86696_Auto_Edited.docx

Diagnostic role of transient elastography in patients with autoimmune liver diseases: A systematic review and meta-analysis

Non-invasive diagnosis of fibrosis in AILD

Hong Chen, Yue Shen, Sheng-Di Wu, Qin Zhu, Cheng-Zhao Weng, Jun Zhang, Mei-Xia Wang, Wei Jiang

Abstract

BACKGROUND

Noninvasive methods have been developed to detect fibrosis in many liver diseases due to the limits of liver biopsy. However, previous studies have focused primarily on chronic viral hepatitis and nonalcoholic fatty liver disease. The diagnostic value of transient elastography for autoimmune liver diseases (AILDs) is worth studying.

AIM

To compare the diagnostic accuracy of imaging techniques with serum biomarkers of fibrosis in autoimmune liver disease.

METHODS

The PubMed, Cochrane Library and EMBASE databases were searched. Studies evaluating the efficacy of noninvasive methods in the diagnosis of AILDs (autoimmune hepatitis [AIH], primary biliary cholangitis [PBC] and primary sclerosing cholangitis [PSC]) were included. The summary area under the receiver operating characteristic

curve (AUROC), diagnostic odds ratio (DOR), sensitivity and specificity were used to assess the accuracy of these noninvasive methods for staging fibrosis.

RESULTS

A total of 60 articles were included in this study, and the number of patients with AIH, PBC, and PSC was 1594, 3126, and 501, respectively. The summary AUROC of transient elastography (TE) in the diagnosis of significant fibrosis, advanced fibrosis and cirrhosis in patients with AIH were 0.84, 0.88, and 0.90, respectively, while those in patients with PBC were 0.93, 0.93, and 0.91, respectively. The AUROC of cirrhosis for patients with PSC was 0.95. However, other noninvasive indices (aspartate aminotransferase to platelet ratio index [APRI], aspartate aminotransferase/alanine aminotransferase ratio [AAR], fibrosis-4 index [FIB-4]) had corresponding AUROC less than 0.80.

CONCLUSION

TE exerts better diagnostic accuracy in AILD patients, especially in PBC patients. The appropriate cutoff values for staging advanced fibrosis and cirrhosis ranged from 9.6 to 10.7 and 14.4 to 16.9 KPa for PBC patients.

Key Words: Liver stiffness; Serum parameter; Liver fibrosis; Non-invasive diagnosis; Transient elastography; Autoimmune liver disease

Chen H, Shen Y, Wu SD, Zhu Q, Weng CZ, Zhang J, Wang MX, Jiang W. Diagnostic role of transient elastography in patients with autoimmune liver diseases: A systematic review and meta-analysis. *World J Gastroenterol* 2023; In press

Core Tip: Onset of autoimmune liver diseases is frequently insidious, and immune cell infiltration and continuous inflammation drive hepatic fibrosis, which gradually progresses to cirrhosis, causing poorer long-term outcomes. Liver biopsy as the reference standard is an invasive procedure, thus, repeated biopsies are difficult to implement.

Consequently, appropriate noninvasive methods are essential to dynamically monitor the degree of liver fibrosis. Our meta-analysis aims to compare the diagnostic accuracy of imaging techniques with serum biomarkers of fibrosis in autoimmune liver disease.

INTRODUCTION

The incidence of autoimmune liver diseases (AILDs), including autoimmune hepatitis (AIH), primary biliary cholangitis (PBC), primary sclerosing cholangitis (PSC), and multiple overlap syndromes, a group of autoimmune diseases associated with the liver and bile duct is increasing[1-2]. Onset is frequently insidious, with nonspecific symptoms. Immune cell infiltration and continuous inflammation drive hepatic fibrosis, which gradually progresses to cirrhosis, causing poorer long-term outcomes in patients[3-5]. Accordingly, accurate identification of high-risk patients for such conditions is essential in clinical care to guide timely treatment and delay disease progression.

For years, liver biopsy has been recognized as the reference standard for the assessment of liver fibrosis. However, biopsy area restrictions, sampling errors, and interobserver variability may affect the diagnostic accuracy[6-7]. Moreover, because biopsy is an invasive procedure with potentially hazardous complications ranging from pain to more severe events and even death, many patients are reluctant to undergo repeat biopsies[8-9]. Consequently, an increasing number of studies have focused on noninvasive methods to identify the ideal approach for dynamically monitoring the degree of liver fibrosis[10].

In recent years, some noninvasive methods, including biochemical tests and imaging techniques, have been widely developed, including the aspartate aminotransferase to platelet ratio index (APRI), aspartate aminotransferase/alanine aminotransferase (ALT) ratio (AAR), fibrosis-4 index (FIB-4), red cell distribution width to platelet ratio (RPR), Mac-2 binding protein (M2BP), platelet count to spleen diameter (PC/SD) ratio, transient elastography (TE), shear wave elastography (SWE), acoustic radiation force impulse (ARFI), magnetic resonance spectroscopy (MRS), and magnetic resonance elastography (MRE). Previous studies have validated that elastography is a reliable method with a

diagnostic accuracy higher than that of blood tests for staging liver fibrosis in chronic viral hepatitis[11-13], nonalcoholic fatty liver disease (NAFLD),[14] and AIH[15]; however, no studies have explored the diagnostic accuracy of noninvasive methods for the other two types of AILDs (PBC and PSC).

Therefore, the present meta-analysis aimed to compare the diagnostic accuracy of biochemical tests and imaging techniques for detecting liver fibrosis in patients with AILD, determine whether the same noninvasive methods show different diagnostic values in the three types of AILDs, and recommend appropriate cut-off values for different fibrosis stages.

MATERIALS AND METHODS

Literature search strategy

Studies on the diagnosis of autoimmune liver disease published between January 2006 and December 2022 were searched for in PubMed, Cochrane Library, and EMBASE databases using the following keywords: autoimmune hepatitis, primary sclerosing cholangitis, primary biliary cholangitis, liver fibrosis, transient elastography, shear wave elastography, magnetic resonance elastography, APRI, FIB-4, and AAR. The detailed search strategy is presented in **Supplementary Table 1**.

Study selection criteria

Original studies that fulfilled the following criteria were enrolled: 1) studies with patient populations with AIH, PBC, or PSC with discrete data that could be separately extracted from the mixed liver disease study cohort; 2) studies in which liver biopsy was used as the gold standard to assess fibrosis based on the Metavir score or another score that could be converted to the Metavir score; 3) studies assessing the performance and utility of APRI, AAR, FIB-4, RPR, M2BP, ARFI, PC/SD ratio, TE, SWE, MRE, or MRS for staging liver fibrosis; and 4) studies directly reporting the true-positive (TP), false-positive (FP), false-negative (FN) and true-negative (TN) values or provided data by which they could be calculated to allow the construction of a 2 × 2 table for each test.

The following studies were excluded: 1) studies exploring the prognostic value of liver stiffness measurement (LSM) for patients with AILD; 2) animal experiments, reviews, protocols, guidelines, case reports, or meta-analyses; 3) studies on liver fibrosis due to other etiologies, including NAFLD, chronic hepatitis B, or chronic hepatitis C; and 4) studies without sufficient data for further analysis or with the same or overlapping group of participants.

Data extraction and quality assessment

Two investigators (Chen H and Shen Y) independently evaluated the eligibility and quality of the included studies and extracted the data. Any disagreements were resolved by a senior researcher (Wu SD). We collated the following parameters in Microsoft Excel 2010: authors, year of publication, country, study period and design, pathological type, diagnostic methods, sample size, patient characteristics (age, sex, body mass index [BMI], ALT level, treatment condition), quality of liver biopsy, and performance of the index test, including cutoff values, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUROC). Two reviewers (Chen H and Shen Y) independently assessed the risk of bias using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool[16].

Data synthesis and statistical analysis

According to the Metavir, Batts-Ludwig, and Scheuer scores, liver fibrosis was classified into five stages (F0, F1, F2, F3, and F4), whereas there were seven stages according to the Ishak score. Given that Shiha et al. proposed that an Ishak score of 3 corresponds to METAVIR score of F2[17]; significant fibrosis (SF), advanced fibrosis (AF), and cirrhosis were defined as stages F2-F4, F3-F4, and F4, respectively. For the data analysis, we constructed 2 × 2 contingency tables with TP, FP, FN, and TN values based on data directly extracted from the original studies or calculated from indirect variables (sensitivity, specificity, and sample size). A bivariate random-effects model was subsequently applied to calculate the diagnostic test accuracy variables, including

summary sensitivity, summary specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio (DOR) with their associated 95% confidence intervals (CIs). We also performed meta-analyses using hierarchical models to produce summary ROC curves, from which we obtained summary AUROC values to evaluate the diagnostic accuracy of the different noninvasive methods. The method was considered to have excellent accuracy if the summary AUROC value was above 0.90; moderate accuracy, if it was greater than 0.80; and poor accuracy, if it was less than 0.80[18].

The heterogeneity was assessed using multiple methods. Spearman's correlation coefficient was calculated to evaluate the threshold heterogeneity of the included studies, whereas Cochran's Q and I^2 values were used to assess non-threshold heterogeneity. If an I^2 value > 50% or P < 0.05 indicated distinct statistical heterogeneity, a random-effects model was used to combine the data. A fixed-effect model was chosen when the I^2 value ≤ 50% or P ≤ 0.05. However, the number of original studies was sufficient to perform a meta-regression to explore the potential heterogeneity of certain index tests. In addition, we conducted a subgroup analysis according to the sample size, treatment conditions, and cutoff value. Deeks' funnel plots were used to evaluate the possible publication bias. The meta-analysis was performed using Stata 12.0, Reviewer Manager Version 5.3, and Meta-Disc Version 1.4.

RESULTS

Characteristics of the included studies and patients

The study selection process is illustrated in **Figure 1**. A total of 1386 studies were retrieved through our search strategy, of which 427 were excluded as duplicates and 602 were removed following the screening of titles, abstracts, and reviews. The remaining 355 potentially eligible studies were selected for further evaluation. Of these, sixty articles were included in the evaluation and analysis (**Supplementary Table 2**). Among them, 22, 29, and six studies were regarding AIH, PBC, and PSC, respectively (two studies focused on both AIH and PBC[19-20], while one study focused on both PBC and PSC[21]). In total,

they included 11 noninvasive index tests. The basic characteristics of the included studies are presented in **Table 1**. We selected articles published between 2006 and 2022, of which 46 (76.7%) will be published between 2016 and 2022. There were 31 (51.7%) retrospective studies, 17 (28.3%) prospective studies, 10 (16.7%) unknown studies, and two studies with both designs. Most included studies were conducted in Asia (28 studies) or Europe (24 studies). A total of 27 studies utilized the Metavir score, 8 used the Batts-Ludwig score, 17 used the Scheuer score, five used the Ishak score, and three used the Ludwig-Scheuer score.

A total of 1594, 3126, and 501 patients with AIH, PBC, and PSC, respectively, were included to analyze the diagnostic performance of non-invasive methods in staging liver fibrosis. Most patients with AIH and PBC were female (72.4% and 87.6%, respectively). In contrast, patients with PSC were predominantly male (73.7%). The average ages of patients with AIH, PBC, and PSC were 47.0, 55.2, and 41.5 years, respectively. Patients with AIH (160.29 IU/mL) had higher ALT levels than patients with PBC (69.81 IU/mL).

Quality assessment of the included studies

The results of the quality assessment based on the QUADAS-2 scale for all 60 eligible studies are shown in **Figure 2** and **Supplementary Figure 1**. Regarding patient selection, eight studies had an unclear risk of bias owing to the lack of information on whether patients were enrolled randomly or consecutively. Regarding the index test, four studies were determined to have an unclear risk of bias because the results of the index test were interpreted without blinded information on the results of the reference standard. Likewise, 22 studies were regarded as having an unclear risk of bias because the results of the reference standard were interpreted without blinded information regarding the results of the index test. In terms of flow and timing, two studies were considered high-risk because not every subject received a reference standard, while 19 studies were considered unclear risk because of an unknown time interval between the index and reference tests.

Performance of noninvasive methods in diagnosing significant fibrosis (F≥2)

Diagnosis of SF for AIH:

Fifteen studies (*n* = 1001) evaluated eight non-invasive methods for detecting SF in patients with AIH. Of these, five (*n* = 459), two (*n* = 129), five (*n* = 459), nine (*n* = 523), and three (*n* = 234) studies focused on APRI, AAR, FIB-4, TE, and SWE separately, whereas only one study each utilized the ARFI, PC/SD ratio, and RPR.

The APRI had moderate summary sensitivity (exceeding 70%) with poor summary specificity (less than 50%), whereas the FIB-4 had the opposite result (**Table 2**). Interestingly, TE had a relatively greater diagnostic performance than the other laboratory tests, with summary sensitivity and specificity values of 0.82 and 0.73, respectively, and cutoff values ranging from 5.8–7 KPa. The summary sensitivity of SWE (0.89; 95%CI, 0.83–0.93) was significantly higher than that of the other six noninvasive methods and slightly greater than that of TE (0.83; 95%CI, 0.78–0.87).

Diagnosis of SF for PBC:

Thirteen studies (*n* = 1389) evaluated nine non-invasive methods for diagnosing SF in patients with PBC. Among them, four (*n* = 584), three (*n* = 323), three (*n* = 462), two (*n* = 87), five (*n* = 446), and two (*n* = 210) studies focused on APRI, AAR, FIB-4, ARFI, TE, and SWE, respectively; however, only one study each utilized the PC/SD ratio, MRE, and M2BP.

As shown in **Table 2**, the APRI and FIB-4 index had relatively good summary sensitivities of 0.84 and 0.85, respectively, with mild summary specificities of 0.63 and 0.77, respectively; the corresponding values for sensitivity and specificity of the AAR were poor (0.69, 0.56). In contrast, both the summary sensitivity (0.81; 95%CI 0.76–0.85) and specificity (0.95; 95%CI 0.89–0.98) of TE were significantly higher than those of the other five noninvasive methods for predicting SF with cutoff values ranging from 5.9–8.8 KPa. Furthermore, **Table 3** shows that the summary DORs of APRI, FIB-4 and TE were 3.9 (95%CI 2.1–7.3), 5.1 (95%CI 3.1–8.5) and 16.8 (95%CI 8.8–32.2), respectively, in patients with AIH, while the summary DORs of APRI and TE were 6.3 (95%CI 3.5–11.2) and 74.5

(95%CI 12.2–455.5), respectively, in patients with PBC. Additionally, the summary AUROC value of TE in patients with PBC (0.93, 95%CI 0.91–0.95) was relatively higher than that of TE in patients with AIH (0.84, 95%CI 0.80–0.87) but significantly higher than that of FIB-4 (0.74) and APRI (0.67) in patients with AIH and APRI (0.77) in patients with PBC (**Table 3** and **Figure 3)**.

Performance of noninvasive methods in diagnosing advanced fibrosis (F≥3)
Diagnosis of AF for AIH:

Twenty studies (*n* = 1435) evaluated 11 noninvasive methods for detecting AF in patients with AIH. Among them, 10 (*n* = 917), seven (*n* = 532), 11 (*n* = 962), seven (*n* = 460), three (*n* = 234), and two (*n* = 191) studies focused on APRI, AAR, FIB-4, TE, SWE, and RPR, respectively; however, only one study each utilized the ARFI, MRS, PC/SD ratio, MRE, and M2BP methods.

As shown in **Table 2,** with a cutoff value of 8.18–9 KPa, both the summary sensitivity and specificity exceeded 80% when TE was used for predicting AF, whereas with a cutoff value of 10.4–12.1 KPa, there was a better summary specificity (0.93; 95%CI 0.86–0.97) with a mild summary sensitivity (0.73; 95%CI 0.60–0.83). Regarding SWE, MRE, and the PC/SD ratio, the summary sensitivity and specificity also exceeded 80%; the specificity of MRE was 1.00, but only one study assessed it[22]. For AAR and FIB-4 index, there was a relatively modest summary specificity (< 0.80) and poor summary sensitivity (< 0.60).

Diagnosis of AF for PBC:

Twenty-eight studies (*n* = 2737) evaluated 11 noninvasive methods for detecting AF in patients with PBC. Of these, 15 (*n* = 1589), six (*n* = 559), 13 (*n* = 1296), two (*n* = 97), 10 (*n* = 874), five (*n* = 362), four (*n* = 370), and two (*n* = 210) studies were focused on APRI, AAR, FIB-4, ARFI, TE, RPR, M2BP and SWE, respectively; only one study each utilized the methods of MRE and RLR.

As shown in **Table 2,** TE had a good summary sensitivity and specificity (0.91, 0.82) with a cutoff value of 9.6–10.7 KPa, while RPR and M2BP had good summary specificity

(0.89 and 0.80, respectively) with poor summary sensitivity (0.49 and 0.68, respectively). Regardless of the cutoff values, the summary sensitivities and specificities of the AAR, APRI, and FIB-4 were less than 0.80.

Furthermore, **Table 3** shows that the summary DORs of AAR, APRI, FIB-4 and TE were 4.9 (95%CI 3.2–7.8), 3.9 (95%CI 2.8–5.3), 4.0 (95%CI 2.4–6.8) and 25.1 (95%CI 9.7–65.3), respectively, in patients with AIH and 4.1 (95%CI 2.0–8.6), 3.7 (95%CI 2.3–6.0), 7.1 (95%CI 4.0–12.8) and 41.8 (95%CI 19.3–91.0), respectively, in patients with PBC. Moreover, the summary DORs of RPR and M2BP in patients with PBC were 8.0 (95%CI 4.0–15.8) and 13.2 (95%CI 4.1–42.4), respectively. As shown in **Table 3** and **Figure 3**, the summary AUROC value of TE for detecting AF was 0.88 (95%CI 0.85–0.90) and 0.93 (95%CI 0.90–0.95) in patients with AIH and PBC, respectively. The value of M2BP was 0.86 (95%CI 0.82–0.88) in patients with PBC, whereas the summary AUROC values for AAR, APRI, and FIB-4 were less than 0.80 in both patients with AIH and PBC, and the value for RPR in patients with PBC was less than 0.60.

Performance of noninvasive methods in diagnosing cirrhosis (F = 4)

Diagnosis of cirrhosis for AIH:

Sixteen studies (*n* = 1076) evaluated ten noninvasive methods for detecting cirrhosis in patients with AIH. Of these, six (*n* = 543), three (*n* = 213), six (*n* = 543), two (*n* = 82), seven (*n* = 415), and four (*n* = 297) studies focused on APRI, AAR, FIB-4, ARFI, TE, and SWE, respectively; only one study each utilized the PC/SD ratio, MRE, RPR, and M2BP.

As shown in **Table 2**, the summary sensitivities and specificities of APRI and FIB-4 were less than 75%, and those of AAR were 0.61 and 0.83, respectively. Moreover, the summary sensitivity and specificity of TE (cutoff value ranging from 11–12.67 KPa) were significantly higher for predicting cirrhosis, with 0.89 (95%CI 0.82–0.94) and 0.88 (95%CI 0.81–0.93), respectively, while the summary sensitivity and specificity of SWE (0.83, 0.86) were close to those of TE. Surprisingly, the summary specificity dramatically rose to 0.97 (95%CI 0.92–0.99) with a cutoff value ranging from 16–19 KPa.

Diagnosis of cirrhosis for PBC:

Sixteen studies (*n* = 1568) evaluated nine non-invasive methods for detecting cirrhosis in patients with PBC. Among them, six (*n* = 852), four (*n* = 407), six (*n* = 852), six (*n* = 483), two (*n* = 210), and two (*n* = 194) studies focused on APRI, AAR, FIB-4, TE, SWE, and M2BP, respectively. However, only one study utilized the ARFI and MRE methods.

As listed in **Table 2**, the summary sensitivities of APRI, AAR, and FIB-4 for predicting cirrhosis were 0.75, 0.81, and 0.87, respectively, and their corresponding summary specificities were 0.51, 0.77, and 0.61, respectively. In contrast, TE had higher summary sensitivity (0.90; 95%CI 0.74–0.98) and specificity (0.93; 95%CI 0.89–0.96) with a cutoff value ranging from 15.6–25.1 KPa.

Diagnosis of cirrhosis for PSC:

Four studies (*n* = 207) evaluated TE as a predictor of cirrhosis in patients with PSC. Because the diagnosis of PSC does not rely on liver biopsy, few related studies have been conducted. As listed in **Table 2**, the summary sensitivity and specificity of TE were 0.82 (95%CI 0.68–0.91) and 0.89 (95%CI 0.83–0.94), respectively.

Furthermore, **Table 3** shows that the summary DORs of APRI and FIB-4 were 3.8 (95%CI 2.2–6.4) and 5.5 (95%CI 2.4–12.6) in patients with AIH and 14.6 (95%CI 1.9–113.8) and 29.8 (95%CI 5.9–150.3) in patients with PBC. In addition, the summary DOR of TE was highest in patients with PBC, with values of 91.8 (95%CI 40.1–201.2), 134.8 (95%CI 33.0–551.8), and 70.6 (95%CI 15.4–322.7) in patients with AIH, PBC, and PSC, respectively. The summary AUROC values of TE for detecting cirrhosis in patients with AIH, PBC, and PSC were 0.90 (95%CI 0.87–0.92), 0.91 (95%CI 0.88–0.93), and 0.95 (95%CI 0.93–0.97), respectively, while the summary AUROC values for APRI and FIB-4 were less than 0.80 in patients with AIH and 0.90 in patients with PBC (**Table 3**, **Figure 3**).

Methodological heterogeneity, subgroup analysis and publication bias

As shown in **Table 4**, threshold heterogeneity was observed only in APRI F2 in both patients with AIH and PBC, whereas non-threshold heterogeneity was observed in most

groups (**Figure 4**, **Supplementary Figure 2** and **Supplementary Figure 3**). Because meta-regression to explore the source of heterogeneity requires the number of original studies to exceed 10, we only conducted a meta-regression for advanced fibrosis in patients with AIH and PBC. The heterogeneity of APRI, FIB-4, and TE accuracy was mainly affected by the cutoff value with regard to specificity, whereas FIB-4 and TE were affected by sample size with regard to sensitivity, according to the meta-regression analysis (**Supplementary Figure 4**).

Subgroup analyses of TE according to sample size, cutoff value, and treatment status are shown in **Table 5**. Because of the limited data, we only conducted an analysis for posttreatment combined with original data for pretreatment (**Supplementary Table 3**). Deeks' funnel plot of these noninvasive methods was generated to assess publication bias. There was a publication bias for APRI in detecting SF ($P = 0.06$) and cirrhosis ($P = 0.08$) in patients with PBC, but not in other methods for detecting SF, AF, and cirrhosis (**Supplementary Figure 5**). Moreover, no publication bias was observed for any non-invasive method in patients with AIH (**Supplementary Figure 6**).

DISCUSSION

In our review, a total of sixty studies (including 1594, 3126, and 501 patients with AIH, PBC, and PSC, respectively) were included to evaluate the diagnostic accuracy of non-invasive methods for predicting SF, AF, and cirrhosis in patients with AILDs. TE had excellent accuracy with summary AUROC values of 0.93, 0.93, and 0.91 for SF, AF, and cirrhosis, respectively, in patients with PBC, while TE had a moderate to excellent accuracy of 0.84, 0.88, and 0.90, respectively, in patients with AIH. Moreover, the summary AUROC was 0.95 for cirrhosis in patients with PSC. In contrast, other non-invasive methods, such as AAR, APRI, FIB-4, and RPR, had poor accuracy, with summary AUROC values of <0.80. In addition, the pooled sensitivity and specificity of TE were higher than those of the other noninvasive methods. Our results indicated that LSM using TE had a better diagnostic performance for staging hepatic fibrosis in AILDs, especially in patients with PBC. Moreover, our results showed that TE had mostly higher

specificity and relatively low sensitivity for the diagnosis of AILDs. [implying that TE was a better noninvasive method for ruling in than for ruling out.] Koizumi et al[23] found that TE has high sensitivity and relatively low specificity for the diagnosis of PBC. However, the optimal cutoff values were higher and the range was wider than those in other studies, indicating that different optimal cutoff values may have an effect on diagnostic accuracy.

Meta-regression analysis is a reliable method for screening heterogeneity. In our study, the sample size, cutoff values, prevalence of significant fibrosis, and scoring system provided heterogeneity in summarizing the test results, consistent with previous studies[24-25]. We conducted subgroup analyses based on the sample size and cutoff values. Our results revealed that TE had a better predictive effect in a larger sample of patients with PBC. LSM by TE is the best surrogate marker for staging in SF and AF with a cutoff ranging from 6.4–9.1 KPa and 9.0–11.0 KPa, respectively, in patients with AIH and staging in AF and cirrhosis with a cutoff ranging from 9.6–10.7 KPa and 14.4–16.9 KPa, respectively, in patients with PBC.

Several previous studies have demonstrated that inflammation in the liver (reflected by elevated ALT levels)[26] and extrahepatic cholestasis (reflected by total bilirubin)[27] may increase stiffness value, causing a decrease in the diagnostic accuracy of TE, whereas ALT and bilirubin levels decline after treatment. Because a limited number of studies have reported results for the ALT subgroup, we only conducted a subgroup analysis of treatment conditions, which showed that the diagnostic accuracy for staging liver fibrosis was comparable between pre- and post-treatment in patients with both PBC and AIH. In other words, this may indicate that ALT levels have no significant effect on diagnostic accuracy. Meanwhile, two scoring systems (International Autoimmune Hepatitis Group (IAIHG) 1999 and IAIHG 2008) proposed by Granito et al for the diagnosis of AIH are not interchangeable[28]. According to our subgroup analysis regarding diagnostic criteria, the IAIHG 2008 showed diagnostic accuracy comparable to that of the IAIHG 1999 in distinguishing patients with AIH (**Supplementary Figure 6**). However, due

to the limited number of studies, further investigation is required to confirm these results.

In addition, some studies have shown that other imaging technologies, including 2D-SWE[29-32], MRE[22,33] and ARFI[19,34], also perform well in staging liver fibrosis (**Supplementary Table 4**). Further, 2D-SWE had excellent accuracy, with a summary AUROC of 0.91 for cirrhosis in patients with AIH (**Table 3**). In comparison, our findings indicated that 2D-SWE and ARFI had good accuracy with higher sensitivity, specificity, and AUROC for AF and cirrhosis in patients with PBC, while the AUROC of MRE was higher in patients with AIH. Interestingly, compared with TE, 2D-SWE produces a two-dimensional grayscale image so that interference from the gallbladder, ascites, and large tubular structures in the liver can be effectively avoided. However, the number of studies on 2D-SWE, MRE, and ARFI included in our analysis was small. Indeed, the diagnostic accuracies of 2D-SWE and MRE require further studies with larger sample sizes to determine the best method for staging fibrosis in patients with AILDs.

However, the overlap syndrome, one of the AILDs, also deserves attention because it exhibits significantly higher rates of various complications, progresses to cirrhosis more rapidly, and has a poor treatment response to ursodeoxycholic acid[35-36]. Hence, the development of non-invasive methods is beneficial for this disease. Wu *et al* reported that the AUROCs of TE for SF, AF, and cirrhosis were 0.837, 0.910, and 0.996, respectively. Yan *et al* reported that the AUROCs of SWE were 0.91, 0.97, and 0.96, respectively[37-38]. These results show that noninvasive imaging techniques have excellent accuracy for overlap syndrome, although more studies are required for further validation.

Our study has some limitations. First, we only included studies published in English language; therefore, a language bias may have influenced the results. Second, we did not consider the confounding factors such as obesity, whereas a previous study proposed that a high BMI may reduce the efficiency of ultrasound-based elastography techniques in detecting fibrosis[39]. However, only a limited number of studies have provided sufficient data to conduct subgroup analyses to explore the potential impact of BMI on the diagnostic effects. Third, it is unknown whether ALT level is responsible for the

difference in the diagnosis of TE between patients with AIH and PBC due to a lack of sufficient data. Moreover, the treatment conditions before inclusion in the study were unknown, and the lack of pretreatment studies made it impossible to compare the effects of treatment on outcomes. Finally, the number of studies on SWE, MRE, and ARFI was inadequate to compare the effects of these imaging technologies and TE.

CONCLUSION

In conclusion, LSM using TE had better diagnostic performance for staging hepatic fibrosis in patients with AILDs compared to other serum biomarkers, especially in patients with PBC. The appropriate cutoff value for staging in AF and cirrhosis ranged from 9.6 to 10.7 KPa and 14.4 to 16.9 KPa, respectively, for patients with PBC. Additional recommended optimal cutoff values warrant further investigation to provide a better reference for clinical applications.

ARTICLE HIGHLIGHTS

Research background

Non-invasive criteria are needed for autoimmune liver diseases (AILDs) to assess liver fibrosis stage for prognosis and treatment decisions.

Research motivation

Results of individual diagnostic test accuracy studies assessing the diagnostic accuracy of transient elastography (TE) for the diagnosis of AILD appear promising, however, previous systematic review and meta-analyses have focused primarily on other liver diseases, which is still lacking in AILD.

Research objectives

Results of individual diagnostic test accuracy studies assessing the diagnostic accuracy of transient elastography (TE) for the diagnosis of AILD appear promising, however,

previous systematic review and meta-analyses have focused primarily on other liver diseases, which is still lacking in AILD.

Research methods

The PubMed, The Cochrane and EMbase databases were searched for literature. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to evaluate the quality. Meta-Disc 1.4 and STATA12.0 software were used to analyze the combined statistics: Sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio (DOR), and area under the curve (AUC) fitted to the total receiver Operating characteristic curve (SROC).

Research results

A total of 60 studies were included to meta-analyze. The summary AUROC values were 0.93, 0.93, and 0.91 for significant fibrosis (SF), advanced fibrosis (AF) and cirrhosis, respectively, in primary biliary cholangitis (PBC) patients, while 0.84, 0.88, and 0.90, respectively, in autoimmune hepatitis (AIH) patients.

Research conclusions

TE is a reliable method for diagnosis in AILD patients, especially in PBC patients. The appropriate cutoff value for staging AF and cirrhosis ranged from 9.6 to 10.7 KPa and 14.4 to 16.9 KPa, respectively.

Research perspectives

We propose a suitable diagnostic threshold for TE in PBC patients, however, future prospective multi-center studies with TE and histopathology protocol are required to validate our results.

86696_Auto_Edited.docx

6 Chen Cheng, Yongguo Yang, Wei Yang, Daomeng Wang, Chen Yao. "The diagnostic value of CEA for lung cancer-related malignant pleural effusion in China: a meta-analysis", Expert Review of Respiratory Medicine, 2021
Crossref
23 words — < 1%

7 Bingtian Dong, Yuping Chen, Yongjian Chen, Huaming Wang, Guorong Lyu. "Diagnostic accuracy of liver stiffness on two-dimensional shear wave elastography for detecting clinically significant portal hypertension: a meta-analysis", Expert Review of Medical Devices, 2022
Crossref
19 words — < 1%

8 www.researchgate.net
Internet
19 words — < 1%

9 Carmen Yanette Suarez-Quintero, Oscar Patarroyo Henao, Oscar Muñoz-Velandia. "Concordance between hepatic biopsy and the APRI index (Ast to Platelet Ratio Index) for the diagnosis of cirrhosis in patients with autoimmune liver disease", Gastroenterología y Hepatología (English Edition), 2021
Crossref
18 words — < 1%

10 mafiadoc.com
Internet
18 words — < 1%

11 www.ncbi.nlm.nih.gov
Internet
17 words — < 1%

12 Li Mingkai, Wan Sizhe, Wu Xiaoying, Lin Ying, Bin Wu. "Diagnostic performance of elastography on liver fibrosis in antiviral treatment-naive chronic hepatitis B patients: a meta-analysis", Gastroenterology Report, 2022
Crossref
16 words — < 1%

13 Shanshan Wu, Zhirong Yang, Jialing Zhou, Na Zeng, Zhiying He, Siyan Zhan, Jidong Jia, Hong You. "Systematic review: diagnostic accuracy of non-invasive tests for staging liver fibrosis in autoimmune hepatitis", Hepatology International, 2018
Crossref

15 words — < 1%

14 richtlijnendatabase.nl
Internet

15 words — < 1%

15 digitalcommons.fiu.edu
Internet

14 words — < 1%

16 www.e-ultrasonography.org
Internet

14 words — < 1%

17 Se Jin Cho, Jae Hyoung Kim, Sung Hyun Baik, Leonard Sunwoo, Yun Jung Bae, Byung Se Choi. "Diagnostic performance of MRI of post-laminar optic nerve invasion detection in retinoblastoma: A systematic review and meta-analysis", Neuroradiology, 2020
Crossref

13 words — < 1%

18 aasldpubs.onlinelibrary.wiley.com
Internet

13 words — < 1%

19 "Abstracts of the 26th Annual Conference of APASL, February 15–19, 2017, Shanghai, China", Hepatology International, 2017
Crossref

12 words — < 1%

20 Guangqin Xiao, Jiayin Yang, Lunan Yan. "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

12 words — < 1%

virus infection: A systemic review and meta-analysis",
Hepatology, 2015
Crossref

21   www.mdpi.com   12 words — < 1%
     Internet