Diagnostic Accuracy of 2-Dimensional Shear Wave Elastography for the Staging of Liver Fibrosis
A Meta-analysis

Wei Zhang, MD, Ya Zhu, MD, Cuncheng Zhang, MD, Haitao Ran, PhD

Objectives—To evaluate the overall accuracy of 2-dimensional shear wave elastography (SWE) for the staging of liver fibrosis.

Methods—Literature databases and conference abstracts were searched from 2000 to September 2017. Sensitivity, specificity, and other information were extracted from the included studies. Methodological quality was assessed with Quality Assessment of Diagnostic Accuracy Studies 2 tools. Data were synthesized by a bivariate hierarchical model.

Results—The summary sensitivity and specificity were 0.85 (95% confidence interval [CI], 0.80–0.89) and 0.79 (95% CI, 0.72–0.85) for fibrosis stage F≥2, 0.90 (95% CI, 0.87–0.93) and 0.85 (95% CI, 0.80–0.89) for F≥3, and 0.89 (95% CI, 0.84–0.93) and 0.92 (95% CI, 0.89–0.95) for F=4, respectively. The areas under the summary receiver operating characteristic curve for F≥2, F≥3, and F=4 was 0.81, 0.77, and 0.84.

Conclusions—Two-dimensional SWE is a good noninvasive method for the diagnosis of substantial liver fibrosis and cirrhosis. Further studies are needed to assess severe fibrosis and to perform head-to-head comparisons of 2-dimensional SWE and other imaging modalities for the evaluation of liver fibrosis.

Key Words—fibrosis staging; liver fibrosis; meta-analysis; 2-dimensional shear wave elastography

Hepatic fibrosis is the repair response and healing process of liver injury caused by chronic liver diseases, leading finally to cirrhosis.1 And the risk of hepatocellular carcinoma increases according to the stage of liver fibrosis.2 There are multiple causes of chronic liver disease including hepatitis B, hepatitis C, non-alcoholic fatty liver disease and alcoholic liver disease.3 A precise estimation of the severity of liver fibrosis and a valid diagnosis of cirrhosis are important steps for determining the prognosis and planning treatment for patients with chronic liver diseases. Liver biopsy has been considered the standard of reference for the evaluation of liver fibrosis. However, it has some limitations, including its invasive nature, incidence of sampling errors, intra- and inter-observer variability and expense.4,5

In recent years, research has focused on the evaluation of noninvasive methods for the assessment of liver fibrosis, and various ultrasound-based elastographic methods have enabled noninvasive,..
Two-dimensional SWE, based on supersonic shear imaging, is a novel elastographic technique among these methods. It can be conveniently performed with the use of an ultrasound scanner and can create a real-time 2D quantitative map of liver tissue stiffness under the guidance of very high-frame rate B-mode imaging. In addition 2D SWE can assess the homogeneity of the liver because it produces color images corresponding to the varying degrees of hardness. Two-dimensional SWE could represent an alternative method to transient elastography and point SWE for the noninvasive assessment of liver fibrosis.

Two-dimensional SWE has proven to be efficient for the evaluation of liver fibrosis in several small to moderate-size trials. In this study, we performed a meta-analysis to assess the overall performance of 2D SWE for the diagnosis of liver fibrosis, including all relevant publications.

**Materials and Methods**

**Inclusion and Exclusion Criteria**

Studies were included if they met each of the following criteria: (1) they evaluated the performance of 2D SWE for the staging of liver fibrosis in adults; (2) the diagnosis of liver fibrosis was made with liver biopsy as the reference standard; (3) the studies used METAVIR as a comparable liver biopsy staging system; and (4) necessary data were provided to extract the true-positive, false-positive, true-negative, and false-negative results or studies with available data to construct 2 × 2 contingency tables. Exclusion criteria were as follows: (1) animal studies; (2) duplicate publications; and (3) reviews, letters, and editorials.

**Literature Search and Data Extraction**

Two authors independently performed a systematic literature search in both the Web of Science and PubMed for original English-language articles from January 2000 to September 2017. The search was performed with the following key words: “elastography and liver,” “shear wave elastography,” “2D SWE,” and “liver fibrosis or hepatic fibrosis.” We also scanned the references of potentially eligible articles. Any discrepancies between the researchers were resolved by an additional author. The data extraction form included the following: first author, year of publication, country or region of origin, sample size, number of male and female patients, mean age, body mass index (BMI), etiology of liver disease, cutoff stage used, and true-positive, false-positive, true-negative, and false-negative results from 2D SWE for each fibrosis stage. All analyses were based on previously published studies; thus, no ethical approval and patient consent were required.

**Quality Assessment**

To assess the quality of the included studies, the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) questionnaire was used. Each study’s risks of bias and concern for applicability were rated as low, high, or unclear for each domain (assessment of applicability was not applied in the domain of flow and timing).

**Analysis**

According to the METAVIR scoring system, liver fibrosis was staged into 5 groups: F0, no fibrosis; F1, portal fibrosis without septa; F2, portal fibrosis and few septa; F3, numerous septa without cirrhosis; and F4, cirrhosis. Then, the estimations of the diagnostic accuracy of F0 versus F1–4, F0/1 versus F2–4, F0–2 versus F3–4, and F0–3 versus F4 were described as F≥1, F≥2, F≥3, and F=4.

We used a bivariate hierarchical model to pool the sensitivities, specificities, and summary receiver operating characteristic curves and to estimate the 95% confidence intervals (CIs) for each hepatic fibrosis. The I² statistic and Q test were used to assess heterogeneity across individual studies. I² values of less than 25%, 25% to 50%, and greater than 50% indicate low, moderate, and high heterogeneity, respectively. Because only 10 studies were included in the main analysis, we only tested 3 moderator variables to assess whether they could explain the heterogeneity among studies. The 3 moderators were the following: (1) the presence of a control group (whether the study included healthy volunteers); (2) composition of etiology (all or most of the patients in the studies with hepatitis B virus infection or another etiology); and (3) mean BMI.
The risk of a publication bias was examined by construction of a funnel plot of the logarithm odds ratios versus their inverse standard errors. Egger tests were conducted to measure the degree of funnel plot asymmetry.\textsuperscript{15}

The quality assessment was made by Review Manager version 5.3 (The Cochrane Collaboration, Copenhagen, Denmark; 2014). All other analyses were performed with R version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria), particularly the R meta4diag package by Jingyi Guo and Andrea Riebler (version 2.0.5,05.07.2016).

**Results**

**Study Selection and Assessment**

The study selection process is illustrated in Figure 1. The initial search yielded 160 articles, 16 of which were excluded because of duplication. The remaining 144 articles were screened by title and abstract; of these, 86 studies were unrelated to the topic; 12 studies were review articles; 14 conference abstracts had insufficient data, and we failed to obtain them from the corresponding authors via e-mail; and 7 were animal studies. A total of 119 studies were excluded finally. After a detailed review of these articles, 10 full-text articles were finally included.\textsuperscript{16–25} The quality of the included articles as assessed according to the QUADAS-2 criteria is reported in Table 1.

**Data Synthesis and Analysis**

The patient characteristics for each study are summarized in Table 2. Ten studies with a total of 2083 patients were included in the meta-analysis. The numbers of studies discriminating $F\geq 1$, $F\geq 2$, $F\geq 3$, and $F=4$ were 3, 10, 9, and 10, respectively. Therefore, we did not synthesize the data for the classification of $F\geq 1$ because of the small sample size.

Figure 2 shows forest plots of summary sensitivities and specificities. For each stage, discrimination in

---

**Table 1. Risk of Bias and Concerns Regarding the Applicability of the Studies in the Analysis (QUADAS-2)**

| Study             | Year | Patient Selection | Index Test | Reference Standard | Flow and Timing | Applicability Concerns | Patient Selection | Index Test | Reference Standard |
|-------------------|------|-------------------|------------|--------------------|-----------------|------------------------|-------------------|------------|---------------------|
| Bende et al\textsuperscript{16} | 2017 | H                 | H          | H                  | L               | L                      | L                 | L          | L                   |
| Bota et al\textsuperscript{17}  | 2015 | L                 | H          | H                  | L               | L                      | L                 | H          | H                   |
| Ferraioli et al\textsuperscript{18} | 2012 | L                 | H          | L                  | L               | L                      | L                 | L          | L                   |
| Guibal et al\textsuperscript{19} | 2015 | L                 | H          | L                  | L               | L                      | L                 | L          | L                   |
| Leung et al\textsuperscript{20}  | 2013 | H                 | L          | L                  | H               | L                      | L                 | L          | L                   |
| Paul et al\textsuperscript{21}   | 2017 | L                 | H          | L                  | L               | L                      | L                 | L          | L                   |
| Samir et al\textsuperscript{22}  | 2015 | L                 | H          | L                  | L               | L                      | L                 | L          | L                   |
| Sporea et al\textsuperscript{23} | 2014 | H                 | L          | L                  | L               | L                      | L                 | L          | L                   |
| Zeng et al\textsuperscript{24}   | 2017 | L                 | H          | L                  | ?               | L                      | L                 | L          | L                   |
| Zheng et al\textsuperscript{25}  | 2015 | L                 | H          | L                  | H               | L                      | L                 | L          | L                   |

L indicates low risk; H, high risk; and ?, unclear risk.
the bivariate meta-analysis led to summary sensitivities of 0.85 (95% CI, 0.80–0.89), 0.90 (95% CI, 0.87–0.93), and 0.89 (95% CI, 0.84–0.93) for F≥2, F≥3, and F=4, respectively. Similarly, the summary specificities for F≥2, F≥3, and F=4 were 0.79 (95% CI, 0.72–0.85), 0.85 (95% CI, 0.80–0.89), and 0.92 (95% CI, 0.89–0.95). The summary receiver operating characteristic curves for F≥2, F≥3, and F=4 are shown in Figure 3. The areas under the curves (AUCs) for F≥2, F≥3, and F=4 were 0.81, 0.77, and 0.84. The I² values of the test results for F≥2, F≥3, and F=4 were 74.27%, 84.08%, and 78.61%, indicating that the heterogeneity was significant among studies. The mean BMI, presence of a control group, and

### Table 2. Characteristics of Included Studies

| Study         | Year | Country    | n for Analysis | Male, % | Mean Age, y | Liver Disease | Fibrosis Cutoff Stage |
|---------------|------|------------|----------------|---------|-------------|---------------------|-----------------------|
| Bende et al16 | 2017 | Romania    | 331            | 317     | 38.4        | 55                 | CHC, CHB, other       |
| Bota et al17  | 2015 | Austria    | 126            | 94      | 50.4        | 52.7 ± 13.3        | CHC, CHB, others      |
| Ferraioli et al18 | 2012 | Italy      | 121            | 118     | 87          | 44.8 ± 11.9        | CHC                   |
| Guibal et al19 | 2015 | France     | 170            | 148     | 64.2        | 54.3 ± 13.2        | NASH, CHB, other      |
| Leung et al20  | 2013 | Hong Kong  | 454            | 226     | 65          | 48.8 ± 12.3        | CHB                   |
| Paul et al21  | 2017 | India      | 256            | 237     | 73.3        | 32.6 ± 11.6        | CHB, CHC              |
| Samir et al22  | 2015 | United States | 138           | 136     | 51.4        | 49                 | CHC, CHB, other       |
| Sporea et al23 | 2014 | Romania    | 383            | 250     | 45.9        | 54                 | CHC, CHB, other       |
| Zeng et al24  | 2017 | China      | 257            | 235     | 77.4        | 36.7 ± 9.4         | CHB                   |
| Zheng et al25  | 2015 | China      | 200            | 167     | 71.2        | 37.7               | CHC, CHB, other       |

Data are presented as mean ± SD where applicable. CHB indicates chronic hepatitis B; CHC, chronic hepatitis C; and NASH, nonalcoholic steatohepatitis.

* Cutoff stage for fibrosis used for the construction of a 2 x 2 contingency table.

**Figure 2.** Forest plots of sensitivity and specificity for F≥2 (A), F≥3 (B), and F=4 (C).
disease type did not significantly explain the heterogeneity for each stage.

Publication Bias

The funnel plot asymmetry tests for each fibrosis stage are shown in Figure 4. The P values from the asymmetry test (Egger test) were as follows: P = .0343 for $F \geq 2$, $P < .0001$ for $F \geq 3$, and $P = .0015$ for $F = 4$; a potential publication bias was likely present.

Discussion

This study evaluated the diagnostic performance of 2D SWE for the staging of liver fibrosis, and 10 full-text studies were included: 2 studies including patients infected with chronic hepatitis B, 18, 20, 24, 1 study including patients infected with chronic hepatitis C, 18, and 7 studies pooling patients with different chronic liver diseases. 16, 17, 19, 21–23, 25 The results of our meta-analysis demonstrated that 2D SWE gave good diagnostic accuracy for detection of liver cirrhosis ($F = 4$) and substantial fibrosis ($F \geq 2$), with pooled sensitivity of 0.89, specificity of 0.92, and AUC of 0.84 for $F = 4$ and sensitivity of 0.85, specificity of 0.79, and AUC of 0.81 for $F \geq 2$. However, it was not highly accurate for diagnosis of severe fibrosis ($F \geq 3$) with an AUC of 0.77. Our results are partly in accordance with the results of a previous meta-analysis, 26 which reported diagnostic accuracies of 0.86, 0.90, and 0.93 for $F \geq 2$, $F \geq 3$, and $F = 4$, respectively; however, the number of patients included in our meta-analysis was almost twice that of the previous study. Thus, the performances of both

Figure 3. Summary receiver operating characteristic (SROC) curves for $F \geq 2$, $F \geq 3$, and $F = 4$. The 95% confidence regions and prediction regions are shown.
meta-analyses confirmed the diagnostic accuracy for liver cirrhosis and substantial fibrosis. With the use of a pretest probability of 50%, 2D SWE had high posttest probabilities of 80.1%, 85.7%, and 91.8% for substantial fibrosis, severe fibrosis, and liver cirrhosis, respectively.

A comparable meta-analysis of SWE revealed overall diagnostic accuracies of 0.93 for the diagnosis of liver cirrhosis, 0.93 for the diagnosis of severe fibrosis, and 0.85 for the diagnosis of substantial fibrosis, encompassing 2 of the SWE technologies: point SWE and 2D SWE. However, the studies in our meta-analysis only included 2D SWE technology. Two-dimensional SWE has been proven to be efficient for the evaluation of liver fibrosis in several studies, reporting equivalent to superior diagnostic accuracies compared to point SWE for the assessment of liver fibrosis. Two-dimensional SWE produced a shear wave that spread less and thus decayed less rapidly with distance than that of a single pushing focus in point SWE. Two-dimensional SWE can create a real-time 2D quantitative map of liver tissue under the guidance of very high-frame rate B-mode imaging.

A previous meta-analysis evaluating point SWE including patients with mixed chronic liver diseases reported accuracies of 0.84 for the diagnosis of substantial fibrosis, 0.89 for the diagnosis of severe fibrosis, and 0.91 for the diagnosis of liver cirrhosis. A recent study reported that 2D SWE showed a better correlation compared to point SWE in a pediatric population. The advantage of 2D SWE is highlighted by the fact that this ultrasound-based elastographic method is integrated in an ultrasound machine that can be used for other examinations. As a new method, 2D SWE has the following advantages compared to acoustic radiation force impulse imaging: qualitative (color coded) and quantitative measurement, easier and more manageable measurement, and stability of the measured value. However, 2D SWE has some limitations, such as the

Figure 4. Funnel plots according to F≥2, F≥3, and F=4. The points represent studies included in the meta-analysis.
subjective nature of the color scale and a potential bias when selecting the region of interest.

Our assessment of the included studies according to QUADAS-2 showed that most of the studies’ index tests were judged to have a high risk of bias because they did not use prespecified cutoff values. In contrast, most studies used an acceptable method to stage liver fibrosis based on liver biopsy. Some studies were assessed to be at a high risk of bias for patient selection.

Our analysis had several limitations. First, significant heterogeneity and a publication bias were found among the individual studies for all fibrosis stages. The heterogeneity had multiple sources. In this analysis, heterogeneity may come from the difference of the sampling protocol, biopsy protocol, machine, or software used for measurement in the studies or a combination of these factors. Meanwhile, the accuracy of 2D SWE mainly relies on the operator’s performance. We assessed only the reasons for different BMIs in patients, studies including a control group or not, and studies with patients infected with hepatitis B virus, hepatitis C virus, or mixed disease. The sensitivity analysis indicated that the different mean BMIs, control group inclusion, and disease type did not significantly explain the heterogeneity for each stage. Due to the small sizes of the included studies, we did not analyze other probable sources. Second, the cutoff values for diagnosing F2, F3, and F4 fibrosis ranged from 6.0 to 18.1 kPa in the different studies. However, we could not perform a subgroup analysis to obtain a single cutoff value because of the limited number of studies, and this factor needs to be confirmed by an analysis including a larger number of studies.

In conclusion, 2D SWE is a good noninvasive method for the diagnosis of substantial liver fibrosis and cirrhosis. It is a promising screening tool for integration into routine imaging protocols for the staging of liver fibrosis and cirrhosis. Further studies with large sample sizes are needed to assess severe fibrosis and perform head-to-head comparisons of 2D SWE with other noninvasive techniques for the evaluation of liver fibrosis.

References

1. Benvegnù L, Gios M, Boccato S, Alberti A. Natural history of compensated viral cirrhosis: a prospective study on the incidence and hierarchy of major complications. Gut 2004; 53:744–749.
2. Ferraioli G, Filice C, Castera L, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography, part 3: liver. Ultrasound Med Biol 2015; 41:1161–1179.
3. Sebastiani G, Castera L, Hallon P, et al. The impact of liver disease aetiology and the stages of hepatic fibrosis on the performance of non-invasive fibrosis biomarkers: an international study of 2411 cases. Aliment Pharmacol Ther 2011; 34:1202–1216.
4. Cadranel JF, Rufat P, Degos F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). Hepatology 2000; 32:477–481.
5. Seeff LB, Everson GT, Morgan TR, et al. Complication rate of percutaneous liver biopsies among persons with advanced chronic liver disease in the HALT-C trial. Clin Gastroenterol Hepatol 2010; 8:877–883.
6. Bamber J, Cosgrove D, Dietrich CF, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography, part 1: basic principles and technology. Ultraschall Med 2013; 34:169–184.
7. European Association for Study of Liver; Asociacion Latinoamericana para el Estudio del Higado. EASL-ALEH clinical practice guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol 2015; 63:237–264.
8. Sandrin I, Fourquet B, Hasquenoph JM, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003; 29:1705–1713.
9. Bercoff J, Tanter M, Fink M. SuperSonic shear imaging: a new technique for soft tissue elasticity mapping. IEEE Trans Ultrason Ferroelectr Freq Control 2004; 51:396–409.
10. Shima T, Nightingale KR, Palmeri ML, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography, part 1: basic principles and terminology. Ultrasound Med Biol 2015; 41:1162–1147.
11. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. J Hepatol 2007; 47:598–607.
12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155:529–536.
13. Arends LR, Hamza TH, van Houwelingen JC, et al. Bivariate random effects meta-analysis of ROC curves. Med Decision Making 2008; 28:621–638.
14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58:982–990.
15. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315:629–634.
16. Bende F, Sporea I, Sirli R, et al. Performance of 2D-SWE.GE for predicting different stages of liver fibrosis, using transient elastography as the reference method. Med Ultrason 2017; 19:143–149.
17. Bota S, Paternostro R, Etschmaier A, et al. Performance of 2-D shear wave elastography in liver fibrosis assessment compared with serologic tests and transient elastography in clinical routine. *Ultrasound Med Biol* 2015; 41:2340–2349.

18. 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.

19. Guibal A, Renosi G, Rode A, et al. Shear wave elastography: an accurate technique to stage liver fibrosis in chronic liver diseases. *Diagn Interv Imaging* 2016; 97:91–99.

20. Leung VY, Shen J, Wong VW, et al. Quantitative elastography of liver fibrosis and spleen stiffness in chronic hepatitis B carriers: comparison of shear-wave elastography and transient elastography with liver biopsy correlation. *Radiology* 2013; 269:910–918.

21. Paul SB, Das P, Mahanta M, et al. Assessment of liver fibrosis in chronic hepatitis: comparison of shear wave elastography and transient elastography. *Abdom Radiol* 2017; 42:2864–2873.

22. Samir AE, Dhyani M, Vij A, et al. Shear-wave elastography for the estimation of liver fibrosis in chronic liver disease: determining accuracy and ideal site for measurement. *Radiology* 2015; 274:888–896.

23. Sporea I, Bota S, Grădinaru-Tașcău O, Sirl I, Popescu A, Jurchiș A. Which are the cut-off values of 2D-shear wave elastography (2D-SWE) liver stiffness measurements predicting different stages of liver fibrosis, considering transient elastography (TE) as the reference method? *Eur J Radiol* 2014; 83:e118–e122.

24. Zeng J, Zheng J, Huang Z, et al. Comparison of 2-D shear wave elastography and transient elastography for assessing liver fibrosis in chronic hepatitis B. *Ultrasound Med Biol* 2017; 43:1563–1570.

25. Zheng J, Gao H, Zeng J, et al. Two-dimensional shear-wave elastography and conventional US: the optimal evaluation of liver fibrosis and cirrhosis. *Radiology* 2013; 275:290–300.

26. Hermann E, de Ledinghen V, Cassinotto C, et al. Assessment of biopsy-proven liver fibrosis by 2D-shear wave elastography: an individual patient data based meta-analysis. *Hepatology* 2018; 67:260–272.

27. Feng JC, Li J, Wu XW, Peng XY. Diagnostic accuracy of SuperSonic shear imaging for staging of liver fibrosis: a meta-analysis. *J Ultrasound Med* 2016; 35:329–339.

28. Cassinotto C, Lapuyade B, Mouriès A, et al. Non-invasive assessment of liver fibrosis with impulse elastography: comparison of SuperSonic shear imaging with ARFI and FibroScan. *J Hepatol* 2014; 61:550–557.

29. Defieux T, Gennisson JL, Bousquet L, et al. Investigating liver stiffness and viscosity for fibrosis, steatosis and activity staging using shear wave elastography. *J Hepatol* 2015; 62:317–324.

30. Nierhoff J, Chavez Ortiz AA, Hermann E, Zeuzem S, Friedrich-Rust M. The efficiency of acoustic radiation force impulse imaging for the staging of liver fibrosis: a meta-analysis. *Eur Radiol* 2013; 23:3040–3053.

31. Belei O, Sporea I, Grădinaru-Tașcău O, et al. Comparison of three ultrasound based elastographic techniques in children and adolescents with chronic diffuse liver diseases. *Med Ultrason* 2016; 18:145–150.