Title: Noninvasive Techniques for the Evaluation and Monitoring of Patients with Chronic Liver Disease

**Professional**

Original Effective Date: September 16, 2016  
Revision Date(s): September 16, 2016; January 18, 2017; December 20, 2017; January 1, 2019; February 1, 2019; March 13, 2019  
Current Effective Date: March 13, 2019

**Institutional**

Original Effective Date: September 16, 2016  
Revision Date(s): September 16, 2016; January 18, 2017; December 20, 2017; January 1, 2019; February 1, 2019; March 13, 2019  
Current Effective Date: March 13, 2019

---

**Title:** Noninvasive Techniques for the Evaluation and Monitoring of Patients with Chronic Liver Disease

**Medical Policy**

**Populations**
- With chronic liver disease

**Interventions**
- Interventions of interest are:
  - FibroSURE serum panels

**Comparators**
- Comparators of interest are:
  - Liver biopsy
  - Noninvasive radiologic methods
  - Other multianalyte serum assays

**Outcomes**
- Relevant outcomes include:
  - Test validity
  - Morbid events
  - Treatment-related morbidity

---

**Populations**
- With chronic liver disease

**Interventions**
- Interventions of interest are:
  - Multianalyte serum assays for liver function assessment other than FibroSURE

**Comparators**
- Comparators of interest are:
  - Liver biopsy
  - Noninvasive radiologic methods
  - Other multianalyte serum assays

**Outcomes**
- Relevant outcomes include:
  - Test validity
  - Morbid events
  - Treatment-related morbidity
DESCRIPTION
Noninvasive techniques to monitor liver fibrosis are being investigated as alternatives to liver biopsy in patients with chronic liver disease. There are 2 options for noninvasive monitoring: (1) multianalyte serum assays with algorithmic analysis of either direct or indirect biomarkers; and (2) specialized radiologic methods, including magnetic resonance elastography (MRE), transient elastography, acoustic radiation force impulse imaging (ARFI), and real-time transient elastography (RTE).

OBJECTIVE
The objective of this policy is to determine whether use of noninvasive techniques for detecting liver fibrosis compared with liver biopsy can improve the net health outcome in patients with chronic liver disease.

BACKGROUND
Biopsy for Chronic Liver Disease
The diagnosis of non-neoplastic liver disease is often made from needle biopsy samples. In addition to establishing a disease etiology, liver biopsy can determine the degree of inflammation present and can stage the degree of fibrosis. The degree of inflammation and fibrosis may be assessed by different scoring schemes. Most of these scoring schemes grade inflammation from 0 to 4 (0 = no or minimal inflammation, 4 = severe) and fibrosis from 0 to 4 (0 = no fibrosis, 4 = cirrhosis). There are several limitations to liver biopsy, including its invasive nature, small tissue sample size, and subjective grading system. Regarding small tissue sample size, liver fibrosis can be patchy and thus missed on a biopsy sample, which includes only 0.002% of the liver tissue. A noninvasive alternative to liver biopsy would be particularly helpful, both to initially assess patients and then as a monitoring tool to assess response to therapy. The implications of using liver biopsy as a reference standard are discussed in the Rationale section.

Hepatitis C Virus
Infection with HCV can lead to permanent liver damage. Liver biopsy is typically recommended before the initiation of antiviral therapy. Repeat biopsies may be performed to monitor fibrosis progression. Liver biopsies are analyzed according to a
Histologic scoring system; the most commonly used one for hepatitis C is the Metavir scoring system, which scores the presence and degree of inflammatory activity and fibrosis. The fibrosis is graded from F0 to F4, with a Metavir score of F0 signifying no fibrosis and F4 signifying cirrhosis (which is defined as the presence throughout the liver of fibrous septa that subdivide the liver parenchyma into nodules and represents the final and irreversible form of disease). The stage of fibrosis is the most important single predictor of morbidity and mortality in patients with hepatitis C. Biopsies for hepatitis C are also evaluated according to the degree of inflammation present, referred to as the grade or activity level. For example, the Metavir system includes scores for necroinflammatory activity ranging from A0 to A3 (A0 = no activity, A1 = minimal activity, A2 = moderate activity, A3 = severe activity).

**Hepatitis B Virus**
Most people who become infected with hepatitis B virus (HBV) recover fully, but a small portion will develop chronic HBV, which can lead to permanent liver damage. As with HCV, identification of liver fibrosis is needed to determine timing and management of treatment, and liver biopsy is the criterion standard for staging fibrosis. The grading of fibrosis in HBV also uses the Metavir system.

**Alcoholic Liver Disease**
Alcoholic liver disease (ALD) is the leading cause of liver disease in most Western countries. Histologic features of ALD usually include steatosis, alcoholic steatohepatitis (ASH), hepatocyte necrosis, Mallory bodies (tangled proteins seen in degenerating hepatocytes), a large polymorphonuclear inflammatory infiltrate, and, with continued alcohol abuse, fibrosis and possibly cirrhosis. The grading of fibrosis is similar to the scoring system used in hepatitis C. The commonly used Laënnec scoring system uses grades 0 to 4, with 4 being cirrhosis.

**Nonalcoholic Fatty Liver Disease**
Nonalcoholic fatty liver disease (NAFLD) is defined as a condition that pathologically resembles ALD but occurs in patients who are not heavy users of alcohol. It may be associated with a variety of conditions, including obesity, diabetes, and dyslipidemia. The characteristic feature of NAFLD is steatosis. At the benign end of the spectrum of the disease, there is usually no appreciable inflammation, hepatocyte death, or fibrosis. In contrast, nonalcoholic steatohepatitis (NASH), which shows overlapping histologic features with ALD, is an intermediate form of liver damage, and liver biopsy may show steatosis, Mallory bodies, focal inflammation, and degenerating hepatocytes. NASH can progress to fibrosis and cirrhosis. A variety of histologic scoring systems have been used to evaluate NAFLD. The NAFLD activity score (NAS) system for NASH includes scores for steatosis (0-3), lobular inflammation (0-3), and ballooning (0-2). Cases with scores of 5 or greater are considered NASH, while cases with scores of 3 and 4 are considered borderline (probable or possible) NASH. The grading of fibrosis is similar to the scoring...
system used in hepatitis C. The commonly used Laënnec scoring system uses grades 0 to 4, with 4 being cirrhosis.

**Noninvasive Alternatives to Liver Biopsy**

**Multianalyte Assays**

A variety of noninvasive laboratory tests are being evaluated as an alternative to liver biopsy. Biochemical tests can be broadly categorized into indirect and direct markers of liver fibrosis. Indirect markers include liver function tests such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), the ALT/AST ratio (also referred to as the AAR), platelet count, and prothrombin index. In recent years, there has been growing understanding of the underlying pathophysiology of fibrosis, leading to direct measurement of the factors involved. For example, the central event in the pathophysiology of fibrosis is activation of the hepatic stellate cell. Normally, stellate cells are quiescent but are activated in the setting of liver injury, producing a variety of extracellular matrix (ECM) proteins. In normal livers, the rate of ECM production equals its degradation, but in the setting of fibrosis, production exceeds degradation. Metalloproteinases are involved in intracellular degradation of ECM, and a profibrogenic state exists when there is either a down regulation of metalloproteinases or an increase in tissue inhibitors of metalloproteinases (TIMP). Both metalloproteinases and TIMP can be measured in the serum, which directly reflects fibrotic activity. Other direct measures of ECM deposition include hyaluronic acid or $\alpha_2$-macroglobulin.

While many studies have been done on these individual markers, or on groups of markers in different populations of patients with liver disease, there has been interest in analyzing multiple markers using mathematical algorithms to generate a score that categorizes patients according to the biopsy score. It is proposed that these algorithms can be used as an alternative to liver biopsy in patients with liver disease. The following proprietary, algorithm-based tests are commercially available in the United States.

**FibroSURE**

**HCV FibroSURE**

HCV FibroSURE uses a combination of 6 serum biochemical indirect markers of liver function plus age and sex in a patented algorithm to generate a measure of fibrosis and necroinflammatory activity in the liver that correspond to the Metavir scoring system for stage (ie, fibrosis) and grade (ie, necroinflammatory activity). The measures are combined using a linear regression equation to produce a score between 0 and 1, with higher values corresponding to more severe disease. The biochemical markers include the readily available measurements of $\alpha_2$-macroglobulin, haptoglobin, bilirubin, $\gamma$-glutamyl transpeptidase (GGT), ALT, and apolipoprotein A1. Developed in France, the test has been clinically available in Europe under the name FibroTest since 2003 and is exclusively offered by LabCorp in the United States as HCV FibroSURE.
ASH FibroSURE
ASH FibroSURE (ASH Test) uses a combination of 10 serum biochemical markers of liver function together with age, sex, height, and weight in a proprietary algorithm and is proposed to provide surrogate markers for liver fibrosis, hepatic steatosis, and ASH. The biochemical markers include \( \alpha_2 \)-macroglobulin, haptoglobin, apolipoprotein A1, bilirubin, GGT, ALT, AST, total cholesterol, triglycerides, and fasting glucose. The test has been available in Europe under the name ASH Test and is exclusively offered by LabCorp in the United States as ASH FibroSURE.

NASH FibroSURE
NASH FibroSURE (NASH Test) uses a proprietary algorithm of the same 10 biochemical markers of liver function in combination with age, sex, height, and weight and is proposed to provide surrogate markers for liver fibrosis, hepatic steatosis, and NASH. The biochemical markers include \( \alpha_2 \)-macroglobulin, haptoglobin, apolipoprotein A1, bilirubin, GGT, ALT, AST, total cholesterol, triglycerides, and fasting glucose. The test has been available in Europe under the name NASH Test™ and is exclusively offered by LabCorp in the United States as NASH FibroSURE.

FIBROSpect II
FIBROSpect II uses a combination of 3 markers that directly measure fibrogenesis of the liver, analyzed with a patented algorithm. The markers include hyaluronic acid, TIMP-1, and \( \alpha_2 \)-macroglobulin. FIBROSpect II is offered exclusively by Prometheus Laboratories. The measures are combined using a logistic regression algorithm to generate a FIBROSpect II index score, ranging from 1 to 100 (or sometimes reported between 0 and 1), with higher scores indicating more severe disease.

Noninvasive Imaging Technologies
Noninvasive imaging technologies to detect liver fibrosis or cirrhosis among patients with chronic liver disease are also being evaluated as an alternative to liver biopsy. The noninvasive imaging technologies include transient elastography (eg, FibroScan), magnetic resonance elastography (MRE), ARFI (eg, Acuson S2000), and real-time tissue elastography (RTE; eg, HI VISION Preirus). Noninvasive imaging tests have been used in combination with multianalyte serum tests such as FibroTest or FibroSURE with FibroScan.

Transient Elastography
Transient elastography (FibroScan) uses a mechanical vibrator to produce mild amplitude and low frequency (50 Hz) waves, inducing an elastic shear wave that propagates throughout the liver. US tracks the wave, measuring its speed, which correlates with liver stiffness. Increases in liver fibrosis also increase liver stiffness and resistance of liver blood flow. Transient elastography does not perform as well in patients with ascites, higher body mass index, or narrow intercostal margins. Although FibroScan may be used to measure fibrosis, unlike liver biopsy, it does not provide information on
necroinflammatory activity and steatosis, nor is it accurate during acute hepatitis or hepatitis exacerbations.

Acoustic Radiation Force Impulse Imaging
ARFI uses an US probe to produce an acoustic “push” pulse, which generates shear waves that propagate in tissue to assess liver stiffness. ARFI elastography evaluates the wave propagation speed to assess liver stiffness. The faster the shear wave speed, the harder the object. ARFI technologies include Virtual Touch Quantification and Siemens Acuson S2000 system. ARFI elastography can be performed at the same time as a liver sonographic evaluation, even in patients with a significant amount of ascites.

Magnetic Resonance Elastography
MRE uses a driver to generate 60-Hz mechanical waves on the patient’s chest well. The MRI equipment creates elastograms by processing the acquired images of propagating shear waves in the liver using an inversion algorithm. These elastograms represent the shear stiffness as a pixel value in kilopascals. MRE has several advantages over US elastography, including: (1) analyzing larger liver volumes; (2) analyzing liver volumes of obese patients or patients with ascites; and (3) precise analysis of viscoelasticity using a 3-dimensional displacement vector.

Real-Time Tissue Elastography
RTE is a type of strain elastography which uses a combined autocorrelation method to measure tissue strain caused by manual compression or a person’s heartbeat. The relative tissue strain is displayed on conventional color B mode US images in real time. Hitachi manufacturers the RTE devices, including one called HI VISION Preirus. The challenge is to identify a region of interest while avoiding areas likely to introduce artifacts, such as large blood vessels, the area near the ribs, and the surface of the liver. Areas of low strain increase as fibrosis progresses and strain distribution becomes more complex. Various subjective and quantitative methods have been developed to evaluate the results. RTE can be performed in patients with ascites or inflammation. This technology does not perform as well in severely obese individuals.

REGULATORY STATUS
On November 13, 2008, Acuson S2000™ Virtual Touch (Siemens AG), which provides ARFI imaging, was cleared for marketing by FDA through the 510(k) process (K072786) (K123622).

On August 12, 2009, AIXPLORER® Ultrasound System (SuperSonic Imagine), which provides shear wave elastography, was cleared for marketing by FDA through the 510(k) process (K091970).
On June 17, 2010, Hitachi HI VISION Preirus Diagnostic Ultrasound Scantier (Hitachi Medical Systems America), which provides real-time tissue elastography, was cleared for marketing by FDA through the 510(k) process (K093466).

On April 5, 2013, FibroScan® (Echosense SA), which uses transient elastography, was cleared for marketing FDA through the 510(k) process (K123806).

In February 2017, ElastQ Imaging shear wave elastography (Royal Phillips) was cleared for marketing by FDA through the 510(k) process (K163120).

FDA product code: IYO

**POLICY**

A. A single FibroSURE multianalyte assay may be considered **medically necessary** for the initial evaluation of patients with chronic liver disease.

B. FibroSURE multianalyte assays are considered **experimental / investigational** for monitoring of patients with chronic liver disease.

C. Other multianalyte assays with algorithmic analyses are considered **experimental / investigational** for the initial evaluation or monitoring of patients with chronic liver disease.

D. Transient elastography (FibroScan) imaging may be considered **medically necessary** for the initial evaluation of patients with chronic liver disease.

E. Transient elastography (FibroScan) imaging is considered **experimental / investigational** for monitoring of patients with chronic liver disease.

F. The use of other noninvasive imaging, including, but not limited to, magnetic resonance elastography, acoustic radiation force impulse imaging (eg, Acuson S2000), or real-time tissue elastography, is considered **experimental / investigational** for the initial evaluation or monitoring of patients with chronic liver disease.

**Policy Guidelines**

1. Multianalyte assays with algorithmic analyses (MAAAs) use the results from multiple assays of various types in an algorithmic analysis to determine and report a numeric score(s) or probability. The results of individual component assays are not reported separately.

2. This policy does not address standard imaging with ultrasound or MRI.
**RATIONALE**

The literature review was for the period through October 4, 2018.

Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.

The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. Evidence reviews assess the evidence on whether a test is clinically valid and clinically useful. Technical reliability is outside the scope of these reviews, and credible information on technical reliability is available from other sources.

**Noninvasive Testing for Chronic Liver Disease**

As noted in the Background, liver biopsy is an imperfect reference standard. There is a high rate of sampling error in the biopsy, which can lead to underdiagnosis of liver disease.\(^1\)\(^2\) These errors will bias estimates of performance characteristics of the noninvasive tests to which it is compared, and therefore such errors must be considered in appraising the body of evidence. Mehta et al (2009) estimated that—even under the best scenario where sensitivity and specificity of liver biopsy are 90%, and the prevalence of significant disease (increased liver fibrosis, scored as Metavir ≥ F2) is 40%—a perfect alternative marker would have calculated area under the receiver operating characteristic (AUROC) curve of 0.90.\(^3\) Therefore, the effectiveness of alternative technologies may be underestimated. In fact, when the accuracy of biopsy is presumed to be 80%, a comparative technology with an AUROC curve of 0.76 may actually have an AUROC curve of 0.93 to 0.99 for diagnosing true disease.

Due to a large number of primary studies published on this topic, this evidence review focuses on systematic reviews when available. The validation of multiple noninvasive tests are assessed individually in the following sections. Although options exist for performing systematic reviews with imperfect reference standards,\(^4\) most available reviews did not use any correction for the imperfect reference.

A systematic review by Crossan et al (2015) was performed for the National Institute for Health Research.\(^5\) The first objective of the review was to determine the diagnostic accuracy of different noninvasive liver tests compared with liver biopsy in the diagnosis and monitoring of liver fibrosis and cirrhosis in patients with hepatitis C virus (HCV), hepatitis B virus (HBV), nonalcoholic fatty liver disease (NAFLD), and alcoholic liver disease (ALD). Reviewers selected 302 publications and presentations from 1998 to April 2012. Patients with HCV were the most common population included in the studies while patients with ALD were the least common. FibroScan and FibroTest were the most commonly assessed tests across liver diseases. Aminotransferase to platelet ratio index (APRI) was also widely assessed in HBV and HCV but not in NAFLD or ALD. The estimates of diagnostic accuracy for each test by disease (as determined by Crossan et al) are discussed in further detail in the following sections. Briefly, for diagnosing significant fibrosis (stage ≥ F2) in HCV, the summary sensitivities and specificities were: FibroScan, 79% and 83%; FibroTest, 68% and 72%; APRI (low cutoff), 82% and 57%; acoustic radiation force impulse (ARFI) imaging, 85% and 89%; HepaScore, 73% and 73%, FIBROSpect II, 78% and 71%; and FibroMeter, 79%
and 73%, respectively. For diagnosing advanced fibrosis in HBV, the summary sensitivities and specificities were: FibroScan, 71% and 84%; FibroTest, 66% and 80%, respectively. There are no established or validated cutoffs for fibrosis stages across the diseases for most tests. For FibroTest, established cutoffs exist but were used inconsistently across studies. Test failures or reference standard(s) were frequently not captured in analyses. Most populations included in the studies were from tertiary care settings who have more advanced disease than the general population, which would overestimate the prevalence of the disease and diagnostic accuracy. These issues likely cause overestimates of sensitivities and specificities. The quality of the studies was generally rated as poor, with only 1.6% receiving a high-quality rating.

Houot et al (2016) reported on a systematic review funded by BioPredictive, the manufacturer of FibroTest. Reviewers included 71 studies published between January 2002 to February 2014 with over 12,000 participants with HCV and HBV comparing the diagnostic accuracy of FibroTest, FibroScan, APRI, and FIB4 index. Reviewers included studies that directly compared the tests and calculated median differences in the AUROC using Bayesian methods. There was no evaluation of the methodologic quality of the included studies. The Bayesian difference in AUROC curve for significant fibrosis (stage ≥ F2) between FibroTest and FibroScan was based on 15 studies and estimated to be 0.06 (95% credible interval [CrI], 0.02 to 0.09) favoring FibroTest. The difference in AUROC curve for cirrhosis for FibroTest vs FibroScan was based on 13 studies and estimated to be 0.00 (95% CrI, 0.04 to 0.04). The difference for advanced fibrosis between FibroTest and APRI was based on 21 studies and estimated to be 0.05 (95% CrI, 0.03 to 0.07); for cirrhosis, it was based on 14 studies and estimated to be 0.05 (95% CrI, 0.00 to 0.11), both favoring FibroTest.

**Multianalyte Serum Assays: Fibrosure (Fibrotest)**

**Clinical Context and Test Purpose**

The purpose of noninvasive testing in individuals with chronic liver disease is to detect liver fibrosis so that patients can avoid the potential adverse events of an invasive liver biopsy and receive appropriate treatment. The degree of liver fibrosis is an important factor in determining the appropriate approach for managing patients with liver disease (eg, hepatitis, ALD, NAFLD).

The question addressed in this portion of the evidence review is: Does the use of the FibroSURE multianalyte serum assay for detecting liver fibrosis improve the net health outcome in patients with chronic liver disease?

The following PICOTS were used to select literature to inform this review.

**Patients**

The relevant population of interest is individuals with chronic liver disease.

**Interventions**

The test being considered is the FibroSURE serum panel.

**Comparators**

The following tests and practices are currently being used to diagnose chronic liver disease: liver biopsy, noninvasive radiologic methods, and other multianalyte serum assays.
Outcomes
The general outcomes of interest are test validity, morbid events, and treatment-related morbidity.

Timing
Follow-up over months to years is of interest for the relevant outcomes.

Setting
Patients are actively managed by gastroenterologists and other specialists in an outpatient setting.

Study Selection Criteria
For the evaluation of clinical validity of this test, studies that meet the following eligibility criteria were considered:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard (describe the reference standard)
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.

Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Hepatitis C Virus
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Initial research into the HCV FibroSURE algorithm involved testing an initial panel of 11 serum markers in 339 patients with liver fibrosis who had undergone liver biopsy. From the original group of 11 markers, 5 were selected as the most informative, based on logistic regression, neural connection, and receiver operating characteristic (ROC) curves. Markers included α2-macroglobulin, haptoglobin, γ-globulin, apolipoprotein AI, γ-glutamyl transpeptidase, and total bilirubin.7 Using an algorithm-derived scoring system ranging from 0 to 1.0, authors reported that a score of less than 0.10 was associated with a negative predictive value (NPV) of 100% (ie, the absence of fibrosis, as judged by liver biopsy scores of Metavir F2-F4). A score greater than 0.60 was associated with a 90% positive predictive value (PPV) of fibrosis (ie, Metavir F2-F4). Authors concluded that liver biopsy might be deferred in patients with a score less than 0.10.

The next step in the development of this test was further evaluation of the algorithm in a cross-section of patients, including patients with HCV participating in large clinical trials before and after the initiation of antiviral therapy. A study by Poynard et al (2003) focused on patients with HCV participating in a randomized study of pegylated interferon and ribavirin.8 From the 1530 participants, 352 patients with stored serum samples and liver biopsies at study entry and at 24-
week follow-up were selected. The HCV FibroSURE score was calculated and then compared with the Metavir liver biopsy score. At a cutoff of 0.30, the HCV FibroSURE score had 90% sensitivity and 88% PPV for the diagnosis of Metavir F2-F4 fibrosis. The specificity was 36%, and the NPV was 40%.

Poynard et al (2004) also evaluated discordant results in 537 patients who underwent liver biopsy and the HCV FibroSURE and ActiTest on the same day; discordance was attributed to either the limitations in the biopsy or serum markers. In this study, cutoff values were used for individual Metavir scores (ie, F0-F4) and for combinations of Metavir scores (ie, F0-F1, F1-F2). The definition of a significant discordance between FibroTest and ActiTest and biopsy scores was at least 2 stages or grades in the Metavir system. Discordance was observed in 29% of patients. Risk factors for failure of HCV FibroSURE scoring system were as follows: the presence of hemolysis, inflammation, possible Gilbert syndrome, acute hepatitis, drugs inducing cholestasis, or an increase in transaminases. Discordance was attributable to markers in 2.4% of patients, to the biopsy in 18%, and nonattributed in 8.2% of patients. As noted in 2 reviews, the bulk of the research on HCV FibroSURE was conducted by researchers with an interest in the commercialization of the algorithm.

An Australian study reported by Rossi et al (2003) attempted to independently replicate the results of FibroSURE in 125 patients with hepatitis C. Using the cutoff of less than 0.1 to identify lack of bridging fibrosis (ie, Metavir stages F0-F1) and greater than 0.6 to identify fibrosis (ie, Metavir stages F2-F4), the NPV for a score of less than 0.1 was 89%, and the PPV of a score greater than 0.6 was 78%.

Poynard et al (2012) assessed the relative accuracy of FibroTest and FibroScan using a method to estimate performance characteristics when no perfect reference standard exists. The study included 1893 subjects retrospectively extracted from 4 prospective cohorts: 3 cohorts with HCV (n=1289) and a cohort of healthy volunteers (n=604). Four different tests (FibroTest, FibroScan, alanine aminotransferase [ALT], liver biopsy) were performed on all patients with HCV. Latent class models with random effects were used to combine the test results to construct a reference standard. Compared with biopsy as the reference standard, the sensitivity and specificity for the diagnosis of advanced fibrosis were 85% and 66% for FibroTest and 93% and 48% for FibroScan, respectively. However, when compared to the latent class reference standard, the specificity and sensitivity for the diagnosis of advanced fibrosis were 93% and 70% for FibroTest and 96% and 45%, for FibroScan, respectively.

In the Crossan (2015) systematic review, FibroTest was the most widely validated commercial serum test. Seventeen studies were included in the pooled estimate of the diagnostic accuracy of FibroTest for significant fibrosis (stage ≥ F2) in HCV. With varying cutoffs for positivity between 0.32 and 0.53, the summary sensitivity in HCV was 68% (95% confidence interval [CI], 58% to 77%) and specificity was 72% (95% CI, 70% to 77%). Eight studies were included for cirrhosis (stage F4) in HCV. The cutoffs for positivity ranged from 0.56 to 0.74 and the summary sensitivity and specificity were 60% (95% CI, 43% to 76%) and 86% (95% CI, 81% to 91%), respectively. Uninterpretable results were rare for tests based on serum markers.
Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials (RCTs).

The primary benefit of the FibroSURE (FibroTest in Europe) for HCV is the ability to avoid liver biopsy in patients without significant fibrosis. Thus, empirical data are needed that demonstrate that the FibroSURE test impacts clinician decision making on whether a biopsy should be performed and that the net effect is to reduce the overall number of biopsies while achieving similar clinical outcomes. There are currently no such published studies to demonstrate the effect on patient outcomes.

The FibroSURE test also has a potential effect on patient outcomes as a means to follow response to therapy. In this case, evidence needs to demonstrate that use of the test for response to therapy impacts decision making and that these changes in management decisions lead to improved outcomes. It is not clear whether the HCV FibroSURE could be used as an interval test in patients receiving therapy to determine whether an additional liver biopsy was necessary.

Alcoholic Liver Disease and Alcoholic Steatohepatitis
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

The diagnostic value of FibroSURE (FibroTest in Europe) has also been evaluated for the prediction of liver fibrosis in patients with ALD and NAFLD. Thabut et al (2006) reported the development of a panel of biomarkers (ASH FibroSURE [ASH Test]) for the diagnosis of alcoholic steatohepatitis (ASH) in patients with chronic ALD. Biomarkers were initially assessed in a training group of 70 patients, and a panel was constructed using a combination of the 6 biochemical components of the FibroSURE-ActiTest plus aspartate aminotransferase (AST). The algorithm was subsequently studied in 2 validation groups (1 prospective study for severe ALD, 1 retrospective study for nonsevere ALD) that included 155 patients and 299 controls. The severity
of ASH (none, mild, moderate, severe) was blindly assessed from biopsy samples. In the validation groups, there were 28 (18%) cases of discordance between the diagnosis of ASH predicted by the ASH Test and biopsy; 10 (36%) were considered false negatives of the ASH Test, and 11 were suspected failures of biopsy. Seven cases were indeterminate by biopsy. The AUROC curves were 0.88 and 0.89 in the validation groups. The median ASH Test value was 0.005 in controls, 0.05 in patients without or with mild ASH, 0.64 in the moderate ASH grade, and 0.84 in severe ASH grade 3. Using a cutoff value of 0.50, the ASH Test had a sensitivity of 80% and specificity of 84%, with PPVs and NPVs of 72% and 89%, respectively.

Several authors have an interest in the commercialization of this test, and no independent studies on the diagnostic accuracy of ASH FibroSURE (ASH Test) were identified. In addition, it is not clear if the algorithm used in this study is the same as that used in the currently commercially available test, which includes 10 biochemicals.

FibroTest has been studied in patients with ALD. In the Crossan (2015) systematic review, 1 study described the diagnostic accuracy of FibroTest for significant fibrosis (stage ≥ F2) or cirrhosis in ALD. With a high cutoff for positivity (0.7), the sensitivity and specificity for advanced fibrosis were 55% (95% CI, 47% to 63%) and 93% (95% CI, 85% to 97%) and for cirrhosis were 91% (95% CI, 82% to 96%) and 87% (95% CI, 81% to 91%), respectively. With a low cutoff for positivity (0.3), the sensitivity and specificity for advanced fibrosis were 84% (95% CI, 77% to 89%) and 65% (95% CI, 55% to 75%), respectively. The sensitivity and specificity for cirrhosis were 100% (95% CI, 95% to 100%) and 50% (95% CI, 42% to 58%), respectively.

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

No studies were identified that assessed clinical outcomes following the use of the ASH FibroSURE (ASH Test).

Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Poynard et al (2006) reported the development of a panel of biomarkers (NASH FibroSURE [NASH Test]) for the prediction of nonalcoholic steatohepatitis (NASH) in patients with NAFLD. Biomarkers were initially assessed with a training group of 160 patients, and a panel was constructed using a combination of 13 of 14 parameters of the currently available test. The algorithm was subsequently studied in a validation group of 97 patients and 383 controls. Patients in the validation group were from a prospective multicenter study with hepatic steatosis at biopsy and suspicion of NAFLD. Histologic diagnoses used Kleiner et al’s scoring system, with 3
classes for NASH (NASH, borderline NASH, no NASH). The main end point was steatohepatitis, defined as a histologic NASH score of 5 or greater. The AUROC curve for the validation group was 0.79 for the diagnosis of NASH, 0.69 for the diagnosis of borderline NASH, and 0.83 for the diagnosis of no NASH. Results showed a sensitivity of 33% and specificity of 94% for NASH, with a PPV and NPV of 66% and 81%, respectively. For borderline NASH or NASH, sensitivity was 88%, specificity 50%, PPV 74%, and NPV 72%. Clinically significant discordance (2 class difference) was observed in 8 (8%) patients. None of the 383 controls was considered to have NASH by NASH FibroSURE (NASH Test). Authors proposed that this test would be suitable for mass screening for NAFLD in patients with obesity and diabetes.

An independent study by Lassailly et al (2011) attempted to prospectively validate the NashTest (along with the FibroTest, SteatoTest, and ActiTest) in a cohort of 288 patients treated with bariatric surgery.24 Included were patients with severe or morbid obesity (body mass index, >35 kg/m²), at least 1 comorbidity for at least 5 years, and resistance to medical treatment. Excluded were patients with current excessive drinking, long-term consumption of hepatotoxic drugs, and positive screening for chronic liver diseases including hepatitis. Histology and biochemical measurements were centralized and blinded to other characteristics. The NashTest provided a 3-category score for no NASH (0.25), possible NASH (0.50), and NASH (0.75). The prevalence of NASH was 6.9%, while the prevalence of NASH or possible NASH was 27%. The concordance rate between histologic NASH score and the NashTest was 43.1%, with a weak κ reliability test (0.14). In 183 patients categorized as possible NASH by the NashTest, 124 (68%) were classified as no NASH by biopsy. In 15 patients categorized as NASH by the NashTest, 7 (47%) were no NASH and 4 (27%) were possible NASH by biopsy. The NPV of the NashTest for possible NASH or NASH was 47.5%. Authors suggested that the power of this study to validate agreement between the NashTest and biopsy was low, due to the low prevalence of NASH. However, the results showed poor concordance between the NashTest and biopsy, particularly for intermediate values.

In the Crossan (2015) systematic review, 4 studies were included in the pooled estimate of the diagnostic accuracy of FibroTest for advanced fibrosis (stage ≥ 3) in NAFLD.5 The summary sensitivities and specificities were 40% (95% CI, 24% to 58%) and 96% (95% CI, 91% to 98%), respectively. Only 1 study included reported accuracy for cirrhosis, with sensitivity and specificity of 74% (95% CI, 54%, to 87%) and 92% (95% CI, 88% to 95%), respectively.

**Clinically Useful**

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

**Direct Evidence:** Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

No studies were identified that assessed clinical outcomes following the use of the NASH FibroSURE (NASH Test).
Hepatitis B Virus

Clinically Valid

A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

While most multianalyte assay studies that have identified fibrosis have been conducted in patients with HCV, studies are also being conducted in patients with chronic HBV. In a study, Park et al (2013) compared liver biopsy with the FibroTest results obtained on the same day from 330 patients who had chronic HBV. Discordance was found in 30 (9.1%) patients for whom the FibroTest underestimated fibrosis in 25 patients and overestimated it in 5 patients. Those with Metavir liver fibrosis stage F3 or F4 (15.4%) had a significantly higher discordance rate than those with stages F1 or F2 (3.0%; p<0.001). The only independent factor for discordance on multivariate analysis was a Metavir stage F3 or F4 on liver biopsy (p<0.001).

Salkic et al (2014) conducted a meta-analysis of studies on the diagnostic accuracy of FibroTest in chronic HBV. Included in the meta-analysis were 16 studies (2494 patients) on liver fibrosis diagnosis and 13 studies (1754 patients) on cirrhosis diagnosis. There was strong evidence of heterogeneity in the 16 fibrosis studies and evidence of heterogeneity in the cirrhosis studies. For significant liver fibrosis (Metavir F2-F4) diagnosis using all of the fibrosis studies, the AUROC curve was 0.84 (95% CI, 0.78 to 0.88). At the recommended FibroTest threshold of 0.48 for a significant liver fibrosis diagnosis, the sensitivity was 60.9%, specificity was 79.9%, and the diagnostic odds ratio was 6.2. For liver cirrhosis (Metavir F4) diagnosis using all of the cirrhosis studies, the AUROC curve was 0.87 (95% CI, 0.85 to 0.9). At the recommended FibroTest threshold of 0.74 for cirrhosis diagnosis, the sensitivity was 61.5%, specificity was 90.8%, and the diagnostic odds ratio was 15.7. While the results demonstrated FibroTest may be useful in excluding a diagnosis of cirrhosis in patients with chronic HBV, the ability to detect significant fibrosis and cirrhosis and exclude significant fibrosis is suboptimal. Xu et al (2014) reported on a systematic review and meta-analysis of studies assessing biomarkers to detect fibrosis in HBV. Included in the analysis on FibroTest were 11 studies (total N=1640 patients). In these 11 studies, AUROC curves ranged from 0.69 to 0.90. Heterogeneity in the studies was statistically significant.

In the Crossan (2015) systematic review, 6 studies were included in the pooled estimate of the diagnostic accuracy of FibroTest for significant fibrosis (stage ≥ F2) in HBV. The cutoffs for positivity ranged from 0.40 to 0.48, and the summary sensitivities and specificities were 66% (95% CI, 57% to 75%) and 80% (95% CI, 72% to 86%), respectively. The accuracy for diagnosing cirrhosis in HBV was based on 4 studies with cutoffs for positivity ranging from 0.58 to 0.74; sensitivities and specificities were 74% (95% CI, 25% to 96%) and 90% (95% CI, 83% to 94%), respectively.

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.
Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

There are no studies evaluating the effect of this test on outcomes for patients with HBV. Of note, some researchers have suggested that different markers (eg, HBV FibroSURE) may be needed for this assessment in patients with hepatitis B.30

Section Summary: FibroSURE (FibroTest)
FibroSURE is the most widely validated of the noninvasive commercial serum tests. It has been studied in populations with viral hepatitis, NAFLD, and ALD. Although there are established cutoffs for positivity for FibroTest, they were not consistently used in validation studies. The methodologic quality of the validation studies was generally poor. There is no direct evidence that FibroSURE (FibroTest) improves health outcomes. However, there is indirect evidence: FibroTest has been used as an alternative to biopsy to establish trial eligibility in terms of fibrosis or cirrhosis in several RCTs that established the efficacy of HCV treatments.

Multianalyte Serum Assays: Other Than Fibrosure
Clinical Context and Test Purpose
The purpose of noninvasive testing in individuals with chronic liver disease is to detect liver fibrosis so that patients can avoid the potential adverse events of an invasive liver biopsy and receive appropriate treatment. The degree of liver fibrosis is an important factor in determining the appropriate approach for managing patients with liver disease (eg, hepatitis, ALD, NAFLD).

The question addressed in a portion of the evidence review is: Does the use of multianalyte serum assays other than FibroSURE for detecting liver fibrosis improve the net health outcome in patients with chronic liver disease?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with chronic liver disease.

Interventions
The tests being considered are multianalyte serum assays for liver function assessment other than FibroSURE. One test, FIBROSpect, consists of measurements of hyaluronic acid, tissue inhibitors of metalloproteinase 1, and α2-macroglobulin.

Comparators
The following tests and practices are currently being used to diagnose chronic liver disease: liver biopsy, noninvasive radiologic methods, and other multianalyte serum assays.

Outcomes
The general outcomes of interest are test validity, morbid events, and treatment-related morbidity.
Timing
Follow-up over months to years is of interest for the relevant outcomes.

Setting
Patients are actively managed by gastroenterologists and other specialists in an clinical setting.

Study Selection Criteria
For the evaluation of clinical validity of this test, studies that meet the eligibility criteria are outlined in indication 1.

FIBROSpect II
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Patel et al (2004) investigated the use of serum markers in an initial training set of 294 patients with HCV and further validated the resulting algorithm in a validation set of 402 patients.\(^3\)\(^1\) The algorithm was designed to distinguish between no or mild fibrosis (F0-F1) and moderate-to-severe fibrosis (F2-F4). With the prevalence of F2-F4 disease of 52% and a cutoff value of 0.36, the PPVs and NPVs were 74.3% and 75.8%, respectively. Using a FIBROSpect II cutoff score of 0.42, Christensen et al (2006) reported a sensitivity of 93%, specificity of 66%, the overall accuracy of 76%, and an NPV of 94% for advanced fibrosis in 136 patients with HCV.\(^3\)\(^2\)

The published studies for this combination of markers continue to focus on test characteristics such as sensitivity, specificity, and accuracy.\(^3\)\(^3\)\^-\(^3\)\(^5\) In Crossan et al (2015), the summary diagnostic accuracy for detecting significant fibrosis (stage ≥ F2) in 5 studies of HCV with FIBROSpect II, with cutoffs ranging from 42 to 72, was 78% (95% CI, 49% to 93%) and the summary specificity was 71% (95% CI, 59% to 80%).\(^5\)

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

The issues of effect on patient outcomes are similar to those discussed for the FibroSURE (FibroTest in Europe). No studies were identified in the published literature in which the results of the FIBROSpect test were actively used in the management of the patient.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Because the clinical validity of FIBROSpect has not been established, a chain of evidence supporting the clinical utility of this test for this population cannot be constructed.
Subsection Summary: FIBROSpect II

FIBROSpect II has been studied in populations with HCV. Cutoffs for positivity varied across studies and were not well validated. The methodologic quality of the validation studies was generally poor. There is no direct evidence that FIBROSpect II improves health outcomes.

Other Multianalyte Scoring Systems

Other scoring systems have been developed. For example, the APRI requires only the serum level of AST and the number of platelets, and uses a simple nonproprietary formula that can be calculated at the bedside to produce a score for the prediction of fibrosis. Using an optimized cutoff value derived from a training set and validation set of patients with HCV, authors have reported that the NPV for fibrosis was 86% and that the PPV was 88%. In Crossan et al (2015), APRI was frequently evaluated and has been tested in HCV, HBV, NAFLD, and ALD. The summary diagnostic accuracies are in Table 1.

Table 1. Diagnostic Accuracy for APRI

<table>
<thead>
<tr>
<th>Disease</th>
<th>Metavir Stage</th>
<th>Cutoff</th>
<th>Studies</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV ≥ F2 (significant)</td>
<td>Low: 0.4-0.7</td>
<td>47</td>
<td>82 (77 to 86)</td>
<td>57 (49 to 65)</td>
<td></td>
</tr>
<tr>
<td>HCV ≥ F2 (significant)</td>
<td>High: 1.5</td>
<td>36</td>
<td>39 (32 to 47)</td>
<td>92 (89 to 95)</td>
<td></td>
</tr>
<tr>
<td>HCV F4 (cirrhosis)</td>
<td>Low: 0.75-1</td>
<td>24</td>
<td>77 (73 to 81)</td>
<td>78 (74 to 81)</td>
<td></td>
</tr>
<tr>
<td>HCV F4 (cirrhosis)</td>
<td>High: 2</td>
<td>19</td>
<td>48 (41 to 56)</td>
<td>94 (91 to 95)</td>
<td></td>
</tr>
<tr>
<td>HBV ≥ F2 (significant)</td>
<td>Low: 0.4-0.6</td>
<td>8</td>
<td>80 (68 to 88)</td>
<td>65 (52 to 77)</td>
<td></td>
</tr>
<tr>
<td>HBV ≥ F2 (significant)</td>
<td>High: 1.5</td>
<td>6</td>
<td>37 (22 to 55)</td>
<td>93 (85 to 97)</td>
<td></td>
</tr>
<tr>
<td>HBV F4 (cirrhosis)</td>
<td>Low: 1</td>
<td>4</td>
<td>58 (49 to 66)</td>
<td>76 (70 to 81)</td>
<td></td>
</tr>
<tr>
<td>HBV F4 (cirrhosis)</td>
<td>High: 2</td>
<td>3</td>
<td>24 (8 to 52)</td>
<td>91 (83 to 96)</td>
<td></td>
</tr>
<tr>
<td>NAFLD ≥ F3 (significant)</td>
<td>0.5 to 1.0</td>
<td>4</td>
<td>40 (7 to 86)</td>
<td>82 (78 to 60)</td>
<td></td>
</tr>
<tr>
<td>NAFLD F4 (cirrhosis)</td>
<td>0.54 and NA</td>
<td>2</td>
<td>78 (71 to 99)</td>
<td>71 (30 to 93)</td>
<td></td>
</tr>
<tr>
<td>ALD ≥ F2 (significant)</td>
<td>Low: 0.5</td>
<td>2</td>
<td>72 (60 to 82)</td>
<td>46 (33 to 60)</td>
<td></td>
</tr>
<tr>
<td>ALD ≥ F2 (significant)</td>
<td>High: 1.5</td>
<td>2</td>
<td>54 (42 to 66)</td>
<td>78 (64 to 88)</td>
<td></td>
</tr>
<tr>
<td>ALD F4 (cirrhosis)</td>
<td>High: 2.0</td>
<td>1</td>
<td>40 (22 to 61)</td>
<td>62 (41 to 79)</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Crossan et al (2015). ALD: alcoholic liver disease; APRI: aspartate aminotransferase–platelet ratio index; CI: confidence interval; HBV: hepatitis B virus; HCV: hepatitis C virus; NA: not available; NAFLD: nonalcoholic fatty liver disease.

Giannini et al (2006) reported that use of the AST/ALT ratio and platelet counts in a diagnostic algorithm would have avoided liver biopsy in 69% of their patients and would have correctly identified the absence or presence of significant fibrosis in 80.5% of these cases. In Crossan et al (2015), the cutoffs for positivity of AST/ALT ratio for diagnosis of significant fibrosis (stage ≥ F2) varied from 0.6 to 1 in 7 studies. Summary sensitivity and specificity were 44% (95% CI, 27% to 63%) and 71% (95% CI, 62% to 78%), respectively. Thirteen studies used a cutoff of 1 to estimate the diagnostic accuracy of cirrhosis with AST/ALT ratio, and summary sensitivity and specificity were 49% (95% CI, 39% to 59%) and 87% (95% CI, 75% to 94%), respectively.

A number of studies have compared HCV FibroSURE (FibroTest) and other noninvasive tests of fibrosis with biopsy using ROC analysis. For example, Bourliere et al (2006) reported on the validation of FibroSURE (FibroTest) and found that, based on ROC analysis, FibroSURE (FibroTest) was superior to APRI for identifying significant fibrosis, with AUROC curves of 0.81 and 0.71, respectively. A prospective multicenter study by Zarksi et al (2012) compared 9 of the best-evaluated blood tests in 436 patients with hepatitis C and found similar performance for HCV FibroSURE (FibroTest), FibroMeter, and HepaScore (ROC curve, 0.84, 0.86, 0.84, respectively).
These 3 tests were significantly superior to the 6 other tests, with 70% to 73% of patients considered well classified according to a dichotomized score (F0/F1 vs ≥F2). The number of “theoretically avoided liver biopsies” for the diagnosis of significant fibrosis was calculated to be 35.6% for HCV FibroSURE (FibroTest). To improve diagnostic accuracy, algorithms that combine HCV FibroSURE (FibroTest) with other tests (eg, APRI) are also being evaluated. One of these, the sequential algorithm for fibrosis evaluation, combines the APRI and FibroTest. Crossan et al (2015) reported that the algorithm has been assessed in 4 studies of HCV for diagnosing both significant fibrosis (stage ≥ F2) and cirrhosis. Summary sensitivity and specificity for significant fibrosis were estimated to be 100% (95% CI, 100% to 100%) and 81% (95% CI, 80% to 83%), respectively. The summary sensitivity and specificity for cirrhosis were 74% (95% CI, 42% to 92%) and 93% (95% CI, 91% to 94%), respectively.

Rosenberg et al (2004) developed a scoring system based on an algorithm combining hyaluronic acid, amino-terminal propeptide of type III collagen, and tissue inhibitors of metalloproteinase 1. The algorithm was developed in a test set of 400 patients with a wide variety of chronic liver diseases and then validated in another 521 patients. The algorithm was designed to discriminate between no or mild fibrosis and moderate-to-severe fibrosis. The NPV for fibrosis was 92%.

Section Summary: Multianalyte Serum Assays Other Than FibroSURE
For multianalyte serum assays other than FibroSURE (eg, FIBROSpect II), there are a number of studies; however, all studies have included varying cutoffs, some of which were standardized and others not validated. Given these limitations and the imperfect reference standard, it is difficult to interpret performance characteristics. There is no direct evidence that other multianalyte serum assays improve health outcomes and a chain of evidence cannot be constructed given the inadequate data on clinical validity.

Noninvasive Imaging: Transient Elastography
Clinical Context and Test Purpose
The purpose of noninvasive testing in individuals with chronic liver disease is to detect liver fibrosis so that patients can avoid the potential adverse events of an invasive liver biopsy and receive appropriate treatment. The degree of liver fibrosis is an important factor in determining the appropriate approach for managing patients with liver disease (eg, hepatitis, ALD, NAFLD).

The question addressed in this portion of the evidence review is: Does the use of transient elastography for detecting liver fibrosis improve the net health outcome in patients with chronic liver disease?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with chronic liver disease.

Interventions
The therapy being considered is transient elastography (eg, FibroScan).
Comparators
The following tests and practices are currently being used to diagnose chronic liver disease: liver biopsy, other noninvasive radiologic methods, and multianalyte serum assays.

Outcomes
The general outcomes of interest are test validity, morbid events, and treatment-related morbidity.

Timing
Follow-up over months to years is of interest for the relevant outcomes.

Setting
Patients are actively managed by gastroenterologists and other specialists in an outpatient setting.

Study Selection Criteria
For the evaluation of clinical validity of this test, studies that meet the eligibility criteria are outlined in indication 1.

Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

There is extensive literature on the use of transient elastography to gauge liver fibrosis and cirrhosis. Summaries of systematic reviews are shown in Tables 2 and 3. Brener (2015) performed a health technology assessment summarizing many of the systematic reviews below.43 The assessment focused on reviews of the diagnostic accuracy and effect on patient outcomes of transient elastography for liver fibrosis in patients with HCV, HBV, NAFLD, ALD, or cholestatic diseases. Fourteen systematic reviews of transient elastography with biopsy reference standard shown below were included in the Brener assessment, summarizing more than 150 primary studies.44-57 There was variation in the underlying cause of liver disease and the cutoff values of transient elastography stiffness used to define Metavir stages in the systematic reviews. There did not appear to be a substantial difference in diagnostic accuracy for one disease over any other. The reviews demonstrated that transient elastography has good diagnostic accuracy compared with biopsy for the assessment of liver fibrosis and steatosis.

Crossan et al (2015) found that FibroScan was the noninvasive liver test most assessed in validation studies across liver diseases (37 studies in HCV, 13 in HBV, 8 in NAFLD, 6 in ALD).5 Cutoffs for positivity for fibrosis staging varied between diseases and were frequently not prespecified or validated: HCV, 5.2 to 10.1 kPa in the 37 studies for Metavir stages ≥ F2; HBV, 6.3 to 8.9 kPa in 13 studies for stages ≥ F2; NAFLD, 7.5 to 10.4 kPa in 8 studies for stages ≥ F3; ALD, 11.0 to 12.5 in 4 studies for stages ≥ F3. Summary sensitivities and specificities by disease are shown in Table 3. The overall sensitivity and specificity for cirrhosis including all diseases (65 studies; cutoffs range, 9.2-26.5 kPa) were 89% (95% CI, 86% to 91%) and 89% (95% CI, 87% to 91%), respectively. The rate of uninterpretable results, when reported, with FibroScan (due to <10 valid measurements; success rate, <60%; interquartile range, >30%) was 8.5% in HCV and 9.6% in NAFLD.
Table 2. Transient Elastography Systematic Review Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Dates</th>
<th>Studies</th>
<th>N</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bota et al (2013)</td>
<td>To May 2012</td>
<td>13</td>
<td>1163</td>
<td>Chronic hepatitis</td>
</tr>
<tr>
<td>Friedric-Rust et al (2008)</td>
<td>2002 to Apr 2007</td>
<td>50</td>
<td>11,275</td>
<td>All causes of liver disease</td>
</tr>
<tr>
<td>Friedric-Rust et al (2012)</td>
<td>To Oct 2010</td>
<td>8</td>
<td>518</td>
<td>All causes of liver disease</td>
</tr>
<tr>
<td>Geng et al (2016)</td>
<td>To Jan 2015</td>
<td>57</td>
<td>10,569</td>
<td>Multiple causes of liver disease</td>
</tr>
<tr>
<td>Njei et al (2016)</td>
<td>To Jan 2016</td>
<td>6</td>
<td>756</td>
<td>HCV/HIV coinfection</td>
</tr>
<tr>
<td>Shi et al (2014)</td>
<td>To May 2013</td>
<td>9</td>
<td>1771</td>
<td>All causes of steatosis</td>
</tr>
<tr>
<td>Steadman et al (2013)</td>
<td>2001 to Jun 2011</td>
<td>64</td>
<td>6028</td>
<td>HCV, HBV, NAFLD, CLD, liver transplant</td>
</tr>
<tr>
<td>Stebbing et al (2010)</td>
<td>NR, prior to Feb 2009</td>
<td>22</td>
<td>4625</td>
<td>All causes of liver disease</td>
</tr>
<tr>
<td>Talwalkar et al (2007)</td>
<td>9</td>
<td>2083</td>
<td>All causes of liver disease</td>
<td></td>
</tr>
<tr>
<td>Tschochatzis et al (2011)</td>
<td>To May 2009</td>
<td>40</td>
<td>7661</td>
<td>All causes of liver disease</td>
</tr>
<tr>
<td>Xu et al (2015)</td>
<td>To Dec 2013</td>
<td>19</td>
<td>3113</td>
<td>HBV</td>
</tr>
</tbody>
</table>

ALD: alcoholic liver disease; CLD: chronic liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus; NAFLD: nonalcoholic fatty liver disease; NR: not reported.

Table 3. Transient Elastography Systematic Reviews Diagnostic Accuracy Results

<table>
<thead>
<tr>
<th>Study</th>
<th>Significant Fibrosis (ie, Metavir Stages F2-F4)</th>
<th>Cirrhosis (ie, Metavir Stage F4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td>Studies/ Sample Size</td>
</tr>
<tr>
<td>Bota et al (2013)</td>
<td>Multiple diseases</td>
<td>10/1016</td>
</tr>
<tr>
<td>Chon et al (2012)</td>
<td>Chronic HBV</td>
<td>12/2000</td>
</tr>
<tr>
<td>Crossan et al (2015)</td>
<td>HCV</td>
<td>37/NR</td>
</tr>
<tr>
<td></td>
<td>HBV</td>
<td>13/NR</td>
</tr>
<tr>
<td></td>
<td>NAFLD</td>
<td>4/NR</td>
</tr>
<tr>
<td>Study</td>
<td>Significant Fibrosis (ie, Metavir Stages F2-F4)</td>
<td>Cirrhosis (ie, Metavir Stage F4)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Friedrich-Rust (2008)&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Multiple diseases 25/3685 0.84 (0.82 to 0.86) 87% (0.93 to 0.95)</td>
<td>NR 25/4557 94% (NR)</td>
</tr>
<tr>
<td>Friedrich-Rust et al (2012)&lt;sup&gt;66&lt;/sup&gt;</td>
<td>No summary statistics reported for transient elastography</td>
<td></td>
</tr>
<tr>
<td>Geng et al (2016)&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Multiple diseases NR 0.93 (0.82 to 0.96) 81% (79% to 83%) 88% (87% to 96%)</td>
<td>NR</td>
</tr>
<tr>
<td>Jiang et al (2018)&lt;sup&gt;59&lt;/sup&gt;</td>
<td>NAFLD 10/NR 0.85 (0.82 to 0.88) 90% (0.93 to 0.97)</td>
<td>NR</td>
</tr>
<tr>
<td>Kwok et al (2014)&lt;sup&gt;68&lt;/sup&gt;</td>
<td>NAFLD 7/800 0.83 (0.79 to 0.87) 92% (82% to 97%)</td>
<td>24/NR 94% (NR) 88% (84% to 90%)</td>
</tr>
<tr>
<td>Li et al (2016)&lt;sup&gt;60&lt;/sup&gt;</td>
<td>HBV 19/NR 0.88 (0.85 to 0.91) 86% (82% to 90%)</td>
<td>77% (71% to 87%)</td>
</tr>
<tr>
<td>Njei et al (2016)&lt;sup&gt;61&lt;/sup&gt;</td>
<td>HCV/HIV 6/756 NR 90% (74% to 91%)</td>
<td>6/756 95% (NR)</td>
</tr>
<tr>
<td>Pavlov et al (2015)&lt;sup&gt;62&lt;/sup&gt;</td>
<td>ALD 7/338 NR 95% (87% to 98%)</td>
<td>7/330 92% (56% to 82%)</td>
</tr>
<tr>
<td>Poynard et al (2008)&lt;sup&gt;49&lt;/sup&gt;</td>
<td>No summary statistics reported for transient elastography</td>
<td></td>
</tr>
<tr>
<td>Poynard et al (2011)&lt;sup&gt;50&lt;/sup&gt;</td>
<td>HBV 4/NR 0.93 (0.87 to 0.99)</td>
<td>86% (82% to 89%) 89% (NR)</td>
</tr>
<tr>
<td>Shaheen et al (2007)&lt;sup&gt;51&lt;/sup&gt;</td>
<td>HCV 4/NR 0.84 (0.78 to 0.89) 86% (82% to 90%)</td>
<td>77% (71% to 87%)</td>
</tr>
<tr>
<td>Shi et al (2014)&lt;sup&gt;62&lt;/sup&gt;</td>
<td>No summary statistics reported. Concluded that transient elastography controlled attenuation parameter has good sensitivity and specificity for diagnosing steatosis, but it has limited utility</td>
<td></td>
</tr>
<tr>
<td>Steadman et al (2013)&lt;sup&gt;53&lt;/sup&gt;</td>
<td>Multiple diseases 45/NR 0.88 (0.84 to 0.90) 90% (0.91 to 0.96)</td>
<td>96% (NR) 86% (82% to 89%)</td>
</tr>
<tr>
<td></td>
<td>HBV 5/710 0.81 (0.78 to 0.84) 86% (87% to 91%)</td>
<td>8/1092 0.86 (NR)</td>
</tr>
<tr>
<td></td>
<td>HCV 13/2732 0.89 (0.86 to 0.91) 86% (87% to 91%)</td>
<td>12/2887 0.94 (NR)</td>
</tr>
<tr>
<td>Stebbing et al (2010)&lt;sup&gt;54&lt;/sup&gt;</td>
<td>Multiple diseases 17/3066 0.78 (0.74 to 0.82) 92% (77% to 98%)</td>
<td>17/4052 95% (NR)</td>
</tr>
<tr>
<td></td>
<td>NAFLD 5/630 0.78 (0.74 to 0.82) 92% (77% to 98%)</td>
<td>4/469 95% (NR)</td>
</tr>
<tr>
<td></td>
<td>Multiple diseases 7/&gt;1100 0.87 (0.83 to 0.91) 96% (0.94 to 0.98)</td>
<td>9/2083 96% (84% to 90%)</td>
</tr>
</tbody>
</table>
### Study

<table>
<thead>
<tr>
<th>Study</th>
<th>Significant Fibrosis (ie, Metavir Stages F2-F4)</th>
<th>Cirrhosis (ie, Metavir Stage F4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsochatzis et al (2011)³⁵</td>
<td>Multiple diseases 31/5919 84% (80% to 88%) 79% (74% to 82%) 78% (72% to 83%)</td>
<td>30/6530 91% (89% to 92%) 83% (79% to 86%) 89% (87% to 91%)</td>
</tr>
<tr>
<td></td>
<td>HCV 14/NR NR 78% (71% to 84%) 80% (71% to 86%)</td>
<td>11/NR NR 83% (77% to 88%) 90% (87% to 93%)</td>
</tr>
<tr>
<td></td>
<td>HBV 4/NR NR 84% (67% to 93%) 78% (68% to 85%)</td>
<td>6/NR NR 80% (61% to 91%) 86% (82% to 94%)</td>
</tr>
<tr>
<td>Tsochatzis et al (2014)³⁷</td>
<td>HCV 37/NR 0.87 (0.83 to 0.90) 79% (74% to 84%) 83% (77% to 88%)</td>
<td>36/NR 0.96 (0.94 to 0.97) 89% (84% to 92%) 91% (89% to 93%)</td>
</tr>
<tr>
<td></td>
<td>HBV 13/NR 0.83 (0.76 to 0.90) 71% (62% to 78%) 84% (74% to 91%)</td>
<td>13/NR 0.92 (0.89 to 0.96) 86% (79% to 91%) 85% (78% to 89%)</td>
</tr>
<tr>
<td></td>
<td>NAFLD 4/NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALD 6/NR 0.90 (0.87 to 0.94) 86% (76% to 92%) 83% (74% to 89%)</td>
<td></td>
</tr>
<tr>
<td>Xu et al (2015)³¹</td>
<td>HBV 14/2318 0.82 (0.78 to 0.86) NR</td>
<td>18/2996 0.91 (0.89 to 0.93) NR</td>
</tr>
</tbody>
</table>

ALD: alcoholic liver disease; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; HBV: hepatitis B virus; HCV: hepatitis C virus; NAFLD: nonalcoholic fatty liver disease; NR: not reported.

### Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

### Direct Evidence

Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

There are currently no published studies that directly demonstrate the effect of transient elastography (eg, FibroScan) on patient outcomes.

FibroScan is used extensively in practice to make management decisions. In addition, FibroScan was used as an alternative to biopsy for to diagnose fibrosis or cirrhosis to establish trial eligibility in several trials (ION-1,-3; VALENCE; ASTRAL-2, -3, -4) that confirmed the efficacy of HCV treatments.¹⁴⁻¹⁹ For example, in the VALENCE trial, cirrhosis could be defined by liver biopsy or a confirmatory FibroTest or FibroScan result at 12.5 kPa or greater. In VALENCE, FibroScan was used to determine cirrhosis in 74% of the participants.
Section Summary: Transient Elastography (FibroScan)
Transient elastography (FibroScan) is the most widely validated of the noninvasive methods. FibroScan has been studied in populations with viral hepatitis, NAFLD, and ALD. FibroScan validation studies have suggested that it can provide good detection of significant fibrosis and good-to-excellent detection of cirrhosis compared with liver biopsy for HCV and HBV. There are limited data on NAFLD and ALD. There are no established or validated cutoffs, and the quality of the validation studies was generally not high. Failures of the test are not uncommon, particularly for those with high body mass index; however, failures were frequently missed in analyses of the validation studies. Newer more sensitive probes may lessen this limitation. There is no direct evidence that FibroScan improves health outcomes. However, FibroScan has been used as an alternative to biopsy to diagnose fibrosis or cirrhosis to establish trial eligibility in several RCTs that established the efficacy of HCV treatments.

Other Noninvasive Imaging
The following noninvasive imaging types are evaluated in this section: magnetic resonance elastography (MRE), ARFI imaging (eg, Acuson S2000), and real-time tissue elastography (RTE; eg, HI VISION Preirus).

Clinical Context and Test Purpose
The purpose of noninvasive testing in individuals with chronic liver disease is to detect liver fibrosis so that patients can avoid the potential adverse events of an invasive liver biopsy and receive appropriate treatment. The degree of liver fibrosis is an important factor in determining the appropriate approach for managing patients with liver disease (eg, hepatitis, ALD, NAFLD).

The question addressed in this portion of the evidence review is: Does the use of noninvasive imaging other than transient elastography for detecting liver fibrosis improve the net health outcome in patients with chronic liver disease?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with chronic liver disease.

Interventions
The tests being considered are noninvasive radiologic methods other than transient elastography for liver fibrosis measurement (eg, MRE, ARFI imaging, RTE).

Comparators
The following tests and practices are currently being used to diagnose chronic liver disease: liver biopsy, other noninvasive radiologic methods, and multianalyte serum assays.

Outcomes
The general outcomes of interest are test validity, morbid events, and treatment-related morbidity.

Timing
Follow-up over months to years is of interest for the relevant outcomes.
Noninvasive Techniques for the Evaluation and Monitoring of Patients with Chronic Liver Disease

Setting
Patients are actively managed by gastroenterologists and other specialists in an outpatient setting.

Study Selection Criteria
For the evaluation of clinical validity of this test, studies that meet the eligibility criteria are outlined in indication 1.

ARFI Imaging
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Tables 4 and 5 summarize the characteristics and results of systematic reviews that have assessed the diagnostic accuracy of ARFI imaging.

Table 4. Characteristics of Systematic Reviews Assessing Acoustic Radiation Force Impulse Imaging

<table>
<thead>
<tr>
<th>Study</th>
<th>Dates</th>
<th>Studies</th>
<th>N</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bota et al (2013)</td>
<td>To May 2012</td>
<td>6</td>
<td>518</td>
<td>Chronic hepatitis</td>
</tr>
<tr>
<td>Hu et al (2017)</td>
<td>To Jul 2014</td>
<td>f7</td>
<td>723</td>
<td>Nonalcoholic fatty liver disease</td>
</tr>
<tr>
<td>Jiang et al (2018)</td>
<td>To Dec 2017</td>
<td>9</td>
<td>982</td>
<td>Nonalcoholic fatty liver disease</td>
</tr>
<tr>
<td>Liu et al (2015)</td>
<td>To Apr 2016</td>
<td>23</td>
<td>2691</td>
<td>Chronic hepatitis B or C</td>
</tr>
</tbody>
</table>

Table 5. Results of Systematic Reviews Assessing the Diagnostic Accuracy of Acoustic Radiation Force Impulse Imaging

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Significant Fibrosis (ie, Metavir Stages F2-F4)</th>
<th>Cirrhosis (ie, Metavir Stage F4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Studies/ Sample Size</td>
<td>AUROC (95% CI)</td>
</tr>
<tr>
<td>Bota et al (2013)</td>
<td>Chronic hepatitis</td>
<td>6/518</td>
<td>0.88 (0.83 to 0.93)</td>
</tr>
<tr>
<td>Crossan et al (2015)</td>
<td>HCV</td>
<td>4/NR</td>
<td>NR</td>
</tr>
<tr>
<td>Guo et al (2015)</td>
<td>Multiple diseases</td>
<td>13/NR</td>
<td>NR</td>
</tr>
<tr>
<td>Hu et al (2017)</td>
<td>HBV, HCV</td>
<td>15/NR</td>
<td>NR</td>
</tr>
<tr>
<td>Jiang et al (2018)</td>
<td>NAFLD</td>
<td>6/NR</td>
<td>0.86 (0.83 to 0.89)</td>
</tr>
<tr>
<td>Liu et al (2015)</td>
<td>NAFLD</td>
<td>7/723</td>
<td>NR</td>
</tr>
</tbody>
</table>
Study | Population | Significant Fibrosis (ie, Metavir Stages F2-F4) | Cirrhosis (ie, Metavir Stage F4) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nierhoff et al (2013)</td>
<td>Multiple diseases</td>
<td>26/NR 0.83 (0.80 to 0.86)</td>
<td>27/NR 0.91 (0.89 to 0.93)</td>
</tr>
</tbody>
</table>

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

There are currently no published studies that directly demonstrate the effect of ARFI imaging on patient outcomes.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Because the clinical validity of ARFI imaging has not been established, a chain of evidence supporting the clinical utility of this test for this population cannot be constructed.

Subsection Summary: ARFI Imaging
The use of ARFI imaging has been evaluated in viral hepatitis and NAFLD. Moreover, many have noted that ARFI imaging has potential advantages over FibroScan—it can be implemented on a standard ultrasound machine, may be more applicable for assessing complications such as ascites and may be more applicable in obese patients. ARFI imaging appears to have similar diagnostic accuracy to FibroScan, but there are fewer data available on performance characteristics. Validation studies have used varying cutoffs for positivity.

Magnetic Resonance Elastography
Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Tables 6 and 7 summarize the characteristics and results of systematic reviews that have assessed the diagnostic accuracy of MRE. MRE has been studied primarily in hepatitis and NAFLD.

Table 6. Characteristics of Systematic Reviews Assessing Magnetic Resonance Elastography
<table>
<thead>
<tr>
<th>Study</th>
<th>Dates</th>
<th>Studies</th>
<th>N</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossan et al (2015)</td>
<td>1998 to Apr 2012</td>
<td>3</td>
<td>Not reported</td>
<td>Chronic liver disease</td>
</tr>
<tr>
<td>Singh et al (2015)</td>
<td>2003 to Sep 2013</td>
<td>12</td>
<td>697</td>
<td>Chronic liver disease</td>
</tr>
<tr>
<td>Singh et al (2016)</td>
<td>To Oct 2014</td>
<td>9</td>
<td>232</td>
<td>Nonalcoholic fatty liver disease</td>
</tr>
</tbody>
</table>
Table 7. Results of Systematic Reviews Assessing the Diagnostic Accuracy of Magnetic Resonance Elastography

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Significant Fibrosis (ie, Stages F2-F4)</th>
<th>Cirrhosis (ie, Stage F4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC (95% CI)</td>
<td>Sensitivity (95% CI)</td>
<td>Specificity (95% CI)</td>
</tr>
<tr>
<td></td>
<td>Studies/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crossan et al (2015)</td>
<td>Chronic liver disease</td>
<td>0.94 (13% to 100%)</td>
<td>92% (72% to 98%)</td>
</tr>
<tr>
<td>Guo et al (2015)</td>
<td>Multiple</td>
<td>0.87 (84% to 90%)</td>
<td>91% (88% to 93%)</td>
</tr>
<tr>
<td>Singh et al (2015)</td>
<td>Chronic hepatitis</td>
<td>0.84 (0.76 to 0.92)</td>
<td>91% (0.90 to 0.94)</td>
</tr>
<tr>
<td>Singh et al (2016)</td>
<td>NAFLD</td>
<td>0.87 (0.82 to 0.93)</td>
<td>91% (0.76 to 0.95)</td>
</tr>
</tbody>
</table>

AUROC: area under the receiver operating characteristic curve; CI: confidence interval; NAFLD: nonalcoholic fatty liver disease; NR: not reported.

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTS.

There are currently no published studies that directly demonstrate the effect of MRE on patient outcomes.

Chain of Evidence: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Because the clinical validity of MRE has not been established, a chain of evidence supporting the clinical utility of this test for this population cannot be constructed.

Subsection Summary: Magnetic Resonance Elastography

MRE has a high success rate and is highly reproducible across operators and time. The diagnostic accuracy also appears to be high. In particular, MRE has high diagnostic accuracy for detection of fibrosis in NAFLD, independent of body mass index and degree of inflammation. However, further validation is needed to determine standard cutoffs and confirm performance characteristics because confidence intervals for estimates are wide. MRE is not widely available.

Real-Time Tissue Elastography (HI VISION 15 Preirus)

Clinically Valid

A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).
Kobayashi et al (2015) published results of a meta-analysis assessing RTE for staging liver fibrosis. They selected 15 studies (total N=1626 patients) published through December 2013, including patients with multiple liver diseases and healthy adults. A bivariate random-effects model was used to estimate summary sensitivity and specificity. The summary AUROC, sensitivity, and specificity were 0.69 (precision NR), 79% (95% CI, 75% to 83%), and 76% (95% CI, 68% to 82%) for detection of significant fibrosis (stage ≥ F2) and 0.72 (precision NR), 74% (95% CI, 63% to 82%), and 84% (95% CI, 79% to 88%) for detection of cirrhosis, respectively. Reviewers found evidence of heterogeneity due to differences in study populations, scoring methods, and cutoffs for positivity. They also found evidence of publication bias based on funnel plot asymmetry.

Hong et al (2014) reported on the results of a meta-analysis evaluating RTE for staging fibrosis in multiple diseases. Thirteen studies (total N=1347 patients) published between April 2000 and April 2014 that used liver biopsy or transient elastography as the reference standard were included. Different quantitative methods were used to measure liver stiffness in the included studies: Liver Fibrosis Index (LFI), Elasticity Index, elastic ratio 1 (ER1), and elastic ratio 2. For predicting significant fibrosis (stage ≥ F2), the pooled sensitivities for LFI and ER1 were 78% (95% CI, 70% to 84%) and 86% (95% CI, 80% to 90%), respectively. The specificities were 63% (95% CI, 46% to 78%) and 89% (95% CI, 83% to 94%) and the AUROCs were 0.79 (95% CI, 0.75 to 0.82) and 0.94 (95% CI, 0.92 to 0.96), respectively. For predicting cirrhosis (stage F4), the pooled sensitivities of LFI, ER1, and elastic ratio 2 were 79% (95% CI, 61% to 91%), 96% (95% CI, 87% to 99%), and 79% (95% CI, 61% to 91%), respectively. The specificities were 88% (95% CI, 81% to 93%) for LFI, 89% (95% CI, 83% to 93%) for ER1, and 88% (95% CI, 81% to 93%) for elastic ratio 2, and the AUROCs were 0.85 (95% CI, 0.81 to 0.87), 0.93 (95% CI, 0.94 to 0.98), and 0.92 (95% CI NR), respectively. Pooled estimates for Elasticity Index were not performed due to insufficient data.

**Clinically Useful**
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

**Direct Evidence**: Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

There are currently no published studies that directly demonstrate the effect of RTE on patient outcomes.

**Chain of Evidence**: Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Because the clinical validity of RTE has not been established, a chain of evidence supporting the clinical utility of this test for this population cannot be constructed.
**Subsection Summary: Real-Time Tissue Elastography**
RTE has been evaluated in multiple diseases with varying scoring methods and cutoffs. Although data are limited, the accuracy of RTE appears to be similar to FibroScan for the evaluation of significant liver fibrosis, but less accurate for the evaluation of cirrhosis. However, there was evidence of publication bias in the systematic review and the diagnostic accuracy may be overestimated.

**Section Summary: Noninvasive Radiological Methods Other Than Transient Elastography**
The available studies have suggested that other radiologic methods (AFRI, MRE, RTE) may have similar performance for detecting significant fibrosis or cirrhosis. However, the studies frequently included varying cutoffs not prespecified or validated. Given these limitations and the imperfect reference standard, it is difficult to interpret performance characteristics. There is no direct evidence that other noninvasive radiologic methods improve health outcomes and an indirect chain cannot be constructed due to the lack of sufficient evidence on clinical validity.

**SUMMARY OF EVIDENCE**

**Multianalyte Serum Assays**
For individuals who have chronic liver disease who receive FibroSURE serum panels, the evidence includes systematic reviews of more than 30 observational studies (>5000 patients). Relevant outcomes are test validity, morbid events, and treatment-related morbidity. FibroSURE has been studied in populations with viral hepatitis, nonalcoholic fatty liver disease, and alcoholic liver disease. There are established cutoffs, although they were not consistently used in validation studies. Given these limitations and the imperfect reference standard, it is difficult to interpret performance characteristics. However, for the purposes of deciding whether a patient has severe fibrosis or cirrhosis, FibroSURE results provide data sufficiently useful to determine therapy. Specifically, FibroSURE has been used as an alternative to biopsy to establish eligibility regarding the presence of fibrosis or cirrhosis in several randomized controlled trials that showed the efficacy of hepatitis C virus treatments, which in turn demonstrated the test can identify patients who would benefit from therapy. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

For individuals who have chronic liver disease who receive multianalyte serum assays for liver function assessment other than FibroSURE, the evidence includes systematic reviews of observational studies. Relevant outcomes are test validity, morbid events, and treatment-related morbidity. Studies have frequently included varying cutoffs, some of which were standardized and others not validated. Given these limitations and the imperfect reference standard, it is difficult to interpret performance characteristics. There is no direct evidence that other multianalyte serum assays improve health outcomes; further, it is not possible to construct a chain of evidence for clinical utility due to the lack of sufficient evidence on clinical validity. The evidence is insufficient to determine the effects of the technology on health outcomes.

**Noninvasive Imaging**
For individuals who have chronic liver disease who receive transient elastography, the evidence includes many systematic reviews of more than 50 observational studies (>10,000 patients). Relevant outcomes are test validity, morbid events, and treatment-related morbidity. Transient elastography (FibroScan) has been studied in populations with viral hepatitis, nonalcoholic fatty liver disease, and alcoholic liver disease. There are varying cutoffs for positivity. Failures of the
test are not uncommon, particularly for those with high body mass index, but these failures often went undetected in analyses of the validation studies. Given these limitations and the imperfect reference standard, it can be difficult to interpret performance characteristics. However, for the purposes of deciding whether a patient has severe fibrosis or cirrhosis, the FibroScan results provide data sufficiently useful to determine therapy. In fact, FibroScan has been used as an alternative to biopsy to establish eligibility regarding the presence of fibrosis or cirrhosis in the participants of several randomized controlled trials. These trials showed the efficacy of hepatitis C virus treatments, which in turn demonstrated that the test can identify patients who would benefit from therapy. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

For individuals who have chronic liver disease who receive noninvasive radiologic methods other than transient elastography for liver fibrosis measurement, the evidence includes systematic reviews of observational studies. Relevant outcomes are test validity, morbid events, and treatment-related morbidity. Other radiologic methods (eg, magnetic resonance elastography, real-time transient elastography, acoustic radiation force impulse imaging) may have similar performance for detecting significant fibrosis or cirrhosis. Studies have frequently included varying cutoffs not prespecified or validated. Given these limitations and the imperfect reference standard, it is difficult to interpret performance characteristics. There is no direct evidence that other noninvasive radiologic methods improve health outcomes; further, it is not possible to construct a chain of evidence for clinical utility due to the lack of sufficient evidence on clinical validity. The evidence is insufficient to determine the effects of the technology on health outcomes.

**CLINICAL INPUT RECEIVED FROM PHYSICIAN SPECIALTY SOCIETIES AND ACADEMIC MEDICAL CENTERS**

While the various physician specialty societies and academic medical centers may collaborate with and make recommendations during this process through the provision of appropriate reviewers, input received does not represent an endorsement or position statement by the physician specialty societies or academic medical centers, unless otherwise noted.

In response to requests, input was received from 3 physician specialty societies and 3 academic medical centers while this policy was under review in 2015. Most reviewers considered noninvasive techniques for the evaluation and monitoring of chronic liver disease to be investigational, both individually and in combination.

**PRACTICE GUIDELINES AND POSITION STATEMENTS**

**Nonalcoholic Fatty Liver Disease**

American Gastroenterological Association et al

The practice guidelines on the diagnosis and management of nonalcoholic fatty liver disease (NAFLD), developed by the American Gastroenterological Association, the American Association for the Study of Liver Diseases, and the American College of Gastroenterology (2018) stated that “NFS [NAFLD fibrosis score] or FIB-4 [Fibrosis-4] index are clinically useful tools for identifying NAFLD patients with higher likelihood of having bridging fibrosis (stage 3) or cirrhosis (stage 4).” It also cited VCTE [vibration-controlled transient elastography] and MRE [magnetic resonance elastography] as “clinically useful tools for identifying advanced fibrosis in patients with NAFLD.”
National Institute for Health and Care Excellence
The National Institute for Health and Care Excellence (NICE; 2016) published guidance on the assessment and management of NAFLD. The guidance did not reference elastography or multianalyte assays with algorithmic analyses. The guidance recommended the enhanced liver fibrosis test to test for advanced liver fibrosis.

American College of Gastroenterology Institute
The American College of Gastroenterology Institute (2017) published guidelines on the role of elastography in chronic liver disease. The guidelines indicated that, in adults with NAFLD, VCTE has the better diagnostic performance for diagnosing cirrhosis than the aspartate aminotransferase to platelet ratio index and Fibrosis-4 (very low quality of evidence). Moreover, the guidelines stated that, in adults with NAFLD, magnetic resonance–guided elastography has little or no increased diagnostic accuracy for identifying cirrhosis compared with VCTE in patients who have cirrhosis, and has higher diagnostic accuracy than VCTE in patients who do not have cirrhosis (very low quality of evidence).

Hepatitis B and C Viruses
National Institute for Health and Care Excellence
NICE (2013) published guidance on the management and treatment of patients with hepatitis B. The guidance recommended offering transient elastography as the initial test in adults diagnosed with chronic hepatitis B, to inform the antiviral treatment decision (see Table 8).

<table>
<thead>
<tr>
<th>Transient Elasticity Score</th>
<th>Antiviral Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;11 kPa</td>
<td>Offer antiviral treatment</td>
</tr>
<tr>
<td>6-10 kPa</td>
<td>Offer liver biopsy to confirm fibrosis level prior to offering antiviral treatment</td>
</tr>
<tr>
<td>&lt;6 kPa plus abnormal (ALT)</td>
<td>Offer liver biopsy to confirm fibrosis level prior to offering antiviral treatment</td>
</tr>
<tr>
<td>&lt;6 plus normal ALT</td>
<td>Do not offer antiviral treatment</td>
</tr>
</tbody>
</table>

ALT: alanine aminotransferase.

As of September 2016, NICE had placed a pause on the development of the guidance on hepatitis C, citing instability and costs in the availability of treatments for the condition.

American Association for the Study of Liver Diseases and Infectious Diseases Society of America
The American Association for the Study of Liver Diseases and Infectious Diseases Society of America (2018) guidelines for testing, managing, and treating hepatitis C virus (HCV) recommended that, for counseling and pretreatment assessment purposes, the following should be completed:

“Evaluation for advanced fibrosis using liver biopsy, imaging, and/or noninvasive markers is recommended in all persons with HCV infection to facilitate an appropriate decision regarding HCV treatment strategy and determine the need for initiating additional measures for the management of cirrhosis (eg, hepatocellular carcinoma screening).
Rating: Class I, Level A [evidence and/or general agreement; data derived from multiple randomized trials, or meta-analyses]”
The guidelines noted that there are several noninvasive tests to stage the degree of fibrosis in patients with hepatitis C. Tests included indirect serum biomarkers, direct serum biomarkers, and vibration-controlled liver elastography. The guidelines asserted that no single method is recognized to have high accuracy alone and careful interpretation of these tests is required.

American College of Gastroenterology Institute
Guidelines published by the American College of Gastroenterology Institute (2017) on the role of elastography in chronic liver disease indicated that, in adults with chronic hepatitis B virus and HCV, VCTE has better diagnostic performance for diagnosing cirrhosis than the aminotransferase to platelet ratio index and Fibrosis-4 (moderate quality of evidence for HCV, low quality of evidence for hepatitis B virus). In addition, the guidelines stated that, in adults with HCV, magnetic resonance–guided elastography has little or no increased diagnostic accuracy for identifying cirrhosis compared with VCTE in patients who have cirrhosis, and has lower diagnostic accuracy than VCTE in patients who do not have cirrhosis (very low quality of evidence).

Chronic Liver Disease
American College of Radiology
The American College of Radiology (2017) appropriateness criteria rated 1-dimensional transient elastography as a 7 (usually appropriate) for the diagnosis of liver fibrosis in patients with chronic liver disease. The criteria noted, “This procedure is less reliable in diagnosing liver fibrosis and cirrhosis in patients with obesity or ascites.”

European Association for the Study of Liver Disease et al
The European Association for the Study of Liver Disease and the Asociacion Latinoamericana para el Estudio del Higado (2015) convened a panel of experts to develop clinical practice guidelines on the use of noninvasive tests to evaluate liver disease severity and prognosis. The publication summarized the advantages and disadvantages of noninvasive techniques (serum biomarkers, imaging techniques). Table 9 summarized the joint recommendations for serum biomarkers and transient elastography.

| Table 9. Recommendations for Serum Biomarkers and Transient Elastography |
|-----------------------|---------|----------|
| **Biomarkers** | QOE | SOR |
| “Serum biomarkers can be used in clinical practice due to high applicability (>95%) and good reproducibility.” | High | Strong |
| “TE can be considered the non-invasive standard for the measure of LS” | High | Strong |
| “Serum biomarkers are well-validated for chronic viral hepatitis... They are less well-validated for NAFLD not validated in other chronic kidney diseases.” | High | Strong |
| “For the diagnosis of significant fibrosis a combination of tests with concordance may provide the highest diagnostic accuracy” | High | Weak |
| “All HCV patients should be screened to exclude cirrhosis by TE [or]... serum biomarkers....” | High | Strong |
| “Non-invasive assessment including serum biomarkers or TE can be used as first line procedure for the identification of patients at low risk of severe fibrosis/cirrhosis” | High | Strong |
| “Follow-up assessment by either serum biomarkers or TE for progression of liver fibrosis should be used for NAFLD patients at a 3 year interval” | Moderate | Strong |

HCV: hepatitis C virus; LS: liver stiffness; NAFLD: nonalcoholic fatty liver disease; QOE: quality of evidence; SOR: strength of recommendation; TE: transient elastography.

U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS
Not applicable.
ONGOING AND UNPUBLISHED CLINICAL TRIALS

Some currently unpublished trials that might influence this review are listed in Table 10.

Table 10. Summary of Key Trials

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT01789008</td>
<td>Transient Elastography in the Determination of Advanced Fibrosis in Alcoholic Liver Disease (FIBR-OH)</td>
<td>300</td>
<td>Aug 2017 (completed)</td>
</tr>
<tr>
<td>NCT02569567</td>
<td>Comparison of Smart-Shear Wave Elastography and Transient Elastography (SMART)</td>
<td>105</td>
<td>Jun 2016 (unknown)</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.

CODING

The following codes for treatment and procedures applicable to this policy are included below for informational purposes. Inclusion or exclusion of a procedure, diagnosis or device code(s) does not constitute or imply member coverage or provider reimbursement. Please refer to the member’s contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

CPT/HCPCS

- 76391 Magnetic resonance (eg, vibration) elastography
- 76981 Ultrasound, elastography; parenchyma (eg, organ)
- 76982 Ultrasound, elastography; first target lesion
- 76983 Ultrasound, elastography; each additional target lesion (List separately in addition to code for primary procedure)
- 81596 Infectious disease, chronic hepatitis C virus (HCV) infection, six biochemical assays (ALT, A2-macroglobulin, apolipoprotein A-1, total bilirubin, GGT, and haptoglobin) utilizing serum, prognostic algorithm reported as scores for fibrosis and necroinflammatory activity in liver
- 83520 Immunoassay for analyte other than infectious agent antibody or infectious agent antigen; quantitative, not otherwise specified
- 83883 Nephelometry, each analyte not elsewhere specified
- 84999 Unlisted chemistry procedure
- 91200 Liver elastography, mechanically induced shear wave (eg, vibration), without imaging, with interpretation and report
- 0002M Liver disease, ten biochemical assays (ALT, A2-macroglobulin, apolipoprotein A-1, total bilirubin, GGT, haptoglobin, AST, glucose, total cholesterol and triglycerides) utilizing serum, prognostic algorithm reported as quantitative scores for fibrosis, steatosis and alcoholicsteatohepatitis (ASH)
- 0003M Liver disease, ten biochemical assays (ALT, A2-macroglobulin, apolipoprotein A-1, total bilirubin, GGT, haptoglobin, AST, glucose, total cholesterol and triglycerides) utilizing serum, prognostic algorithm reported as quantitative scores for fibrosis, steatosis and nonalcoholic steatohepatitis (NASH)

Diagnoses

Experimental / Investigational for all diagnoses related to this medical policy.
## REVISIONS

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-16-2016</td>
<td>Policy added to the bcbks.com website on 08-17-2016.</td>
</tr>
<tr>
<td>01-18-2017</td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
<tr>
<td>12-20-2017</td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>• Added CPT code 0346T.</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
<tr>
<td>01-01-2019</td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>• Added new CPT codes: 76391, 76981, 76982, 76983, 81596.</td>
</tr>
<tr>
<td></td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>In Policy section:</td>
</tr>
<tr>
<td></td>
<td>• Added new Item A, “A single FibroSURE multianalyte assay may be considered medically necessary for the initial evaluation of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>• In new Item B (previous Item A), removed “with algorithmic analyses” and “the evaluation or” and added “FibroSURE” to read, “FibroSURE Multianalyte assays are considered experimental / investigational for monitoring of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>• Added new Item C, “Other multianalyte assays with algorithmic analyses are considered experimental / investigational for the initial evaluation or monitoring of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>• In new Item D (previous Item B), removed “ARFI” and added “initial” to read, “The use of noninvasive imaging, including, but not limited to, transient elastography (eg, FibroScan), magnetic resonance elastography, acoustic radiation force impulse imaging (eg, Acuson S2000), or real-time tissue elastography, is considered experimental / investigational for the initial evaluation or monitoring of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>• Removed CPT code: 0346T (deleted January 1, 2019).</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
<tr>
<td>03-13-2019</td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>In Policy section:</td>
</tr>
<tr>
<td></td>
<td>• Added new Item D, “Transient elastography (FibroScan) imaging may be considered medically necessary for the initial evaluation of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>• Added new Item E, “Transient elastography (FibroScan) imaging is considered experimental / investigational for monitoring of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>• In Item F (previously Item D), removed “transient elastography (eg, FibroScan)” to read, “The use of other noninvasive imaging, including, but not limited to, magnetic resonance elastography, acoustic radiation force impulse imaging (eg, Acuson S2000), or real-time tissue elastography, is considered experimental / investigational for the initial evaluation or monitoring of patients with chronic liver disease.”</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>• Removed CPT code: 0001M.</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
</tbody>
</table>
REFERENCES


Other References
1. Blue Cross and Blue Shield of Kansas Radiology Liaison Committee Consent Ballot, July 2018.
2. Blue Cross and Blue Shield of Kansas Radiology Liaison Committee, April 2016; July 2017.
3. Blue Cross and Blue Shield of Kansas Internal Medicine Liaison Committee, August 2018; January 2019.