 1.90$ for ruling in significant fibrosis, $APRI \geq 2.00$ and $FIB-4 \geq 2.31$ for ruling in cirrhosis)⁸ were used as the reference. If not available, the initially defined thresholds ($FIB-4 \geq 3.25$ ¹⁶ for ruling in advanced fibrosis) were used. The APRI score was derived for detecting significant fibrosis and cirrhosis, and thus, no reference threshold was available for ruling in advanced fibrosis in this study.¹⁷ To obtain laboratory characteristics (including ALT, AST, platelet count) of the subjects from Nanning dataset, the routine laboratory tests were performed using overnight fasting blood samples which were collected within 2 weeks before liver biopsies. Specifically, platelet count was detected with a Beckman Coulter LH 750 (Beckman Coulter, Inc.); ALT and AST were determined via a Hitachi 7600 automatic biochemical

analyzer. The FIB-4 and APRI scores of Huai'an dataset and the external validation dataset were obtained from the original publication.

2.3 | Liver biopsy

Liver tissue was obtained by ultrasonography-guided liver biopsy for histopathological examination. The Scheuer¹⁸ or METAVIR¹⁹ scoring systems were adopted as the pathological diagnosis standard of hepatic fibrosis. Significant fibrosis, advanced fibrosis, and cirrhosis were defined as Scheuer F2-4 or METAVIR F2-4, Scheuer F3-4 or METAVIR F3-4, and Scheuer F4 or METAVIR F4, respectively. The fibrosis stage of all patients included in this analysis was evaluated by at least two pathologists who were blinded to clinical data.^{14,15}

2.4 | Statistical analysis

Correlations between fibrosis scores and fibrosis stages were examined using Spearman's rank correlation coefficient. The areas under the receiver operating characteristic curve (AUROC), specificity, sensitivity, PPV, misdiagnosis rate (ie, 1-specificity), positive likelihood ratio (PLR), and diagnostic odds ratio (DOR)²⁰ were calculated to evaluate the diagnosis accuracy. The violin plots were selected to visualize the distribution of fibrosis scores within different fibrosis stages. For the fibrosis scores to be useful in practice, more than 10% of patients ought to be ruled in by the newly defined thresholds. In the discovery cohort, 10% of patients have FIB-4 scores of more than 3.56, and 10% of patients have APRI scores of more than 2.13. Therefore, we calculated the diagnostic performance metrics for FIB-4 cutoffs ranging from 0.01 to 3.56 by increments of 0.01, and for APRI cutoffs ranging from 0.01 to 2.13 by increments of 0.01, respectively. We tried to optimize the thresholds of fibrosis scores in the discovery dataset, aiming for accurately (ie, with high PPVs and low misdiagnosis rates) ruling in as more patients with hepatic fibrosis as possible. To define the optimal thresholds for ruling in hepatic fibrosis, PPVs and misdiagnosis rates were plotted against hypothetical cutoffs using Microsoft Excel software (version 2019; Microsoft Corporation). To address the issue of overfitting, the diagnosis performance of the newly defined cutoffs was internally validated by bootstrap methods²¹ with 500 replicates. Besides, further validation was also performed in the external validation dataset. Statistical analyses were conducted with the R software (version 3.6.1).

3 | RESULTS

3.1 | Patient baseline characteristics

A total of 1001 patients were enrolled in this analysis. The discovery dataset consisted of 433 patients from two cities (Nanning and Huai'an); the external validation dataset consisted of 568 patients

from another two cities (Shanghai and Xiamen). The patient characteristics of both discovery dataset and external validation dataset are detailed in Table 1. There was no statistical difference in FIB-4 scores between discovery dataset and external validation dataset ($P = .482$). The APRI scores of external validation dataset were higher than those of discovery dataset ($P < .001$). The median lengths of biopsy samples in Nanning cohort and Huai'an cohort were 1.6 cm (interquartile range: 1.2-2.2 cm) and 1.3 cm (interquartile range: 1.0-1.5 cm), respectively. The distribution of fibrosis stages between the discovery dataset and external validation dataset showed statistically significant differences ($P < .001$).

As shown in Figure 1A,B, significant differences of FIB-4 and APRI scores were observed between stage F0-1 and stage F2-4 (all P values $< .001$). Similar results were obtained when discriminating stage F0-2 and stage F3-4 (shown in Figure 1C,D) and discriminating stage F0-3 and stage F4 (shown in Figure 1E,F).

3.2 | Diagnostic performance of FIB-4 scores and APRI scores with pre-defined thresholds

For ruling in significant fibrosis, the AUROCs of FIB-4 and APRI were 0.727 and 0.731, respectively. According to the FIB-4 threshold calculated at the 90% specificity level (ie, 1.90), 28.6% patients had a FIB-4 score in the classifiable range with a PPV of 87.9% and a misdiagnosed rate of 9.9% (shown in Table 2). When the APRI threshold calculated at the 90% specificity level (ie, 1.74) was applied to rule in significant fibrosis, the PPV, misdiagnosed rate, and the proportion of classifiable patients were 88.9%, 4.0%, and 12.4%, respectively (shown in Table 3).

For ruling in advanced fibrosis, the AUROCs of FIB-4 and APRI were 0.733 and 0.718, respectively. The FIB-4 with the initially defined threshold (ie, 3.25) ruled in advanced fibrosis with a PPV of 81.1%, a misdiagnosis rate of 4.5%, and a small proportion of classifiable patients (ie, 12.2%) (shown in Table 2).

For ruling in cirrhosis, the AUROCs of FIB-4 and APRI were 0.763 and 0.707, respectively. Although 23.1% of patients could be identified according to the FIB-4 threshold (ie, 2.31) calculated at the 90% specificity level, the PPV of 50.0% was unacceptably low (ie, half of the patients ruled in as having cirrhosis were misclassified). Similarly, only 10.4% patients could be identified according to the APRI threshold (ie, 2.00) calculated at the 90% specificity level, and the PPV of 44.4% was also unacceptably low (shown in Table 4).

3.3 | Exploration of the optimal FIB-4 and APRI thresholds for ruling in hepatic fibrosis

Both the FIB-4 scores and APRI scores were positively correlated with hepatic fibrosis stages ($r = .47$ and $r = .43$, respectively, both P values $< .0001$). Therefore, the thresholds of the FIB-4 and APRI scores for ruling in specific fibrosis scale (eg, significant fibrosis) increased within a particular range, along with higher PPVs, lower

TABLE 1 Baseline of patients included in this analysis

	Discovery dataset (n = 433)		External validation dataset (Shanghai + Xiamen, n = 568)	P value*
	Nanning (n = 177)	Huai'an (n = 256)		
Male (% , n)	68.4% (121/177)	79.7% (204/256)	71.4% (411/576) ^a	NA
Age (years) (median, IQR)	41.0 (34.0-47.0)	38.0 (29.0, 46.0)	35.0 (28.0-44.2)	NA
AST (U/L) (median, IQR)	28.0 (22.0-40.0)	36.0 (26.0, 58.5)	57.0 (37.0-100.2)	NA
ALT (U/L) (median, IQR)	34.0 (24.0-47.0)	42.0 (28.0, 79.8)	99.0 (49.8-204.2)	NA
PLT (10 ⁹ /L) (median, IQR)	174.9 (149.0-217.6)	148.5 (110.5,190.5)	166.5 (125.8-201.2)	NA
APRI (median, IQR)	0.5 (0.3-1.0)		1.0 (0.6-1.9)	<.001
FIB-4 (median, IQR)	1.2 (0.8-2.2)		1.2 (0.8-2.4)	.482
Fibrosis stage				
F0(% , n)	15.0% (65)		10.7% (61)	<.001
F1(% , n)	19.9% (86)		33.1% (188)	
F2(% , n)	16.6% (72)		25.7% (146)	
F3(% , n)	23.3% (101)		11.6% (66)	
F4(% , n)	25.2% (109)		18.8% (107)	

Note: NA, difference test could not be done, because raw data of some variables (including male, age, AST, ALT, PLT) from Huai'an dataset were not available.

Abbreviations: ALT, alanine aminotransferase; AST, aminotransferase; PLT, platelet count.

^aEight patients with missing data in the validation cohort are deleted in the analysis except for sex ratios.

*P value compare patients from discovery dataset against those from external validation dataset (Wilcoxon rank-sum test for continuous variables; Chi-square test for categorical variables).

misdiagnosis rates but smaller proportions of classifiable patients (ie, the patients with fibrosis scores higher than positive thresholds). In this situation, the optimal thresholds were determined in correspondence to the best compromise between the clinical diagnosis accuracy (PPV and misdiagnosis rate) and the proportion of classifiable patients based on Figures 1 and 2.

As shown in Figure 2A and Table 2, the PPVs reached levels of above 90% (ie, 91.3%), and misdiagnosis rates reached levels of below 10% (ie, 6.0%) with a FIB-4 of 2.25 or more for ruling in significant fibrosis, exhibiting little change for higher FIB-4. Similarly, the PPVs reached levels of above 80% (ie, 80.6%), and misdiagnosis rates reached levels of below 10% (ie, 5.4%) with a FIB-4 of 3.00 or more for ruling in advanced fibrosis, exhibiting little change for higher FIB-4 (shown in Figure 2C and Table 2). Therefore, the FIB-4 of 2.25 and 3.00 were regarded as the optimal thresholds for ruling in significant fibrosis and advanced fibrosis, respectively. The diagnosis performance estimated by bootstrap resampling methods agreed with those calculated from discovery dataset. The PPVs and misdiagnosis rates derived from external validation dataset were worse slightly, but still acceptable for use in clinical practice (shown in Table 2).

For ruling in significant fibrosis, the newly defined FIB-4 cutoff (ie, 2.25) exhibited better diagnostic accuracy (PPV 91.3% vs 87.9%, misdiagnosis rate 6.0% vs 9.9%, DOR 8.0 vs 5.7, PLR 5.7 vs 3.9) as compared with the pre-defined FIB-4 cutoff (ie, 1.90). For ruling in advanced fibrosis, FIB-4 score at the newly defined cutoff (ie, 3.00) offered more patients (14.3% vs 12.2%) with similar diagnostic accuracy (PPV 80.6% vs 81.1%, misdiagnosis rate 5.4% vs

4.5%, DOR 5.5 vs 5.5, PLR 4.4 vs 4.6) than did at the pre-defined FIB-4 cutoff (ie, 3.25).

The optimal thresholds of APRI for ruling in significant fibrosis and advanced fibrosis were presented in Table 3. Interestingly, the APRI of 1.15 was chosen as the optimal threshold for ruling in significant fibrosis and advanced fibrosis (shown in Figure 2B,D). When ruling in significant fibrosis and advanced fibrosis with APRI at a threshold of 1.15, the proportions of classifiable patients, PPVs, and misdiagnosis rates were 19.4% and 19.4%, 90.5% and 81.0%, and 5.3% and 7.2%, respectively. The diagnosis metrics calculated from internal validation tests showed a high degree of similarity to those derived from the discovery dataset. Unfortunately, although the proportions of classifiable patients in the external validation dataset increased markedly, the PPVs reduced severely, and misdiagnosis rates increased significantly (shown in Table 3).

As shown in Figure 2E,F, the highest PPVs for ruling in cirrhosis with FIB-4 and APRI scores could only reach levels of about 60% and 50%, respectively, which were insufficiently precise in clinical diagnostics. Consequently, there was no need to explore the optimal thresholds of both FIB-4 and APRI scores for diagnosing cirrhosis in CHB patients.

4 | DISCUSSION

Chronic hepatitis B virus infection continues to be prevalent all around the world.¹ Accurate diagnosis of hepatic fibrosis is critical for CHB patient treatment, surveillance, and prognosis.^{2,3} In

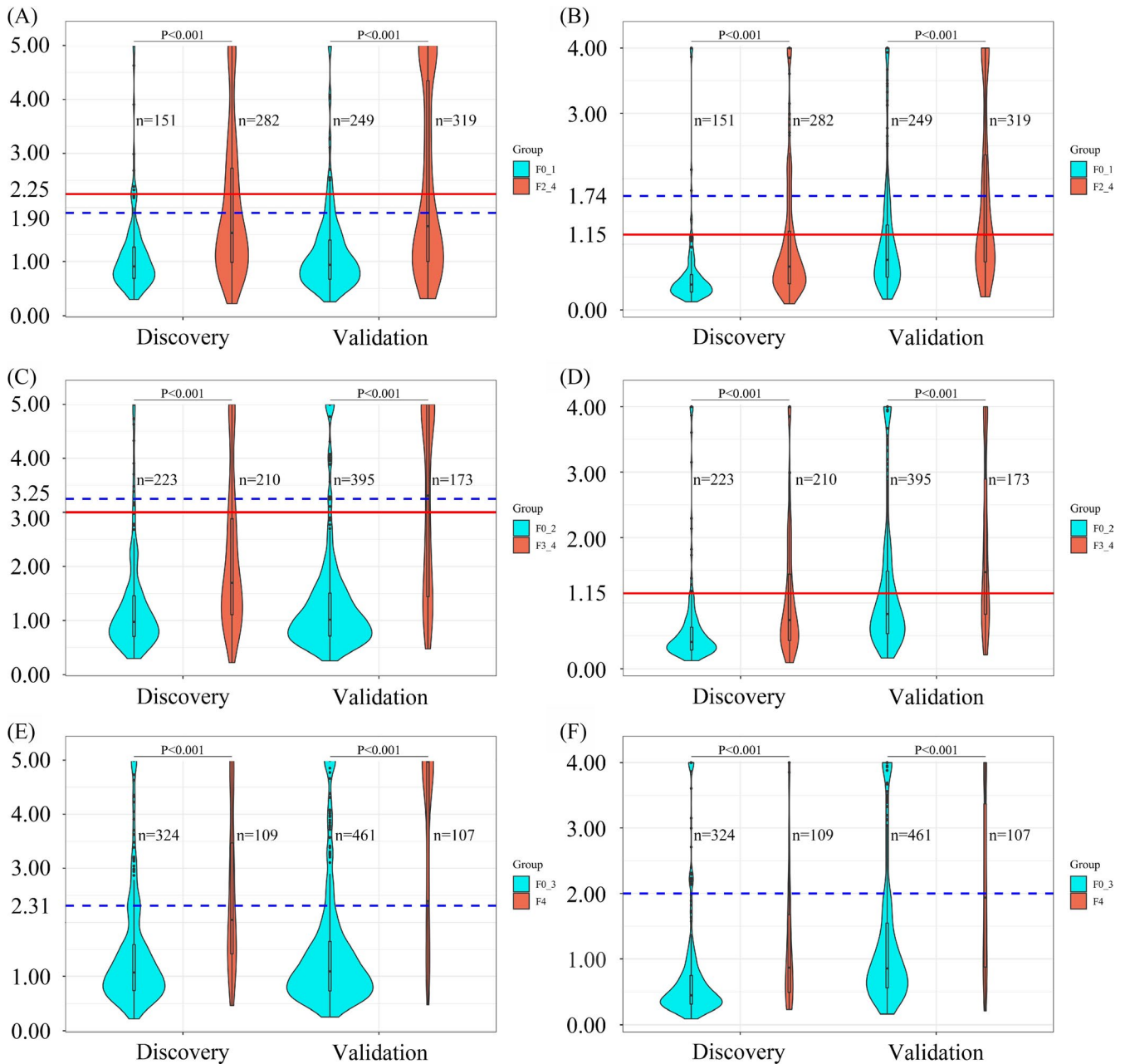


FIGURE 1 Violin plots showing the distribution of FIB-4 and APRI scores within different hepatic fibrosis stages, with overlay boxplots displaying medians and upper and lower quartiles. (A) FIB-4 scores between stage F0-1 and stage F2-4, (B) APRI scores between stage F0-1 and stage F2-4, (C) FIB-4 scores between stage F0-2 and stage F3-4, (D) APRI scores between stage F0-2 and stage F3-4, (E) FIB-4 scores between stage F0-3 and stage F4, and (F) APRI scores between stage F0-3 and stage F4. In each figure, the area of the violin plot is proportional to the number of patients. The blue dotted lines indicate pre-defined thresholds, and solid red lines indicate newly defined thresholds. The P values were calculated using the Wilcoxon rank-sum test

the past decades, non-invasive diagnosis methods have been a research hotspot of diagnosing hepatic fibrosis stage. Unfortunately, the most widely used serum test formulas in clinical practice (ie, FIB-4 and APRI scores) proved to be ineffective in predicting CHB related liver fibrosis, because a large number of CHB patients without fibrosis were misdiagnosed as having mild fibrosis or significant fibrosis.⁸ In clinical settings, a false-positive result can lead to unnecessarily or prematurely antiviral therapy and associated risk of subsequent potential drug resistance and drug

toxicities.⁷ Compared to liver biopsy, the two non-invasive serum test formulas could obtain short-term cost savings. Still, the long-term consequences of misdiagnosis in light of treatment costs and health outcomes may outweigh the short-term gains in convenience and diagnostic fee. Although numerous researches^{8,10-12} have been conducted to re-establish the diagnostic thresholds of FIB-4 and APRI scores for improving their diagnostic accuracy, few studies have focused on reducing the chance of misdiagnosis. To re-define the optimal thresholds of FIB-4 and APRI scores for

TABLE 2 Diagnostic performance of FIB-4 score for ruling in significant fibrosis and advanced fibrosis

	Discovery (n = 433)		External validation (n = 568)	Internal validation ^d	
				Mean, SD	95% CI
Significant fibrosis					
Cutoff	≥1.90 ^a	≥2.25 ^c	≥2.25	≥2.25	
Classifiable patients (% (n/N))	28.6% (124/433)	24.0% (104/433)	25.9% (147/568)	24.0%, 2.0% (103.8/433, 8.8/433)	19.4% ~ 27.8% (84.0/433 ~ 120.0/433)
Specificity	90.1%	94.0%	93.2%	94.0%, 1.9%	90.0% ~ 97.3%
Sensitivity	38.7%	33.7%	40.8%	33.9%, 2.9%	27.8% ~ 39.7%
PPV	87.9%	91.3%	88.4%	91.3%, 2.8%	85.2% ~ 96.2%
Misdiagnosis rate (% (n/N))	9.9% (15/151)	6.0% (9/151)	6.8% (17/249)	6.0%, 1.9% (9.1/151, 2.9/151)	2.7% ~ 10.0% (4.1/151, 15.1/151)
DOR	5.7	8.0	9.4	9.1, 4.1	4.4 ~ 19.4
PLR	3.9	5.7	6.0	6.2, 2.3	3.2 ~ 12.2
Advanced fibrosis					
Cutoff	≥3.25 ^b	≥3.00 ^c	≥3.00	≥3.00	
Classifiable patients (% (n/N))	12.2% (53/433)	14.3% (62/433)	21.0% (119/568)	14.3%, 1.6% (61.9/433, 7.1/433)	11.4% ~ 17.6% (49.5/433 ~ 76.0/433)
Specificity	95.5%	94.6%	92.9%	94.7%, 1.4%	91.9% ~ 97.4%
Sensitivity	20.5%	23.8%	52.6%	24.1%, 3.1%	17.6% ~ 30.5%
PPV	81.1%	80.6%	76.5%	80.9%, 5.0%	71.0% ~ 90.6%
Misdiagnosis rate (% (n/N))	4.5% (10/223)	5.4% (12/223)	7.1% (28/395)	5.3%, 1.4% (12.4/223, 3.3/223)	2.6% ~ 8.1% (6.1/223, 18.9/223)
DOR	5.5	5.5	14.6	6.2, 2.4	3.1 ~ 12.8
PLR	4.6	4.4	7.4	4.9, 1.8	2.6 ~ 9.4

Note: Classifiable patients, the proportion of patients with APRI or FIB-4 scores higher than the positive thresholds.

Abbreviations: DOR, diagnostic odds ratio; PLR, positive likelihood ratio; PPV, positive predictive value.

^athreshold calculated by 90% specificity.⁸

^bthreshold initially defined.¹⁶

^cthreshold newly defined.

^d500 bootstrap replicates were generated by using resampling with replacement, and averages of these samples were calculated to evaluate the stability of the newly defined thresholds.

ruling in CHB related fibrosis, further investigations with a more reasonable design are needed.

Several previous studies^{8,11,12} described a seemingly improved strategy (ie, based on the specificity of 90%) to re-establish the diagnostic thresholds, aiming to reduce the misdiagnosis rate to 10%. However, the discriminative measurement (ie, specificity) cannot directly show how well the positive test result for individual CHB patient predicts the probability of liver fibrosis in that patient (ie, PPV).²² In daily clinical work, high PPV values for detecting liver fibrosis would provide useful references for clinicians in making proper medical decisions.³ For the newly defined thresholds to be useful in practice, the optimal cutoffs of fibrosis scores should be determined by not only the discriminative measurements (ie, sensitivity and specificity) but also the predictive values (ie, PPV and NPV).¹³ Besides, for the newly defined cutoffs could be generalized beyond the study, the internal and external validation tests are required to conduct.^{21,23} In the present study, we tried to identify and validate the optimal FIB-4/APRI cutoffs based on multicenter data, aiming for

accurately (ie, with high PPVs and low misdiagnosis rates) ruling in as more CHB patients with hepatic fibrosis as possible.

As shown in Figures 1 and 2, it is evident that the higher the thresholds within a specific range, the lower misdiagnosis rates, the higher PPVs, but the smaller proportions of classifiable patients. To accurately rule in CHB related fibrosis, we defined the optimal positive thresholds based on the best compromise between the clinical diagnosis accuracy (PPV and misdiagnosis rate) and the proportion of classifiable patients. For the FIB-4 score, we identified the cutoffs of 2.25 and 3.00 as the optimal thresholds for ruling in significant fibrosis and advanced fibrosis, respectively. These newly defined FIB-4 thresholds had roughly similar PPVs and misdiagnosis rates in both internal and external verification tests. For APRI score, it is surprising that the APRI cutoff of 1.15 showed excellent PPVs and misdiagnosis rates for ruling in significant fibrosis as well as advanced fibrosis. Though promising, the PPVs and misdiagnosis rates of newly defined APRI thresholds showed apparent deterioration in the external validation test.

TABLE 3 Diagnostic performance of APRI score for ruling in significant fibrosis and advanced fibrosis

	Discovery (n = 433)		External validation (n = 568)	Internal validation ^c	
				Mean, SD	95% CI
Significant fibrosis					
Cutoff	≥1.74 ^a	≥1.15 ^b	≥1.15	≥1.15	
Classifiable patients (% (n/ N))	12.4% (54/433)	19.4% (84/433)	40.7% (231/568)	19.3%, 1.9% (83.6/433, 8.4/433)	15.5% ~ 23.2% (67.0/433 ~ 100.5/433)
Specificity	96.0%	94.7%	71.1%	94.8%, 1.8%	91.2% ~ 97.9%
Sensitivity	17.0%	27.0%	49.8%	27.0%, 2.6%	22.2% ~ 32.4%
PPV	88.9%	90.5%	68.8%	90.5%, 3.2%	84.1% ~ 96.4%
Misdiagnosis rate (% (n/N))	4.0% (6/151)	5.3% (8/151)	28.9% (72/249)	5.2%, 1.8% (7.9/151, 2.7/151)	2.1% ~ 8.8% (3.2/151 ~ 13.3/151)
DOR	5.0	6.6	2.4	7.7, 3.9	3.6 ~ 19.3
PLR	4.3	5.1	1.7	5.8, 2.5	2.9 ~ 12.9
Advanced fibrosis					
Cutoff	NA	≥1.15 ^b	≥1.15	≥1.15	
Classifiable patients (% (n/ N))	NA	19.4% (84/433)	40.7% (231/568)	19.3%, 1.9% (83.6/433, 8.4/433)	15.5% ~ 23.2% (67.0/433 ~ 100.5/433)
Specificity	NA	92.8%	66.3%	92.9%, 1.6%	89.6% ~ 95.8%
Sensitivity	NA	32.4%	56.6%	32.4%, 3.2%	26.4%, 39.3%
PPV	NA	81.0%	42.4%	81.0%, 4.2%	72.9% ~ 89.2%
Misdiagnosis rate (% (n/N))	NA	7.2% (16/223)	33.7% (133/395)	7.1%, 1.6% (15.8/223, 3.6/223)	4.2% ~ 10.4% (9.4/223 ~ 23.2/223)
DOR	NA	6.2	2.6	6.7, 2.2	3.7 ~ 11.9
PLR	NA	4.5	1.7	4.8, 1.5	2.9 ~ 8.2

Note: NA, no reference threshold was available for ruling in advanced fibrosis in this study, because the APRI score was derived for detecting significant fibrosis and cirrhosis.¹⁷ Classifiable patients, the proportion of patients with APRI or FIB-4 scores higher than the positive thresholds. Abbreviations: DOR, diagnostic odds ratio; PLR, positive likelihood ratio; PPV, positive predictive value.

^athreshold calculated by 90% specificity.⁸

^bthreshold newly defined.

^c500 bootstrap replicates were generated by using resampling with replacement, and averages of these samples were calculated to evaluate the stability of the newly defined thresholds.

TABLE 4 Diagnostic performance of APRI and FIB-4 scores for ruling in cirrhosis

	Cutoff	Classifiable patients (% (n/N))	Specificity	Sensitivity	PPV	Misdiagnosis rate (% (n/N))	DOR	PLR
APRI	2.00 ^a	10.4% (45/433)	92.3%	18.3%	44.4%	7.7% (25/324)	2.7	2.4
FIB-4	2.31 ^b	23.1% (100/433)	84.6%	45.9%	50.0%	15.4% (50/324)	4.6	3.0

Note: Classifiable patients, the proportion of patients with APRI or FIB-4 scores higher than the positive thresholds.

Abbreviations: DOR, diagnostic odds ratio; PLR, positive likelihood ratio; PPV, positive predictive value.

^athreshold calculated by 90% specificity⁸ and recommended in various clinical guidelines and expert consensuses.³⁻⁷

^bthreshold calculated by 90% specificity.⁸

The diagnostic performance difference between discovery dataset and external validation dataset could be explained by three reasons. Firstly, the external validation dataset of current study enrolled not only treatment-naïve patients but also treatment-experienced patients. It has been confirmed that serum test formulas perform worse in treatment-experienced patients than in treatment-naïve patients because the serum aminotransferases and platelet count

were inevitably influenced by antiviral treatment.^{24,25} Secondly, the serum aminotransferases in our external validation dataset were apparently higher than those in discovery dataset (shown in Table 1), and high levels of serum aminotransferases may overestimate the non-invasive fibrosis scores, particularly APRI score.²⁶ This phenomenon has also been confirmed in the present study. The APRI score of external validation cohort was higher than that of discovery cohort

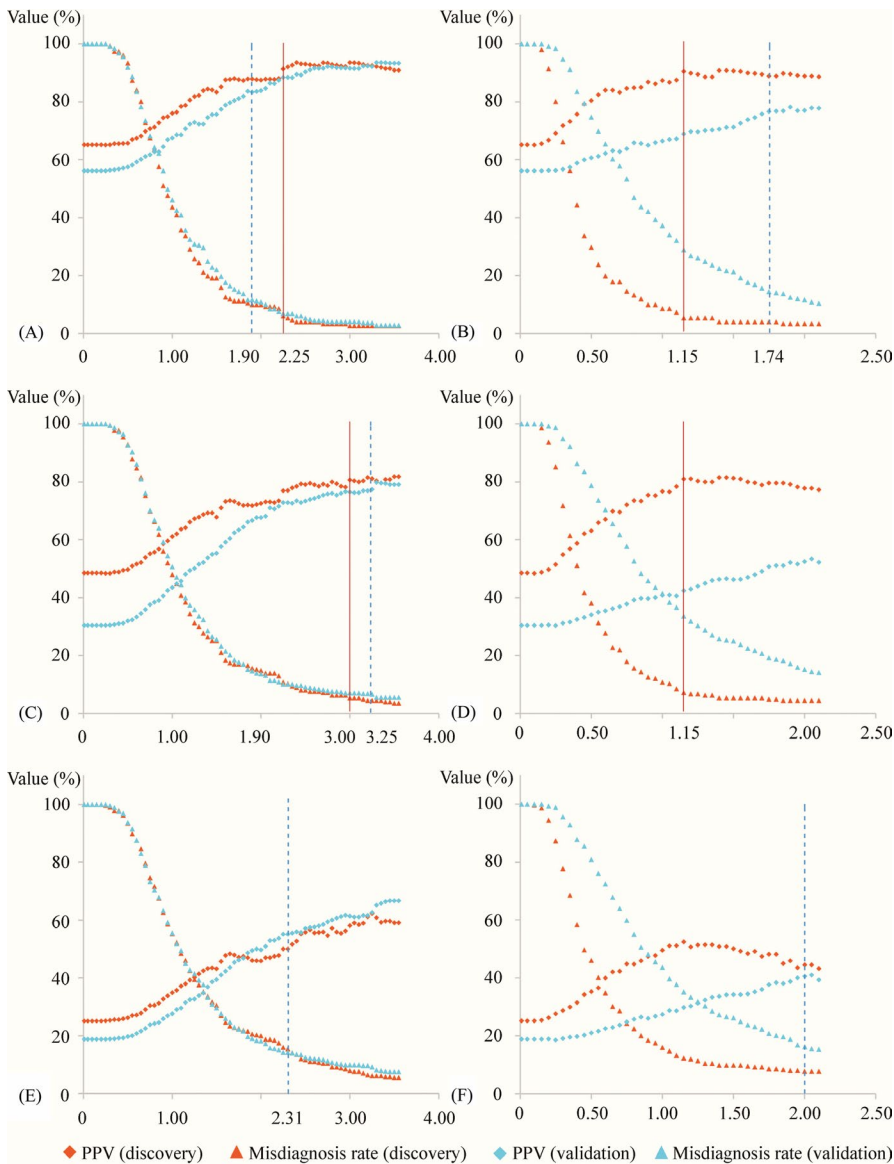


FIGURE 2 Scatterplots showing PPVs and misdiagnosis rates based on hypothetical cutoffs. (A) FIB-4 score for ruling in stage F2-4, (B) APRI score for ruling in stage F2-4, (C) FIB-4 score for ruling in stage F3-4, (D) APRI score for ruling in stage F3-4, (E) FIB-4 score for ruling in stage F4, and (F) APRI score for ruling in stage F4. The blue dotted lines indicate pre-defined thresholds, and solid red lines indicate newly defined thresholds

(median 1.0 vs 0.5, $P < .001$, shown in Table 1). Correspondingly, the discriminatory ability of APRI score was poorer in the external validation cohort (AUROCs change from 0.731 to 0.658 for detecting significant fibrosis; from 0.718 to 0.673 for detecting advanced fibrosis, shown in the Table S3 and Figure S2). Lastly, the diagnostic accuracy in our study was defined by PPV because a high PPV is crucially important for the use of the non-invasive serum test formula as a diagnostic test. It is well known that PPV is influenced by disease prevalence. The relative lower prevalence of hepatic fibrosis in the external validation cohort (significant fibrosis 56.1% vs 65.1%, advanced fibrosis 30.4% vs 48.5%) could explain, at least in part, the slightly lower PPVs of newly defined FIB-4 thresholds (significant fibrosis 88.4% vs 91.3%, advanced fibrosis 76.5% vs 80.6%).

According to the World Health Organization guideline,⁷ transient elastography is the preferred non-invasive test for assessing liver fibrosis in settings where it is available, and cost is not a major limitation. Based on the sensitivity and specificity (ie, 71.2% and 73.9%) from a large-scale multicenter study,²⁷ the re-calculated

PPV and misdiagnosis rate of transient elastography were 83.6% and 26.1%, respectively, in the same prevalence setting with our discovery cohort (ie, 65.1% for significant fibrosis). It is encouraging that the PPV (91.3%, 95% CI 85.2% ~ 96.2%) and misdiagnosis rate (6.0%, 95% CI 2.7% ~ 10.0%) at the newly defined FIB-4 threshold (≥ 2.25) for diagnosing significant fibrosis were comparable with those of transient elastography. Moreover, when extrapolating the newly defined FIB-4 threshold (≥ 2.25) to the external validation cohorts, the PPV and misdiagnosis rate did not deteriorate significantly (PPV changes from 91.3% to 88.4%; misdiagnosis rate changes from 6.0% to 6.8%). Our research was based on multicenter data from five tertiary hospitals, including patients with different prevalence of hepatic fibrosis, with different transaminase levels, and at different treatment statuses (treatment-naïve vs treatment-experienced). The substantial differences between our discovery and external validation cohorts provide persuasive evidence that the performance of newly defined FIB-4 thresholds was reproducible and reliable. It has become a consensus that CHB

patients with moderate-to-severe liver fibrosis should start antiviral treatment.^{4,5,28} The excellent PPV for diagnosing significant fibrosis implies that about one-quarter of treat-naïve adult CHB patients could avoid liver biopsy and initiate antiviral treatment, due to their FIB-4 scores of 2.25 or more. Although the sensitivity (33.7%, 95% CI 27.8%~39.7%) of this newly defined FIB-4 threshold was not very high, its excellent and reproducible diagnostic accuracy makes it of great value in clinical application, especially for resource-limited areas.

In various clinical guidelines and expert consensus,³⁻⁷ APRI score with a threshold of 2.00 is recommended to detect cirrhosis despite inadequate evidence for clinical diagnostic accuracy in CHB patients. Based on the data from a previous meta-analysis,²⁹ not only was the summary sensitivity of APRI score at the cutoff of 2.00 for detecting cirrhosis low (31.3%) but also the estimated PPV was only 38.7% (ie, 61.3% of the patients ruled in as having cirrhosis were misclassified). In the current study, the best PPVs and corresponding sensitivities for ruling in cirrhosis were only about 60% and 30% with FIB-4 score, respectively, and only about 50% and 40% with APRI score, respectively (data not shown). In contrast, conventional sonography of the upper abdomen, a cheap and easily accessible examination in China, could diagnose cirrhosis with higher sensitivities (77.8%-91.1%) and PPVs (77.8%-82.5%, re-calculated based on the same prevalence of 25.2% as our discovery cohort).^{30,31} Obviously, both FIB-4 and APRI scores were not ideal non-invasive alternatives to liver biopsy for diagnosing cirrhosis and not suitable for use in daily clinical practice.

The present study was performed and reported strictly following the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement,²¹ but still has some limitations. First, although the hepatic histopathological evaluation was performed by at least two pathologists blindly, differences in hepatic fibrosis scoring between pathologists of different hospitals may remain, which might induce inter-observer bias and thus influence the reliability of our results. Second, the liver biopsy sample length of some patients included in this study was lower than 15 mm. Fortunately, the sensitivity analysis suggested that the liver biopsy length did not significantly influence the evaluation of liver fibrosis or the discriminative ability of the non-invasive serum tests for staging liver fibrosis in this study (shown in the Table S1 and Table S2). Third, although the excellent performance of newly defined thresholds was supported by an internal validation test, their performance became slightly worse in the external validation test. However, we have reason to believe that the treatment-experienced patients included in the external validation dataset have influenced the results, and the newly defined FIB-4 thresholds are robust in treatment-naïve patients. Fourth, performing subgroup analyses based on age and ALT level may be interesting but were infeasible in the current study because the raw data of these variables from Huai'an dataset were not available. We hope that more researchers could share results and raw data to further validate the validity of our proposed thresholds, which will allow better use of the inexpensive and more accessible serum test formulas in daily clinical practice.

In conclusion, for staging hepatic fibrosis in CHB patients, the newly defined thresholds of FIB-4 score exhibited better diagnostic performance than its pre-defined thresholds in clinical practice. A FIB-4 score of 2.25 or more could be used to identify those patients with significant fibrosis accurately, so about one-quarter of treat-naïve adult CHB patients could avoid liver biopsy and initiate antiviral treatment in tertiary care settings.

AUTHOR CONTRIBUTIONS

Huang JA and Liu KC proposed the study. Liu KC and Qin MB performed the research and wrote the article. Liu KC, Tao KL, Liang ZH, Cai FQ, Zhao LB, Qin MB, Peng P, Liu SQ, and Zou J collected and analyzed the data. All authors contributed to the design and interpretation of the results and have approved the final article.

DATA AVAILABILITY STATEMENT

The raw data of Nanning dataset can be found in Supplementary material.

ORCID

Kecheng Liu  <https://orcid.org/0000-0003-2126-7030>

REFERENCES

- Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet*. 2015;386(10003):1546-1555.
- Shiha G, Ibrahim A, Helmy A, et al. Asian-Pacific Association for the Study of the Liver (APASL) consensus guidelines on invasive and non-invasive assessment of hepatic fibrosis: a 2016 update. *Hepatol Int*. 2017;11(1):1-30.
- EASL-ALEH Clinical Practice Guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol*. 2015;63(1):237-264.
- Sarin SK, Kumar M, Lau GK, et al. Asian-Pacific clinical practice guidelines on the management of hepatitis B: a 2015 update. *Hepatol Int*. 2016;10(1):1-98.
- Terrault NA, Bzowej NH, Chang KM, Hwang JP, Jonas MM, Murad MH. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology*. 2016;63(1):261-283.
- [Consensus on the diagnosis and therapy of hepatic fibrosis in]. *Zhonghua Gan Zang Bing Za Zhi*. 2019;27(9):657-667.
- Guidelines for the Prevention, Care and Treatment of Persons with Chronic Hepatitis B Infection*. Geneva, Switzerland: World Health Organization; 2015.
- Kim WR, Berg T, Asselah T, et al. Evaluation of APRI and FIB-4 scoring systems for non-invasive assessment of hepatic fibrosis in chronic hepatitis B patients. *J Hepatol*. 2016;64(4):773-780.
- 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.
- Yin Z, Zou J, Li Q, Chen L. Diagnostic value of FIB-4 for liver fibrosis in patients with hepatitis B: a meta-analysis of diagnostic test. *Oncotarget*. 2017;8(14):22944-22953.
- Wong GL, Wong VW, Choi PC, Chan AW, Chan HL. Development of a non-invasive algorithm with transient elastography (Fibroscan) and serum test formula for advanced liver fibrosis in chronic hepatitis B. *Aliment Pharmacol Ther*. 2010;31(10):1095-1103.

12. Wang W, Zhao X, Li G, et al. Diagnostic thresholds and performance of noninvasive fibrosis scores are limited by age in patients with chronic hepatitis B. *J Med Virol*. 2019;91(7):1279-1287.
13. Sonneveld MJ, Brouwer WP, Chan HL, et al. Optimisation of the use of APRI and FIB-4 to rule out cirrhosis in patients with chronic hepatitis B: results from the SONIC-B study. *Lancet Gastroenterol Hepatol*. 2019;4(7):538-544.
14. Huang R, Wang G, Tian C, et al. Gamma-glutamyl-transpeptidase to platelet ratio is not superior to APRI, FIB-4 and RPR for diagnosing liver fibrosis in CHB patients in China. *Sci Rep*. 2017;7(1):8543.
15. Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine*. 2018;35:124-132.
16. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317-1325.
17. Wai CT, Greenson JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518-526.
18. Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. *J Hepatol*. 1991;13(3):372-374.
19. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*. 1996;24(2):289-293.
20. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56:1129-1135.
21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2014;350:g7594.
22. Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res*. 1999;8:113-134.
23. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem*. 2008;54(4):729-737.
24. Basar O, Yimaz B, Ekiz F, et al. Non-invasive tests in prediction of liver fibrosis in chronic hepatitis B and comparison with post-antiviral treatment results. *Clin Res Hepatol Gastroenterol*. 2013;37(2):152-158.
25. Dong XQ, Wu Z, Zhao H, Wang GQ. Evaluation and comparison of thirty noninvasive models for diagnosing liver fibrosis in chinese hepatitis B patients. *J Viral Hepat*. 2019;26(2):297-307.
26. Wang L, Fan YX, Dou XG. Declining diagnostic accuracy of non-invasive fibrosis tests is associated with elevated alanine aminotransferase in chronic hepatitis B. *World J Clin Cases*. 2018;6(12):521-530.
27. Seo YS, Kim MY, Kim SU, et al. Accuracy of transient elastography in assessing liver fibrosis in chronic viral hepatitis: a multicentre, retrospective study. *Liver Int*. 2015;35(10):2246-2255.
28. Lampertico P, Agarwal K, Berg T, et al. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol*. 2017;67(2):370-398.
29. 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.
30. Hung CH, Lu SN, Wang JH, et al. Correlation between ultrasonographic and pathologic diagnoses of hepatitis B and C virus-related cirrhosis. *J Gastroenterol*. 2003;38(2):153-157.
31. Simonovsky V. The diagnosis of cirrhosis by high resolution ultrasound of the liver surface. *Br J Radiol*. 1999;72(853):29-34.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Liu K, Qin M, Tao K, et al.

Identification and external validation of the optimal FIB-4 and APRI thresholds for ruling in chronic hepatitis B related liver fibrosis in tertiary care settings. *J Clin Lab Anal*. 2021;35:e23640. <https://doi.org/10.1002/jcla.23640>