



Precise fibrosis staging with shear wave elastography in chronic hepatitis B depends on liver inflammation and steatosis

Junzhao Ye¹ · Wei Wang² · Shiting Feng³ · Yang Huang² · Xianhua Liao¹ · Ming Kuang⁵ · Xiaoyan Xie² · Bing Liao⁴ · Bihui Zhong¹

Received: 28 October 2019 / Accepted: 19 January 2020 / Published online: 20 February 2020
© Asian Pacific Association for the Study of the Liver 2020

Abstract

Background Two-dimensional shear wave elastography (2D-SWE) is the latest generation of ultrasound elastography for the non-invasive assessment of liver fibrosis in chronic hepatitis B (CHB). We aimed to identify confounders of 2D-SWE in fibrosis grading.

Methods A prospective cohort of 440 CHB patients (286 with liver biopsy and 154 with clinical decompensated cirrhosis) was consecutively enrolled from a clinical trial (registration number: ChiCTR-D-15006000) aimed at optimizing 2D-SWE assessments from 2015 to 2018. All patients underwent 2D-SWE examination, anthropometric measurement, and serum biomarker assessment. Steatosis was graded by the magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF).

Results Overall, the prevalence of incorrect fibrosis staging by 2D-SWE was 26.1% ($n = 115$), with 43.5% of patients under-staged and 56.5% over-staged. In multivariate analysis, the steatosis degree was an independent predictor of 2D-SWE discordance in the overall cohort, with moderate–severe steatosis for underestimation (odds ratio, [OR]=4.3, 95% confidence interval [CI] 1.2–18.2, $p = 0.049$) and overestimation (OR = 8.2, 95% CI 2.9–23.5, $p < 0.001$), and mild steatosis for overestimation (OR = 3.7, 95% CI 1.5–9.0, $p = 0.004$). In patients with liver biopsy, both histological inflammation activity over 2 (OR = 5.0, 95% CI 2.0–25.3, $p = 0.048$) and moderate–severe steatosis (OR = 5.2, 95% CI 2.1–13.4, $p < 0.001$) were independent factors associated with discordance. For the risk of 2D-SWE mis-staging, a nomogram that integrated these confounders was established and the area under the receiver operating characteristic curve of the model was 0.861.

Conclusions Steatosis and inflammation activities were confounders for 2D-SWE. The combination of these confounders could predict mis-staging risks of CHB-related fibrosis with 2D-SWE and may be valuable to decision-making on liver biopsy for fibrosis staging.

Keywords 2D-shear wave elastography · Diagnostic accuracy · Inflammation · Steatosis · Chronic hepatitis B · Magnetic resonance imaging-derived proton density fat fraction · Fibrosis staging · Liver biopsy · Predictive model · Liver stiffness measurements · Alanine aminotransferase · Glutamy transferase

Junzhao Ye and Wei Wang contributed equally to the study.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12072-020-10017-1>) contains supplementary material, which is available to authorized users.

✉ Bing Liao
bing1168@163.com

✉ Bihui Zhong
sophiazhong@hotmail.com

Extended author information available on the last page of the article

Introduction

Chronic hepatitis B (CHB) infection remains the predominant causes of liver-related mortality, with approximately 248 million carriers globally [1]. CHB-related fibrosis has been recognized as the hallmark of progression from mild hepatitis to decompensation manifestations [2], the staging of which well correlates with the incidence of portal hypertension, liver failure, and tumorigenesis [3]. Fibrosis stage underestimation may delay antiviral therapy and worsen liver damage, while the overestimation of fibrosis stage may lead to treatment of uncertain benefits and increase the healthcare burden [2, 3]. Although liver biopsy is the gold

standard for fibrosis staging, it is invasive and is accompanied by complications including hemorrhage, infection, and pain. Therefore, considering biopsy only for those at a high mis-staging risk of non-invasive assessments would minimize unnecessary biopsies.

Liver stiffness measurements (LSMs) with ultrasound elastography, which reflect the fibrosis severity based on the velocity of shear waves captured by ultrasound imaging devices, such as transient elastography (TE) or point shear wave elastography (pSWE), and two-dimensional shear wave elastography (2D-SWE) [4], are emerging as noninvasive tools for fibrosis detection. These technologies have been extensively recommended by the guidelines of CHB management due to their painlessness, diagnostic performance, and convenience of operation [2, 5]. However, correlations between fibrosis severity and LSM with TE or pSWE may decline when hepatic steatosis, obesity, abnormal aminotransferase, hyperbilirubinemia, or ascites are present [4]. These confounding factors have hindered elastography from becoming an ideal alternative to liver biopsy for fibrosis assessments.

As the latest generation ultrasound elastography [4], 2D-SWE possesses higher reproducibility, accuracy [6–8], and the four advantages transcending the aforementioned shortcomings of TE and pSWE: (a) real-time visualization of the fibrosis distribution; (b) an acquisition time within a few milliseconds to minimize interruptions from patient or operator movement; (c) a larger detection region (up to 700 mm²); (d) the generation of a shear wave directly from the tissue, enabling reliable LSMs from patients with ascites or obesity. All of these improvements were devised to obtain LSM that were more indicative of the fibrosis severity. However, the confounding factors of 2D-SWE require validation [9–11].

In this study, we aimed to investigate the factors that are linked to the diagnostic performance of 2D-SWE in patients with CHB. Furthermore, we integrated these factors to predict the accuracy of 2D-SWE for fibrosis staging, which may improve risk stratification for patients who might benefit from liver biopsy.

Patients and methods

Patients and study design

This was a prospective study (registered in the Chinese Clinical Trial Register [ChiCTR-DCD-15006000]) that included consecutive CHB patients with and without nonalcoholic fatty liver diseases in the First Affiliated Hospital, Sun Yat-sen University between February 2015 and December 2018. The study was approved by the institutional ethics committee and informed consent was obtained from all patients. All

authors approved the final version of the article, including the authorship list.

The inclusion criteria included (a) HBV DNA positivity in the blood serum for over 6 months; (b) decompensated cirrhosis; (c) antiviral treatment-naïve patients. The exclusion criteria were as follows: (a) another hepatitis virus coinfection or autoimmune liver disease, (b) alcohol consumption of > 140 g/week in male or > 70 g/week in female, and (c) the presence of severe extrahepatic diseases or pregnancy.

Clinically decompensated cirrhosis was diagnosed by the radiologic evidence (computed tomography or magnetic resonance imaging consistent with cirrhosis) and at least one feature of ascites, variceal bleeding, hepatopulmonary syndrome, hepatorenal syndrome, and encephalopathy [2].

Clinical and laboratory examinations

Patient demographics, alcohol consumption, history of HBV infection, and medications were collected. Weight, height, waist circumference, and hip circumference were measured. The waist-to-hip ratio was calculated as the waist circumference/hip circumference, and the body mass index (BMI) was calculated as weight (kg)/height² (m²). Serum samples were collected after the patients had fasted overnight (8 h) for the following measurements: alanine aminotransferase (ALT), aspartate transaminase (AST), glutamyl transferase (GGT), alkaline phosphatase, total bilirubin, direct bilirubin, albumin, and prothrombin time, platelet, HBV DNA, HBsAg (hepatitis B surface antigen), HBeAg (hepatitis B envelope antigen), total cholesterol, triglycerides, low-density lipoprotein cholesterol, serum glucose, insulin levels, and serum uric acid. The homeostatic model assessment of insulin resistance (HOMA-IR) was also calculated [12].

Magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF) assessment for ultrasonographic steatosis

Conventional abdominal ultrasounds were performed for all subjects with the same US machine as 2D-SWE (SuperSonic Imagine, Aix-en-Provence, France) by two fixed ultrasound physicians who had over 5 years of experience with ultrasound in NAFLD and 3 years' experience of 2D-SWE and were blinded to the aim of the study. Patients fasted for at least 8 h and were required to remain in a supine position while maximally abducting the right arm during the examination. Fatty liver on ultrasound was defined as the presence of liver and kidney echo discrepancy, with or without the presence of posterior attenuation of ultrasound beam, vessel blurring, difficult visualization of the gallbladder wall, and difficult visualization of the diaphragm.

Within 2 weeks of ultrasound steatosis assessment, the liver fat content was assessed using MRI-PDFF with DIXON-fat–water-separation MR imaging (SIEMENS 3.0 T MAGNETOM Verio) at 3.0 T, which was decided as a protocol for all ultrasonographic steatosis patients. The scanning parameters were [13] TE1 2.5 ms, TE2 3.7 ms, repetition time 5.47 ms, 5° flip angle, ± 504.0 kHz per pixel receiver bandwidth, and a slice thickness of 3.0 mm. The fat content was calculated in an irregular-shaped region of interest (ROI) covering the liver in 21 consecutive slices (max-area centered) for each patient placed by two trained radiologists who were blinded to the aim of this study. The liver fat content fraction was categorized as mild (5–10%) or moderate–severe ($\geq 10\%$) that were validated in previous clinical trials on NAFLD [14].

2D-SWE measurements

Right after abdominal ultrasound examinations, 2D-SWE examinations were conducted using a Supersonic Imagine system (Aix-en-Provence, France) by the same operator and with the same transducer. We chose a right intercostal approach for collecting images. SWE was performed in the dual mode. The target area of the liver was chosen under the guidance of conventional, real-time B-mode imaging. Then the patient was asked to hold breath for approximately 5 s after quiet breathing. A rectangular electronic ROI (approximately $4 \times 3 \times 3$ cm and set 1–2 cm under the liver surface) was displayed on the best static SWE image, in which a circular ROI (the diameter set about 2 cm) was set for analysis (Supplementary Fig. 1). Then the LSM means, minimum, maximum, and standard deviation (SD) were calculated. Special attention was paid to avoid any focal lesion, vessels, biliary tracts, or artifacts from nearby lung gas or cardiac movement. The mean value was considered representative of the LSM for each 2D-SWE image. Five consecutive 2D-SWE images were obtained for each patient, and the average values used for analysis. Measurement failure was defined as obtaining no color-coded elasticity images after five trials.

Histological assessment

An ultrasound-guided percutaneous liver biopsy was performed with an 18-gauge needle (Bard, USA) soon after 2D-SWE examination. Biopsy specimens with a length of at least 20 mm were obtained and then processed via formalin fixation, paraffin embedding, hematoxylin–eosin staining, and Masson staining. Two experienced liver pathologists blinded to the laboratory results and LSMs with 2D-SWE evaluated the specimens with at least ten portal tracts using METAVIR scores [15]. Any disagreement in fibrosis or

inflammation staging was re-evaluated and a third pathologist participated in the discussion to achieve a final consensus. Clinically compensated cirrhosis was categorized as F4 in METAVIR stages. Steatosis was semi-quantitatively scored with the ratio of hepatocytes containing visible macrovesicular lipid droplets, from S0 (<5%); S1 (5–32%); S2 (33–66%) to S3 (> 66%).

Statistical analyses

Variables are reported as the means with SD, medians with interquartile ranges, or relative frequencies. Student's *t* tests, Mann–Whitney *U* tests, and χ^2 tests were performed for the comparative analysis. Receiver operating characteristic (ROC) curves were generated using Bootstrap resampling (times = 500) to identify the optimal cut-off values for discriminating fibrosis stages as \geq F2 (significant fibrosis), \geq F3 (advanced fibrosis), and F4 (cirrhosis). The sensitivity, specificity, positive predictive value, negative predictive value, areas under the curve (AUCs), and Obuchowski index (weighted by the relative proportion of the fibrosis stages) [16] were utilized to evaluate the performance of 2D-SWE using hepatic pathology and decompensated cirrhosis as standards. Comparisons between the AUCs were performed with the DeLong test [17]. To evaluate variables influencing the staging accuracy of 2D-SWE, multivariate backward stepwise logistic regression analysis was performed with 2D-SWE misdiagnosis defined as a discordance of at least one stage for fibrosis in the METAVIR scoring system. The performance of single confounders and their combination for predicting 2D-SWE staging misdiagnosis was evaluated by AUC.

All statistical analyses were performed using R statistics version 3.2.5. (The R Project for Statistical Computing, Vienna, Austria) and a two-sided value of $p < 0.05$ was considered significant.

Results

Patient characteristics

A total of 472 eligible individuals were enrolled, of whom 6 patients were excluded for 2D-SWE failure and 26 patients were excluded because of inadequate data. Ultimately, 440 patients were in the final analysis (Table 1 summarizes the characteristics). The majority (70.7%) of the patients were male, and the mean age was 44 years. For the 85 patients with biopsy-proven liver steatosis, 76 patients (89.4%) were detected with abdominal ultrasound and all were confirmed by MRI-PDFF. The accordance rate of steatosis staging differences between biopsy-proven

Table 1 Demographic, virologic, metabolic, and histological characteristics in the overall cohort

	All patients (<i>n</i> = 440)	Concordance (<i>n</i> = 325)	Discordance (<i>n</i> = 115)	<i>p</i>
Demographics				
Age, years	43.8 ± 14.1	42.7 ± 14.0	44.5 ± 13.6	0.23
Male, <i>n</i> (%)	311 (70.7)	235 (72.3)	76 (66.1)	0.21
Anthropometry				
Body mass index (kg/m ²)	23.0 ± 3.5	22.9 ± 3.4	23.2 ± 4.0	0.51
Waist–hip ratio	0.9 (0.8–0.9)	0.9 (0.8–0.9)	0.8 (0.8–0.9)	0.83
Liver biochemistry				
ALT (U/L)	35.0 (23.0–50.0)	34.0 (22.0–48.0)	38.0 (24.8–56.8)	0.51
AST (U/L)	34.0 (26.0–53.0)	34.0 (26.0–53.0)	33.5 (26.8–46.8)	0.77
GGT (U/L)	33.0 (21.0–75.8)	33.0 (20.2–78.2)	30.0 (20.0–66.0)	0.28
ALP (U/L)	89.9 ± 45.6	87.3 ± 36.0	88.8 ± 53.9	0.60
Albumin (mg/L)	45.3 ± 31.2	46.5 ± 36.9	43.8 ± 4.6	0.20
Globulin (mg/L)	31.3 ± 13.6	30.7 ± 6.4	32.3 ± 24.9	0.78
Tbil (μmol/L)	20.2 ± 18.8	19.6 ± 13.4	18.3 ± 10.7	0.35
Dbil (μmol/L)	3.6 (2.4–6.0)	3.6 (2.4–5.9)	3.3 (2.5–5.5)	0.23
Platelet (× 10 ⁹ /L)	174.2 ± 86.7	172.2 ± 88.8	187.0 ± 76.8	0.12
HBV virology				
HBV DNA (Log ₁₀ IU/mL)	4.5 ± 1.4	4.6 ± 2.2	4.0 ± 2.1	0.01
HBeAg positive	153 (34.7)	119 (36.6)	34 (29.6)	0.21
HBsAg (Log ₁₀ IU/mL)	2.9 ± 1.3	3.0 ± 1.3	2.5 ± 1.4	0.003
AFP (ng/mL)	3.4 (2.3–7.2)	3.5 (2.4–7.9)	3.2 (2.3–5.1)	0.11
Metabolic characteristics				
Fasting glucose (mmol/L)	5.2 ± 1.5	5.1 ± 1.2	5.4 ± 2.2	0.44
FINS (μUI/L)	6.9 (4.6–10.5)	7.0 (4.7–10.5)	6.7 (4.4–9.9)	0.54
HOMA-IR	1.5 (0.9–2.5)	1.5 (1.0–2.5)	1.5 (0.8–2.2)	0.78
Cholesterol (mmol/L)	4.8 ± 1.1	4.7 ± 1.1	4.9 ± 1.2	0.22
Triglycerides (mmol/L)	1.2 ± 0.7	1.2 ± 0.7	1.4 ± 0.9	0.02
HDL (mmol/L)	1.4 (1.1–1.7)	1.4 (1.1–1.7)	1.3 (1.1–1.7)	0.86
LDL (mmol/L)	2.8 (2.1–3.4)	2.8 (2.2–3.4)	2.8 (2.0–3.5)	0.59
Uric acid (mmol/L)	340.4 ± 98.1	338.3 ± 96.5	346.9 ± 99.9	0.43
Ultrasonographic fatty liver, <i>n</i> (%)	92(20.9)	54 (16.6)	38 (33.0)	<0.001
Decompensated cirrhosis, <i>n</i> (%)	154 (35.0)	136 (41.8)	18 (15.6)	<0.001
Fibrosis stage^a				
F0–1, <i>n</i> (%)	186(40.9)	159 (41.5)	27 (39.1)	<0.001
F2, <i>n</i> (%)	56 (12.7)	13 (11.4)	43 (16.5)	
F3, <i>n</i> (%)	19 (8.6)	8 (2.5)	11 (26.1)	
F4, <i>n</i> (%)	179 (37.7)	144 (44.6)	35 (18.3)	
Inflammation^b				
	<i>n</i> = 286	<i>n</i> = 199	<i>n</i> = 87	<0.001
A0, <i>n</i> (%)	13 (4.5)	11 (5.5)	2 (2.3)	
A1, <i>n</i> (%)	173 (60.5)	141 (70.9)	32 (36.8)	
A2, <i>n</i> (%)	86 (30.1)	40 (20.1)	46 (52.9)	
A3, <i>n</i> (%)	14 (4.9)	7 (3.5)	7 (8.0)	
Steatosis^b				
	<i>n</i> = 286	<i>n</i> = 199	<i>n</i> = 87	
S0, <i>n</i> (%)	201 (70.3)	155 (77.9)	46 (52.9)	
S1, <i>n</i> (%)	53 (18.5)	31 (15.6)	22 (25.3)	
S2, <i>n</i> (%)	22 (7.7)	10 (5.0)	12 (13.8)	
S3, <i>n</i> (%)	10 (3.5)	3 (1.5)	7 (8.0)	
2D-SWE (kpa)	8.0 (5.8, 14.0)	7.8 (5.4, 15.3)	8.1 (6.8, 9.8)	0.68

Normally and non-normally distributed variables were expressed as mean ± standard deviation and median (25–75% quantiles), respectively
 ALT alanine aminotransferase, AST aspartate transaminase, GGT γ-glutamyl-transferase, ALP alkaline phosphatase, Tbil total bilirubin, Dbil

Table 1 (continued)

direct bilirubin, *HBV* hepatitis B virus, *DNA* deoxyribonucleic acid, *HBsAg* hepatitis B virus surface antigen, *AFP* alpha fetal protein, *FBG* fasting blood glucose, *FINS* fasting insulin, *HOMA-IR* homeostatic model assessment of insulin resistance, *HDL* high-density lipoprotein, *LDL* low-density lipoprotein, *2D-SWE* two-dimensional shear wave electrography liver stiffness measurement

^aFibrosis was calculated as a range of 0–4 according to Metavir system. The distribution of fibrosis stages was compared in patients with or without mis-staging fibrosis by 2D-SWE using the Chi-square test

^bInflammation was calculated as a range of 0–3 according to Metavir system in a biopsy cohort of 286 patients. Steatosis was S scored with the ratio of hepatocytes containing visible macrovesicular steatosis, from S0 to S3. The proportion across inflammation activities or steatosis was compared between patients with and without mis-staging fibrosis by 2D-SWE using the Chi-square test

liver steatosis and MRI-PDFF achieved 97.4% (74/76). In patients with decompensated cirrhosis, 33.8% and 47.4% of cases were with ascites and recent variceal bleeding, respectively. There was no significant mean LSM value between those with or without ascites (16.8 ± 7.3 kPa vs 15.7 ± 6.6 kPa, $p = 0.37$), nor patients with or without recent variceal bleeding (17.1 ± 9.4 vs 14.9 ± 6.8 kPa, $p = 0.10$). Using biopsy-proven liver fibrosis and

decompensated cirrhosis as a reference, 2D-SWE with 7.8 kPa (AUC = 0.919, 95% CI 0.88–0.94, $p < 0.001$, Obuchowski index = 0.79, 95% CI 0.73–0.84, $p < 0.001$) was identified as the best cutoff value for predicting significant fibrosis ($\geq F2$), with 8.6 kPa [AUC = 0.896, 95% CI 0.87–0.93, $p < 0.001$, Obuchowski index = 0.83, 95% CI 0.75–0.90, $p < 0.001$] for severe fibrosis ($\geq F3$), and 10.1 kPa (AUC = 0.900, 95% CI 0.87–0.93, $p < 0.001$,

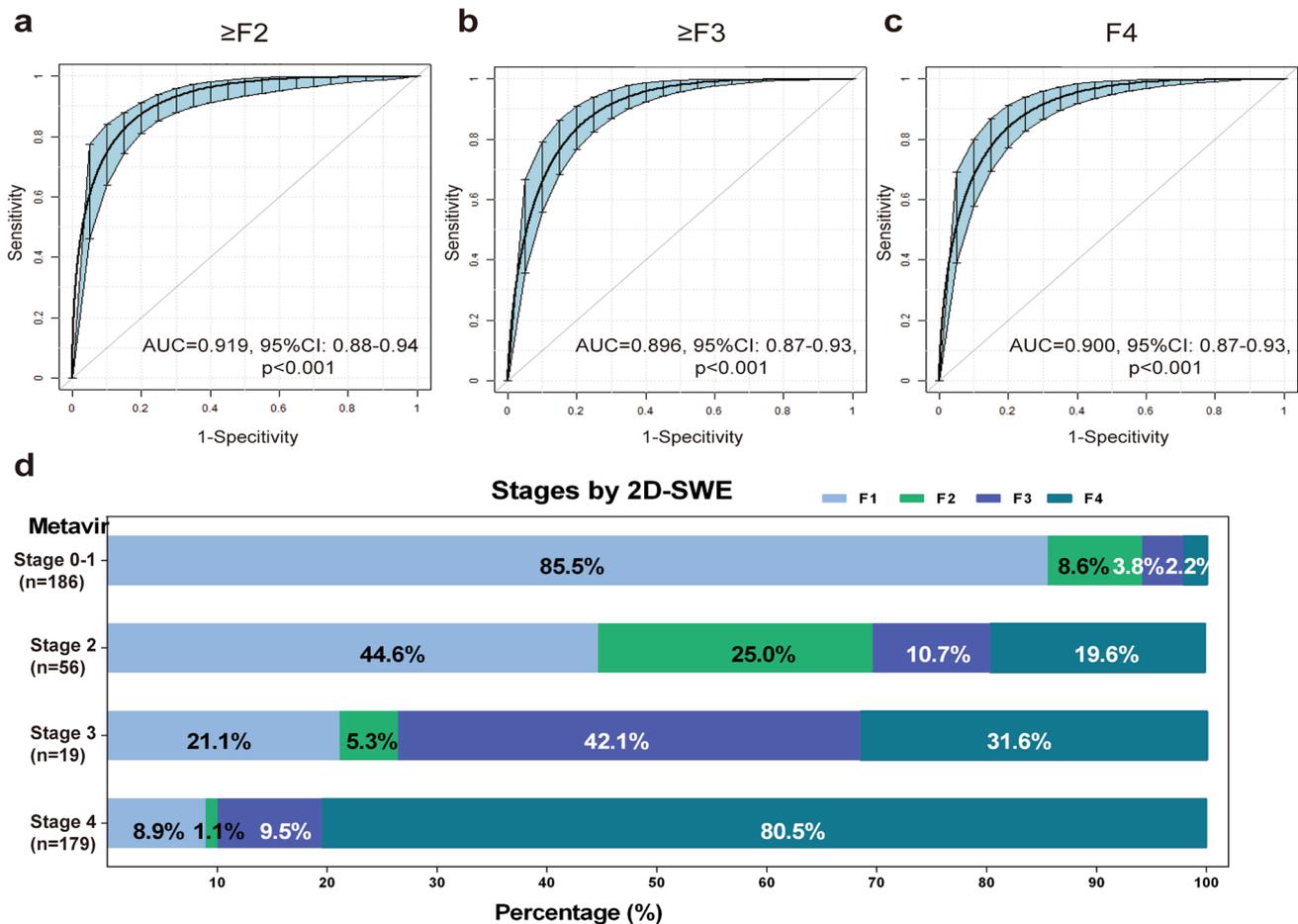


Fig. 1 Accuracy of 2D-SWE. ROC curves using bootstrap resampling (times = 500) for 2D-SWE for discriminating fibrosis stages as follows: **a** $\geq F2$ (significant fibrosis), **b** $\geq F3$ (advanced fibrosis), and **c** F4 (cirrhosis). Green shading shows the bootstrap estimated 95%

CI with the AUC. AUC, area under the curve. **d** Distribution of discordant stages based on 2D-SWE and pathologic scores. The fibrosis stages evaluated by 2D-SWE or pathologic stages was compared in CHB patients using the Chi-square test

Table 2 Logistic multivariable regression analysis of the risk of 2D-SWE staging discordance in all patients (N = 440)

Variables	Mis-staging		Over-staging		Under-staging	
	Univariate OR (95% CI), p	Multivariate OR (95% CI), p	Univariate OR (95% CI), p	Multivariate OR (95% CI), p	Univariate OR (95% CI), p	Multivariate OR (95% CI), p
Age, years	1.0 (0.9, 1.0), 0.23		1.0 (0.8, 1.0), 0.64		1.0 (1.0, 1.1), 0.09	1.0 (1.0, 1.1), 0.35
Female	1.3 (0.8–2.1), 0.21		1.7 (1.0–3.2), 0.07	2.1 (1.0–4.3), 0.06	1.3 (0.7–2.2), 0.43	
BMI > 25 kg/m ²	1.4 (0.9–2.1), 0.21		1.4 (0.8–2.6), 0.25		1.5 (0.8–2.6), 0.19	
WHR > 0.9	1.0 (0.6–1.5), 0.95		1.5 (0.9–2.8), 0.15		0.7 (0.4–1.3), 0.32	
ALT ^a						
< 1 ULN	Reference	Reference	Reference	Reference	Reference	Reference
1–3 ULN	1.3 (0.8–2.1), 0.28	1.3 (0.7–2.2), 0.36	1.1 (0.6–2.3), 0.70	1.0 (0.4–2.2), 0.93	1.3 (0.7–2.3), 0.44	
3–5 ULN	1.0 (0.4–2.3), 0.92	1.0 (0.4–2.5), 0.72	1.5 (0.5–4.1), 0.48	1.3 (0.4–4.1), 0.66	0.5 (0.1–2.0), 0.30	
≥ 5 ULN	2.1 (0.9–4.8), 0.09	1.7 (0.6–4.6), 0.29	3.0 (1.1–7.9), 0.03	3.8 (1.3–11.7), 0.02	1.0 (0.3–3.5), 0.96	
AST > 2 ULN	1.2 (0.5–2.5), 0.68		1.7 (0.7–4.1), 0.24		0.6 (0.2–2.0), 0.40	
GGT > 1 ULN	0.7 (0.4–1.29), 0.26		1.1 (0.6–2.1), 0.75		0.5 (0.2–1.0), 0.05	0.3 (0.1, 0.8), 0.012
ALP > 1 ULN	1.2 (0.5–2.8), 0.69		2.9 (1.0–6.0), 0.06	2.0 (0.6–6.7), 0.24	0.2 (0.0–1.8), 0.17	
TBIL > 2 ULN	0.4 (0.2–1.0), 0.06	0.6 (0.2–1.5), 0.24	0.6 (0.2–1.8), 0.38		0.2 (0.1–1.0), 0.06	
DBIL > 2 ULN	1.0 (0.1–9.4), 0.98		2.1 (0.2–20.6), 0.52		0.0 (0.0–), 0.99	
HBeAg positive	0.7 (0.5–1.2), 0.19		1.3 (0.7–2.4), 0.33		0.4 (0.2–0.8), 0.01	0.5 (0.2, 1.0) 0.064
HOMA-IR > 3	1.0 (0.5–1.8), 0.92		1.0 (0.4–2.2), 0.94		1.0 (0.5–2.1), 0.94	
CHOL > 5.7 mmol/L	1.7 (1.0–2.9), 0.07	1.6 (0.9–2.7), 0.13	1.7 (0.8–3.5), 0.13		1.9 (1.0–3.7), 0.06	1.4 (1.0–3.7), 0.06
TG > 1.7 mmol/L	1.4 (0.8–2.6), 0.26		1.7 (0.8–3.7), 0.14		1.4 (0.6–2.9), 0.42	
HDL > 1.6 mmol/L	1.2 (0.7–2.2), 0.56		1.8 (0.9–3.8), 0.09	0.7 (0.3–1.7), 0.45	1.0 (0.4–2.1), 0.94	
LDL > 2.6 mmol/L	0.8 (0.5–1.4), 0.45		0.7 (0.3–1.4), 0.33		1.0 (0.5–2.0), 0.95	
Hyperuricemia	1.3 (0.8–2.1), 0.35		1.7 (0.9–3.2), 0.12		1.1 (0.6–2.2), 0.75	
Steatosis degree						
None	Reference	Reference	Reference	Reference	Reference	Reference
Mild	1.9 (1.0–3.5), 0.04	1.6 (0.9–3.2), 0.14	2.9 (1.4–6.1), 0.01	3.7 (1.5–9.0), 0.004	1.2 (0.5–2.7), 0.68	1.8 (0.7–4.3), 0.20
Moderate–severe	5.3 (2.5–11.2), < 0.001	4.4 (2.0–9.8), < 0.001	5.9 (2.5–13.7), < 0.001	8.2 (2.9–23.5), < 0.001	3.7 (1.5–9.1), 0.004	4.3 (1.2–18.2), 0.049

2D-SWE two-dimensional shear-wave elastography, OR odds ratio, 95% CI 95% confidence interval, BMI body mass index, WHR waist-hip ratio, ALT alanine aminotransferase, ULN upper limits of normal, AST aspartate transaminase, GGT γ -glutamyl-transferase, ALP alkaline phosphatase, TBIL total bilirubin, DBIL direct bilirubin, HOMA-IR homeostasis model assessment-insulin resistance, CHOL cholesterol, TG triglyceride, HDL high-density lipoprotein, LDL low-density lipoprotein

^aNormal ALT levels represent 40 U/L for female and male

Obuchowski index = 0.78, 95% CI 0.71–0.85, $p < 0.001$) for cirrhosis (=F4) (Fig. 1a–c, Table 2, Supplementary Table 1).

Staging disagreement of 2D-SWE vs pathologic scores

The anticipated fibrosis staging with 2D-SWE for different METAVIR fibrosis stages is shown in Fig. 1d. The diagnostic discordancy rate of fibrosis staging was 26.1% (115/440 patients), while 10% (44/440) of patients had a discordance of over 2 stages. The majority (56.5%) of discordances were attributed to overstaging with 2D-SWE, and the remainder (43.5%) were attributed to understaging. The concordant group contained higher mean HBV DNA levels (4.6 ± 2.2 vs 4.0 ± 2.1 Log₁₀IU/mL, $p = 0.01$), higher mean HBsAg levels (3.0 ± 1.3 vs 2.5 ± 1.4 Log₁₀IU/mL, $p = 0.003$), and proportionately less ultrasonographic fatty liver (8.3% vs 19.1%, $p < 0.001$) compared with the misdiagnosed group.

Factors associated with incorrect 2D-SWE staging: discordance, underestimation, and overestimation

For discordance, multivariate logistic regression analyses identified the steatosis degree (moderate–severe steatosis, OR = 4.4, 95% CI 2.0–9.8, $p < 0.001$) as an independent factor. With respect to underestimation, moderate–severe steatosis and GGT levels remained significant after multivariate analysis, with ORs of 4.3 (95% CI 1.2–18.2, $p = 0.049$) and 0.3 (95% CI 0.1–0.8, $p = 0.012$), respectively (Table 2). Mild steatosis (OR = 3.7, 95% CI 1.5–9.0, $p = 0.004$) and ALT levels over five times above the upper of limit (ULN) (OR = 3.8, 95% CI 1.3–11.7, $p = 0.02$) were found to predict the overestimation of the fibrosis stages by 2D-SWE.

In the subgroup of patients with liver biopsy ($n = 286$) (Supplementary Table 2), histological inflammation activity over 2 was revealed as an independent factor for discordance (OR = 5.0, 95% CI 2.0–25.3, $p = 0.048$) in multivariate analysis. Moderate–severe steatosis remained predictive for discordance (OR = 5.2, 95% CI 2.1–13.4, $p < 0.001$),

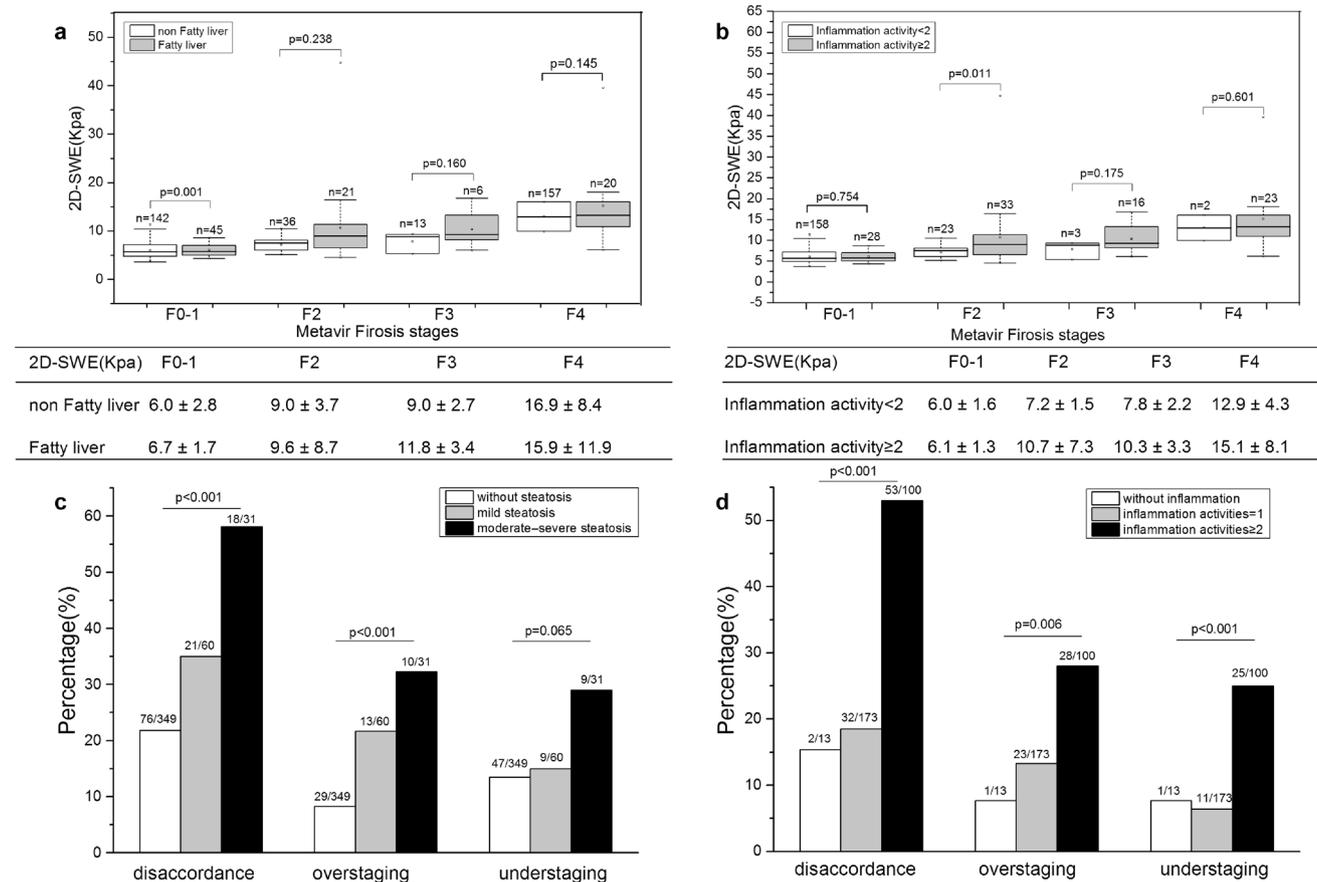


Fig. 2 a Steatosis and **b** inflammation severity and the distribution of 2D-SWE for METAVIR stage. The top and bottom of the box are the 25th and 75th percentiles. The length of the box represents the interquartile range and the median (50th percentile) is the line drawn

through the box. **c** Steatosis, **d** inflammation grade and the prevalence of staging discordance of 2D-SWE. The trend of the proportion of mis-staging fibrosis with 2D-SWE across steatosis and inflammation severity was evaluated using the Chi-square test

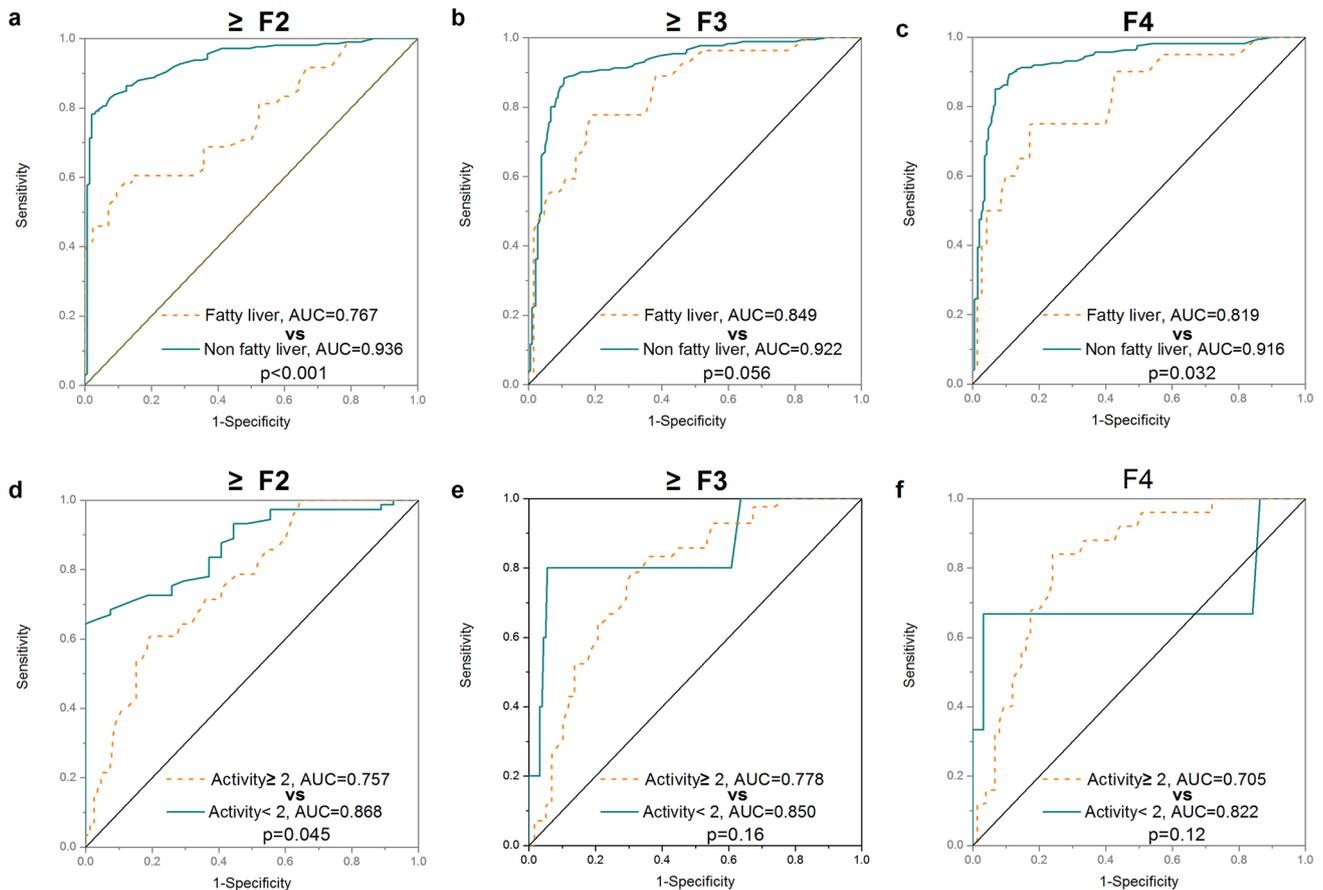


Fig. 3 Impacts of steatosis and inflammation on the diagnostic performance of 2D-SWE for discriminating the different fibrosis stages. For the overall cohort ($n=440$), steatosis and the accuracy of

2D-SWE in discriminating **a** \geq F2, **b** \geq F3, and **c** F4. For the subgroup with biopsy ($n=286$), inflammation activities and the accuracy of 2D-SWE in discriminating **d** \geq F2, **e** \geq F3, and **f** F4

overestimation (OR = 4.6, 95% CI 1.4–14.3, $p=0.009$) or underestimation (OR = 3.8, 95% CI 1.4–10.2, $p=0.0085$).

Impacts of steatosis and inflammation degree on 2D-SWE staging accuracy

In patients with ultrasonographic steatosis ($n=92$), the LSM values were higher for F0-1 (6.7 ± 1.7 vs 6.0 ± 2.8 kPa, $p=0.001$, Fig. 2a). The performance of 2D-SWE for the diagnosis of significant and severe fibrosis was significantly better in nonsteatosis cases than in patients with steatosis by comparison of AUCs (0.936 vs 0.767, $p<0.001$ for \geq F2 and 0.916 vs 0.819, $p=0.032$ for F4, respectively) (Fig. 3a, c, Supplementary Table 1). After quantifying ultrasonographic steatosis with MRI-PDFF, mild steatosis and moderate–severe steatosis patients had higher rates of mis-staging (35.0% and 58.1% vs 21.8%, $p<0.001$), overstaging (21.7% and 32.3% vs 8.3%, $p<0.001$) but not understaging (15.0% and 29.0% vs 13.5%, $p=0.065$) than patients without steatosis (Fig. 2c).

By comparing the LSM of 2D-SWE of different histological inflammation activities in the same fibrosis stage, the subjects with inflammation activities over 2 had a higher mean LSM of 2D-SWE than those groups in F2 group (10.7 ± 7.3 vs 7.2 ± 1.5 kPa, $p=0.011$), while no difference was observed in the other fibrosis stages (all $p>0.05$). In patients with inflammation activities ≥ 2 , lower AUC (0.868 vs 0.757, $p=0.045$) for \geq F2 (Fig. 3d, f) and higher percentages with 2D-SWE mis-staging (53.0% vs 15.4% and 18.5%, $p<0.001$), overstaging (28.0% vs 7.7% and 13.3%, $p=0.006$), and understaging (25.0% vs 7.7% and 6.4%, $p<0.001$) (Fig. 2d) were observed compared with those with inflammation activities = 0 or 1.

Predictive value of confounders for 2D-SWE staging accuracy

For identifying fibrosis mis-staging with 2D-SWE, we built a new model based on impactors that were significantly related to the discordance between 2D-SWE based and pathological fibrosis stages using logistic regression. The combined

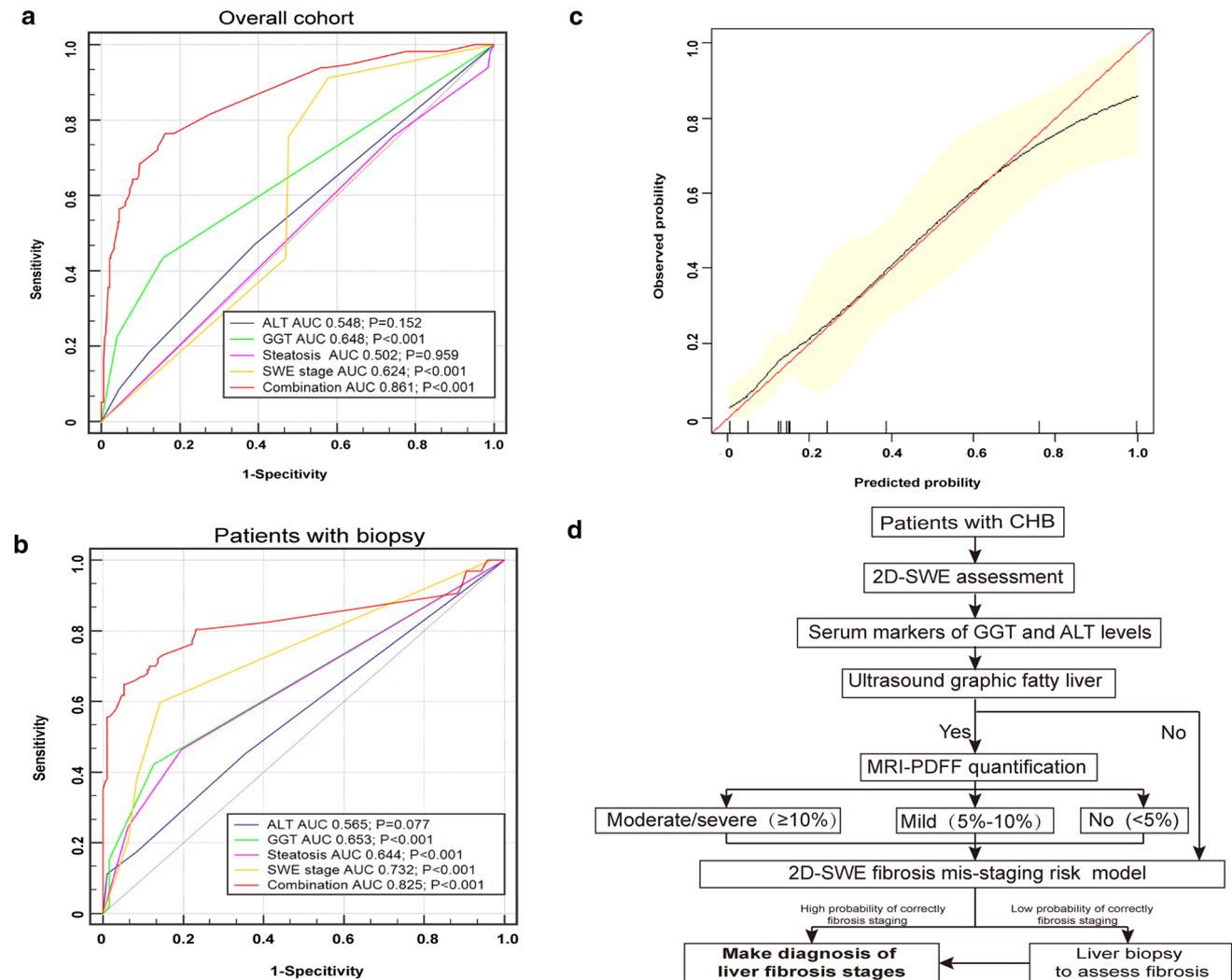


Fig. 4 The diagnostic performances of the confounders for predicting risk of fibrosis mis-staging by 2D-SWE in **a** the overall cohort and **b** the subset with biopsy. **c** Calibration plot of the model (bootstrap resampling times = 500). The *x*-axis plots model-predicted probability of fibrosis mis-staging by 2D-SWE, and the *y*-axis plots observed fibrosis mis-staging by 2D-SWE. The red line represents a perfect predic-

tion by an ideal model. The black line represents the performance of the model, which is a closer fit to the diagonal dotted line, representing improved prediction. The 95% confidence intervals of observed probabilities are plotted as yellow shadows. **d** Diagram shows workflow and confounder-based model for 2D-SWE management of fibrosis staging in chronic hepatitis B

prediction formula was $\sum \text{coefficient}_{\text{impactor}} \times \text{value}_{\text{impactor}}$ (0.06, -1.4, and 0.5 weighted for ALT levels being 1–3, 3–5 and ≥ 5 ULN; 0.8 and 1.8 weighted for mild and moderate–severe steatosis; 1.8 weighted for GGT elevation, and 2.3, 3.2 and 0.3 weighted for SWE predicted of F2, F3 and F4, accordingly) with an AUC of 0.861 (Fig. 4a), significantly higher than single confounders of 2D-SWE predictive fibrosis stages (0.624), degree of steatosis (0.502), ALT (0.548), and GGT levels (0.648) (all $p < 0.001$). The optimal threshold value was 0.45 with 76.5% sensitivity and 83.7% specificity. In patients with biopsy-proven fibrosis, the AUC was 0.825 with 64.9% sensitivity and 94.7% specificity (Supplementary Table 3, Fig. 4b). With 500 times of bootstrap resampling, calibration curves presented visually

good agreement between the predictive model and actual probabilities (Fig. 4c). Over 82% (361 of 440) of patients with CHB could benefit from this model and, therefore, correctly make a diagnosis of fibrosis staging with 2D-SWE and avoid liver biopsy (Fig. 4d).

Discussion

This prospective study explored the associations between a variety of anthropometric, biochemical, metabolic, and pathologic parameters and 2D-SWE staging accuracy. 2D-SWE mis-staging was observed in more than 20% of the cases and markedly correlated with GGT levels, steatosis degree,

and inflammation activities. We also created a noninvasive model combining these determinants to facilitate identifying patients with a higher probability of 2D-SWE mis-staging.

There are growing concerns about the influence of high transaminase levels, cholestasis, ascites, BMI, and steatosis in liver stiffness measurement (LSM) assessed by TE and p-SWE [4]. The prevalence of NAFLD in patients with CHB ranges from 18 to 40% [18], which is similar to that of our results (20.9%). We observed that fatty liver on ultrasound was related to a significant decrease in the AUC of 2D-SWE, independent of the BMI and the waist-to-hip ratio. Moreover, we evaluated the confounding effect of hepatic steatosis by further quantifying with MRI-PDFF, which demonstrated that mild steatosis was associated with overestimation of fibrosis stage, while moderate–severe steatosis was linked to both underestimation and overestimation. These results were consistent with previous findings based on TE that both higher or lower LSM values were associated with severe steatosis [18, 19]. These discrepant results are partially explained, as the accumulation and distribution of fat droplets in the liver may alter with the shear wave velocity, potentially interfering with the 2D-SWE measurement. For patients with moderate to severe steatosis, they have a high prevalence of obesity, with more subcutaneous and prehepatic fat, and thicker abdominal walls [20], which would attenuate ultrasound reflection and interfere with 2D-SWE measurements, as 2D-SWE relies on the ultrasound technique. A study suggested that both the fourth quartiles of abdominal wall thickness and third and fourth quartiles of the non-muscular layer thickness to abdominal wall thickness ratio were significantly associated with increased variation of multiple 2D-SWE measurements (OR = 2.103, 1.753 and 1.695, respectively) [21]. Higher variation in 2D-SWE represented unstable results coexisting with both higher and lower values. We thus speculated that this situation had a non-negligible effect on both over- and under-estimation of 2D-SWE. More researches would be needed to further identify the specific mechanisms.

The bias caused by liver inflammation in 2D-SWE remains unclarified. Multiple existing reports reported that all ALT, AST and GGT levels independently correlated with LSM by 2D-SWE in biopsied patients with CHB using multivariate linear models [22, 23], while another study reported lack of these correlations [24]. However, the mild changes of LSM values within the cut-off values might not alter the fibrosis staging assessment. Our finding of high ALT and GGT levels, which are extensively used to predict the escalation of liver histological necroinflammation, have been consistent with a recent study using liver biopsy as a reference reported that the sensitivity and specificity of 2D-SWE were significantly different at different ALT levels ($< 2 \times \text{ULN}$ and $\geq 2 \times \text{ULN}$), suggesting that ALT levels over $2 \times \text{ULN}$ predicted lower diagnostic efficacy for

2D-SWE [25]. Furthermore, both our and another study did not observe that the grade of liver inflammation affects the performance of LSMs by 2D-SWE in patients with CHB at the F4 stage [26]. These results indicated a significant association between inflammation activities and 2D-SWE incorrect staging and such association should be evaluated considering both ALT, GGT levels, and fibrosis severity.

For patients with normal levels of serum fibrosis markers during chronic infection, it is difficult to make a diagnosis of significant fibrosis according to clinical data and antiviral therapy is not needed unless there is other evidence [26]. Because a superior diagnostic performance of 2D-SWE was achieved for fibrosis assessments than transient elastography or other serum markers [23, 26], 2D-SWE has become valuable in excluding fibrosis non-invasively. However, its accuracy decreases by different degrees when varied confounders coexist in CHB. Therefore, a model incorporating confounders to stratify mis-staging risks may be important to the judgement of significant fibrosis or biopsy, lessening the possibility of delaying antiviral treatment. Our model presented good predicting values by ROC analysis in the overall and the biopsy cohort, with higher sensitivity than using single indices alone. Based on this model, patients with the high risk of 2D-SWE mis-staging should be recommended for biopsy and monitored more frequently.

Steatosis severity served as a key parameter in our nomogram model. Although histologic steatosis staging is the gold standard, it is invasive and the steatosis grading may vary by different liver fat distribution and pathologists [27]. Abdominal ultrasound is the preferred first-line screening method in NAFLD for its inexpensiveness and accessibility, while MRI-PDFF has been emerging as a noninvasive, quantitative, and sensitive measure of the entire liver fat content [13, 14]. For other imaging modalities, such as controlled attenuated parameter (CAP), there is a substantial limitation that it does not represent the whole liver steatosis distribution. The CAP measurements identified patients with biopsy-based hepatic steatosis grade ≥ 2 with an AUC of 0.73 (95% CI 0.64–0.81), which was lower than that identified with MRI-PDFF (AUC = 0.90, 95% CI 0.82–0.97; $p < 0.001$) [28]. Therefore, we use ultrasound as an early screening test to select which patients should receive further MRI-PDFF measurements, which would decrease the number of false-positive results and lessen the burden of subsequent testing.

Several limitations existed in this study. First, inflammatory activity analysis was restricted to biopsy patients. Second, comparisons between steatosis detected among histology, controlled attenuation parameter and MRI-PDFF were not performed. Third, the distribution of fibrosis stages was uneven although we have adopted Obuchowski correction to partly decrease this bias when calculating AUCs. Moreover, the predicting values of

these confounders warranted further validation in a large multicenter cohort.

In this study, steatosis and inflammation were identified as independent predictors of 2D-SWE inaccuracy. Combining these factors demonstrated high diagnostic accuracy for predicting correct 2D-SWE staging, which may assist in selecting candidates for liver biopsy after 2D-SWE assessments.

Funding This study is supported by China foundation for hepatitis prevention and control (TQGB20140083), Special Project on the Integration of Industry, Education and Research of Guangzhou, China (201604020155), Science and Technology Program of Guangdong province, China (2013B021800290, 2014A020212118, 2017A020215015) and National Natural Science Foundation of China (81870404, 81670518, 81170392).

Compliance with ethical standards

Conflict of interest Junzhao Ye, Wei Wang, Shiting Feng, Yang Huang, Xianhua Liao, Ming Kuang, Xiaoyan Xie, Bing Liao, and Bihui Zhong declare that they have no conflict of interest.

Ethical approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Informed consent was obtained from all patients for being included in the study. This study does not contain any studies with animals performed by the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Schweitzer A, Horn J, Mikolajczyk RT. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet* 2015;386:1546–1555
- 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. *Hepatology* 2017;11:1–30
- Wu JF, Song SH, Lee CS, et al. Clinical predictors of liver fibrosis in patients with chronic hepatitis B virus infection from children to adults. *J Infect Dis* 2018;217:1408–1416
- European Association for Study of Liver, Asociacion Latinoamericana para el Estudio del Hgado. EASL-ALEH Clinical Practice Guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015;63:237–264
- European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol* 2017;67:370–398
- Gao Y, Zheng J, Liang P, et al. Liver fibrosis with two-dimensional US shear-wave elastography in participants with chronic hepatitis B: a prospective multicenter study. *Radiology* 2018;289:407–415
- Ferraioli G, Tinelli C, Dal Bello B, et al. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology* 2012;56:2125–2133
- Ferraioli G, Tinelli C, Zicchetti M, et al. Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. *Eur J Radiol* 2012;81:3102–3106
- Suh CH, Kim SY, Kim KW, et al. Determination of normal hepatic elasticity by using real-time shear-wave elastography. *Radiology* 2014;271:895–900
- Huang Z, Zheng J, Zeng J, et al. Normal liver stiffness in healthy adults assessed by real-time shear wave elastography and factors that influence this method. *Ultrasound Med Biol* 2014;40:2549–2555
- Wu T, Wang P, Zhang T, et al. Comparison of two-dimensional shear wave elastography and real-time tissue elastography for assessing liver fibrosis in chronic hepatitis B. *Dig Dis Sci* 2016;34:640–649
- Ye J, Hu X, Wu T, et al. Insulin resistance exhibits varied metabolic abnormalities in nonalcoholic fatty liver disease, chronic hepatitis B and the combination of the two: a cross-sectional study. *Diabetol Metab Syndr* 2019;11:45
- Dong Z, Luo Y, Zhang Z, et al. MR quantification of total liver fat in patients with impaired glucose tolerance and healthy subjects. *PLoS One* 2014;24:e111283
- Ye J, Wu Y, Li F, et al. Effect of orlistat on liver fat content in patients with nonalcoholic fatty liver disease with obesity: assessment using magnetic resonance imaging-derived proton density fat fraction. *Ther Adv Gastroenterol* 2019;12:1756284819879047
- Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289–293
- Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* 2018;63(7):07TR01
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845
- Shen F, Mi YQ, Xu L, et al. Moderate to severe hepatic steatosis leads to overestimation of liver stiffness measurement in chronic hepatitis B patients without significant fibrosis. *Aliment Pharmacol Ther* 2019;50:93–102
- Gaia S, Carezzi S, Barilli AL, et al. Reliability of transient elastography for the detection of fibrosis in non-alcoholic fatty liver disease and chronic viral hepatitis. *J Hepatol* 2011;54:64–71
- Sogabe M, Okahisa T, Hibino S, et al. Usefulness of differentiating metabolic syndrome into visceral fat type and subcutaneous fat type using ultrasonography in Japanese males. *J Gastroenterol* 2012;47:293–299
- Cho YS, Lim S, Kim Y, et al. Abdominal wall thickness affects liver stiffness measurements by 2-D shear wave elastography in patients with chronic liver disease. *Ultrasound Med Biol* 2019;45(10):2697–2703
- Herrmann E, de Lédinghen V, Cassinotto C, et al. Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: an individual patient data-based meta-analysis. *Hepatology* 2018;67:260–272
- Zhuang Y, Ding H, Zhang Y, et al. Two-dimensional shear-wave elastography performance in the noninvasive evaluation of liver fibrosis in patients with chronic hepatitis B: comparison with serum fibrosis indexes. *Radiology* 2017;283(3):873–882
- Liu J, Li Y, Yang X, et al. Comparison of two-dimensional shear wave elastography with nine serum fibrosis indices to assess liver fibrosis in patients with chronic hepatitis B: a prospective cohort study. *Ultraschall Med* 2019;40(2):237–246
- Zeng J, Zheng J, Jin JY, et al. Shear wave elastography for liver fibrosis in chronic hepatitis B: adapting the cut-offs to alanine aminotransferase levels improves accuracy. *Eur Radiol* 2019;29:857–865

26. Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicenter study. *Gut* 2019;68:729–741
27. Bedossa P. Current histological classification of NAFLD: strength and limitations. *Hepatology Int* 2013;7:765–770
28. Imajo K, Kessoku T, Honda Y, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology* 2016;150(3):626–637

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Junzhao Ye¹ · Wei Wang² · Shiting Feng³ · Yang Huang² · Xianhua Liao¹ · Ming Kuang⁵ · Xiaoyan Xie² · Bing Liao⁴ · Bihui Zhong¹ 

¹ Department of Gastroenterology, The First Affiliated Hospital, Sun Yat-Sen University, No. 58 Zhongshan Road II, Yuexiu District, Guangzhou 510080, Guangdong, People's Republic of China

² Department of Ultrasound, The First Affiliated Hospital, Sun Yat-Sen University, Guangdong, People's Republic of China

³ Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, Guangdong, People's Republic of China

⁴ Department of Pathology, The First Affiliated Hospital, Sun Yat-Sen University, Yuexiu District, No. 58 Zhongshan Road II, Guangzhou 510080, Guangdong, People's Republic of China

⁵ Department of Hepatobiliary Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangdong, People's Republic of China