Healthcare Improvement Scotland | SHTG Advice on health technologies

Evidence Note
Number 82 July 2018

In response to an enquiry from the Scottish Clinical Biochemistry Managed Diagnostic Network

What is the most accurate and cost-effective direct test (ELF test™, hyaluronic acid, P3NP, Fibroscan® or ARFI elastography) for detecting and staging liver fibrosis and cirrhosis in patients with diagnosed or suspected non-alcoholic fatty liver disease, alcohol-related liver disease, or viral hepatitis?

What is an evidence note?

Evidence notes are rapid reviews of the evidence on health technologies under consideration by decision makers within NHSScotland. They are intended to provide information quickly to support time-sensitive decisions. Information is available to the topic referrer within a 6-month period and final publication of the associated advice is usually complete within 6–12 months. Evidence notes are not comprehensive systematic reviews. They are based on the best evidence that Healthcare Improvement Scotland could identify and retrieve within the time available. Evidence notes do not make recommendations for NHSScotland, however the Scottish Health Technologies Group (SHTG) produces an Advice Statement to accompany all evidence reviews.

This evidence note is based upon a review of the published secondary clinical and cost effectiveness literature, and has been peer reviewed by experts across NHS Scotland.

Key points

- The ELF test™, hyaluronic acid and P3NP are direct serum biomarker tests, and Fibroscan® and ARFI elastography are direct imaging tests, for the detection and staging of liver fibrosis and cirrhosis.
- All included studies evaluated test performance in a secondary or tertiary care setting.

- All five direct tests of interest had sensitivity ≥70% for detection of fibrosis and cirrhosis in patients with hepatitis C or non-alcoholic fatty liver disease (NAFLD).
- In a meta-analysis of 13 studies (n=1,163) the direct imaging tests Fibroscan® and ARFI elastography had similar sensitivity and specificity for the detection of significant fibrosis and cirrhosis.
- In a meta-analysis of four studies (n=612) Fibroscan® imaging was slightly more sensitive and more specific for diagnosing cirrhosis compared with the ELF test™.
- In a prospective diagnostic cohort study Fibroscan® had lower sensitivity and higher specificity for detection of significant fibrosis, and similar sensitivity with higher specificity for the detection of advanced fibrosis, compared with the ELF test™. A second prospective diagnostic study reported that Fibroscan® had statistically significantly higher sensitivity and specificity for the detection of cirrhosis compared with serum hyaluronic acid.
- The ELF test™ had higher sensitivity and lower specificity than the indirect APRI and FIB4 scores for detection of significant fibrosis. Fibroscan® had higher sensitivity than the indirect APRI and FIB4 scores for detection of fibrosis and cirrhosis.
- No studies were identified that compared serum biomarkers with ARFI elastography, P3NP with any other non-invasive liver fibrosis test, or serum biomarkers with each other.
- In HBeAg-positive patients, testing with hyaluronic acid resulted in 11.66 QALYs gained per patient at a total cost of £79,084. Compared with not testing or treating anyone, which resulted in 9.64 QALYs at a cost of £37,831, hyaluronic acid was associated with an ICER of £20,422 per QALY.
- In HBeAg-negative patients the test/treat no one strategy offered 8.83 QALYs for a cost of £37,579, while the treat everyone without testing strategy offered 10.92 QALYs for a cost of £96,525. This gives an ICER of £28,204 per QALY for treating everyone compared to treating no one.
- Treating all hepatitis C patients without testing was cost-effective with an ICER of £8,573 per QALY compared with ARFI elastography.
- In patients with alcohol-related liver disease testing with Fibroscan® provided 9.02 QALYs at an overall cost of £20,009, whereas treating everyone without testing offered 9.50 QALYs for a cost of £31,004. This gives an ICER of £22,906 per QALY for the treat everyone without testing strategy compared with Fibroscan®.
- The NAFLD cost-effectiveness analysis was limited to incremental cost per correct diagnosis. ARFI elastography provided the highest rate of true positives at a cost of £8,600 per additional case detected compared with the next best alternative (hyaluronic acid at £600/case). The indirect NAFLD fibrosis score offered the highest rate of true negatives detected at a cost of £183 per additional case compared with the next best alternative (the FIB4 score at £80/case).

- In an analysis comparing the cost-effectiveness of tests for detecting cirrhosis irrespective of aetiology, the ELF test™, which had an ICER of £10,333 per QALY, seems to be the most cost-effective choice given commonly accepted willingness-to-pay thresholds.

1. Definitions

Definitions of terms relating to the accuracy of diagnostic tests are provided in appendix 1.

Biomarker: a biological molecule present in the blood, body fluids or tissues that indicates a normal or abnormal process, or the presence/absence of a condition or disease[1].

Liver fibrosis: scarring of the liver as a result of the accumulation of excess connective tissue caused by repeated liver injury from metabolic dysfunction, alcohol abuse, viral hepatitis or autoimmune disease[2].

Cirrhosis: end-stage liver disease characterised by nodules of liver tissue surrounded by advanced fibrosis and consequent liver dysfunction[3].

Extendedly dominated (a term used in health economics): an extendedly dominated strategy has an incremental cost-effectiveness ratio (ICER) higher than that of the next most effective strategy; therefore an extendedly dominated strategy produces additional gains in effectiveness at incremental costs higher than those of the next most effective strategy[4]. Funding such as strategy is not an efficient use of healthcare resources since the same gain could potentially be achieved at a lower cost through a combination of the previous and the next most effective strategies.

A complete list of abbreviations used in the evidence note is provided in appendix 2.

2. Literature search

A systematic search of the secondary literature was carried out between 16 and 19 November 2017 to identify systematic reviews, health technology assessments and other evidence based reports. Medline, Medline in process, Embase, Cinahl and Web of Science databases were also searched for systematic reviews and meta-analyses.

Key websites were searched for guidelines, policy documents, clinical summaries and economic studies.

The primary literature was systematically searched between 12 and 13 December 2017 using the following databases: Medline (All), Embase and Web of Science. Results were limited to English language and publication date 2007–present.

Concepts used in all searches included: liver/hepatic fibrosis, enhanced liver fibrosis (ELF), serum hyaluronic acid (HA), Fibroscan®, transient elastography and biomarkers. A full list of resources searched and terms used are available on request.

Studies identified in the literature search were excluded if they:

- focused on sequential testing with multiple tests or explored the effect of combining liver fibrosis tests,

- did not report sensitivity and specificity for all tests in a comparison,
- included only East Asian patients (viral hepatitis variants, prevalence of chronic liver diseases, and performance of serum biomarker tests all differ between Asian and white patient populations)[5], or
- only reported the use of liver fibrosis tests for predicting patient prognosis.

3. Introduction

Patients with chronic liver disease are at risk of developing liver fibrosis and cirrhosis[6]. The most common chronic liver diseases associated with liver fibrosis and cirrhosis are non-alcoholic fatty liver disease (NAFLD), alcohol-related liver disease and viral hepatitis. Much of the morbidity and mortality associated with chronic liver disease occurs in those patients who also have advanced liver fibrosis or cirrhosis[7].

Liver fibrosis is a progressive condition where excess connective tissue builds up in the liver causing scarring and loss of liver function[6]. This accumulation of excess connective tissue and scarring is caused by repeated liver injury from metabolic dysfunction, alcohol misuse, viral hepatitis infection or autoimmune disease, followed by hepatic regeneration[2]. Cirrhosis develops as a final irreversible stage in the progression of liver fibrosis[8].

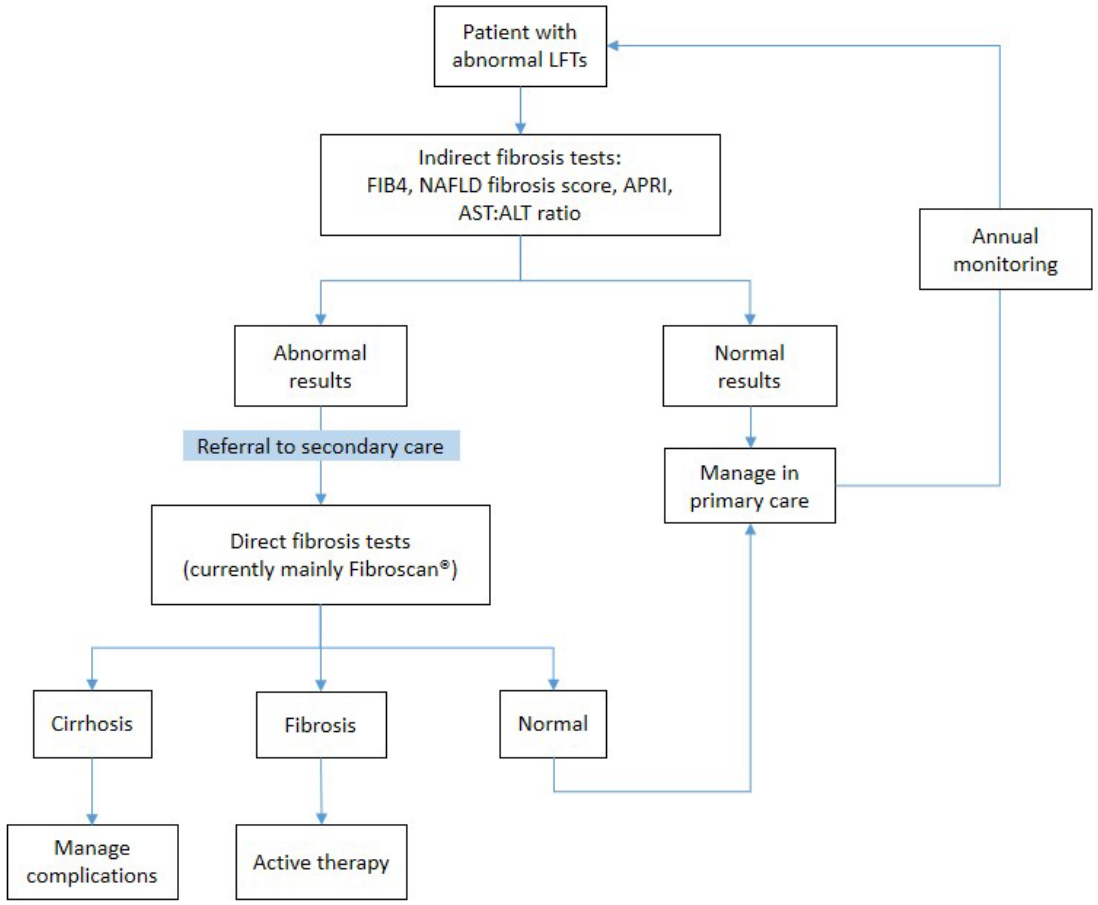
There are several systems for defining stages of liver fibrosis and cirrhosis. The most commonly reported classification system is the METAVIR score which defines liver fibrosis stages based on histopathology from liver biopsies (table 1). Early detection and staging of fibrosis helps inform treatment decisions and monitoring of disease progression[9]. In some cases, such as antiviral therapy for patients with hepatitis C, early intervention can halt and even reverse liver fibrosis[8].

Table 1: METAVIR score for defining stages of liver fibrosis and cirrhosis

METAVIR score	Fibrosis stage
F0-1	No or mild fibrosis
F≥2	Significant fibrosis
F≥3	Severe or advanced fibrosis

Figure 1 shows a typical clinical pathway for diagnosis, monitoring and treatment of patients with suspected liver fibrosis or cirrhosis in Scotland (Dr A MacGilchrist, Consultant Hepatologist, NHS Lothian. Personal communication, 11 December 2017). Patients initially presenting to primary care with liver disease or non-specific symptoms that suggest impaired liver function are offered blood tests (Dr S Jenks, Clinical Biochemistry Registrar, NHS Lothian. Personal communication, 29 November 2017). These liver function tests (LFTs) consist of a panel of blood tests that together give an indication of liver function. However, these tests are poorly correlated with liver disease or dysfunction[10]. Consequently many patients without liver disease undergo further investigations while some patients with liver disease remain undiagnosed until presenting at a hospital with potentially life-threatening complications of liver cirrhosis.

Figure 1: typical clinical pathway for patients with abnormal liver functions tests (LFTs) in Scotland



Liver biopsy is the current reference standard for diagnosing liver fibrosis and cirrhosis. This is an invasive procedure and an imperfect reference standard because it has:

- a risk of severe complications, such as major bleeding or pain following the procedure, resulting in approximately 1–5% of patients being hospitalised[2, 11, 12],
- low patient acceptability, particularly as a test for repeat monitoring of fibrosis progression[11],
- intra- and inter-observer discrepancy rates of approximately 10–20%[12], and
- been associated with incorrect staging in up to 30% of cases where an adequate sample was obtained[13].

For the above reasons, liver biopsy is not used to diagnose or stage liver fibrosis in routine clinical practice in Scotland (Dr A MacGilchrist, Consultant Hepatologist, NHS Lothian. Personal communication, 11 December 2017).

As a consequence of the invasive nature of liver biopsy and the growing burden of chronic liver disease a number of non-invasive liver fibrosis tests have been developed (table 2)[2, 11]. These tests can be categorised as indirect or direct tests. Indirect fibrosis tests combine routine LFT results with patient characteristics, such as age and weight, in an algorithm to determine the likelihood of fibrosis and cirrhosis. Direct fibrosis tests measure biological changes in the blood (serum biomarkers) or liver tissue stiffness (imaging tests) which occur as a result of the fibrosis process, thereby providing a more direct measure of fibrosis and cirrhosis.

Table 2: categories of liver fibrosis tests

Category	Tests*
Indirect tests	Fibrosis-4 (FIB4) score, aspartate transaminase: alanine transaminase (AST:ALT) ratio, AST:platelet ratio index (APRI) score, NAFLD fibrosis score
Serum biomarkers (direct tests)	Enhanced liver fibrosis (ELF) test™, hyaluronic acid, N-terminal propeptide type III procollagen (P3NP or PIIINP)
Imaging tests (direct tests)	Transient elastography (Fibroscan®), acoustic radiation force impulse (ARFI) elastography
Reference standard	Liver biopsy

**Not all non-invasive tests are listed in this table – only tests currently relevant to the Scottish healthcare system are included*

This evidence note aims to address the question:

- What is the most accurate and cost-effective direct test (ELF test™, hyaluronic acid, P3NP, Fibroscan® or ARFI elastography) for detecting and staging liver fibrosis and cirrhosis in patients with diagnosed or suspected non-alcoholic fatty liver disease, alcohol-related liver disease or viral hepatitis?

Interest in direct serum biomarker tests (ELF test™, hyaluronic acid, P3NP) lies in the value they could add when used for further investigations in secondary care following referral based on indirect tests, or as replacements for indirect tests in the care pathway (figure 1) if the diagnostic accuracy and cost-effectiveness of these biomarker tests proves greater than that of indirect tests. Due to the need for specialist equipment, direct imaging tests (Fibroscan® and ARFI elastography) are currently unlikely to be performed in primary care. Therefore direct imaging tests could be used in secondary care as further investigations following referral from primary care. Direct fibrosis tests (biomarkers or imaging) could also add value as tools for monitoring patients.

4. Health technology description

The ELF test™ is a commercially produced panel of three serum markers of liver metabolism: hyaluronic acid, tissue inhibitor of metalloproteinases-1 (TIMP1) and P3NP[7]. The test manufacturer (Siemens Healthcare Ltd) recommends using an ELF score ≥7.7 and <9.8 for diagnosing moderate fibrosis, and ≥9.8 for diagnosing advanced fibrosis; higher ELF scores indicate more severe fibrosis. The ELF test™ is not currently used for routine liver fibrosis testing in Scotland.

Hyaluronic acid and P3NP are serum biomarkers that are elevated in patients with liver fibrosis or cirrhosis[12]. The hyaluronic acid level considered 'normal' is ≤75ng/ml (Prof P Hayes, Consultant Hepatologist, University of Edinburgh. Personal communication, 25 January 2018). The reference range for P3NP is approximately 1.3–4.2mcg/l[14]. Hyaluronic acid is currently used in NHS Lothian, and P3NP used in NHS Tayside, for liver fibrosis testing.

Transient elastography (Fibroscan®) and ARFI elastography are ultrasound-based imaging techniques that measure liver stiffness[15]. Fibroscan® measures liver stiffness in kilopascals (kPa) and ARFI elastography measures liver stiffness in metres per second (m/s). Fifteen out of 24 hospitals in Scotland had access to

Fibroscan® devices in 2013[16]. The availability of ARFI elastography in Scotland is unclear but it may be an option on many existing ultrasound machines.

5. Epidemiology

The prevalence of liver fibrosis and cirrhosis in the chronic liver disease patient population is uncertain. Estimated incidence of chronic liver diseases in Scotland and the proportion of those patients thought to develop fibrosis and cirrhosis are presented in table 3. It is likely that the figures in table 3 are underestimates, as many people with liver disease, fibrosis or cirrhosis are undiagnosed[10].

Table 3: estimated incidence of chronic liver disease in Scotland and the proportion of liver disease patients who may develop fibrosis and cirrhosis

Condition	Estimated disease incidence in Scotland	Estimated % developing fibrosis (any stage)	Estimated % developing cirrhosis
NAFLD	2,211 new patients (2016-17)*	-	9% to 20%[17]
Hepatitis C	1,594 new cases (2016)[18]	Up to 50%[19]	Up to 20% within 20 years of infection[20]
Hepatitis B	418 new cases (2015)[21]	Up to 36%[22]	Up to 18%[22]
Alcohol-related liver disease	1,088 new hospital admissions (2016-17)[23]	Up to 40%[24]	10% to 20%[24]

**R Gordon, Senior Information Analyst, ISD Scotland. Personal communication, 3 March 2018.*

The progression of fibrosis can be very gradual, occurring over many years. For example, studies have suggested little fibrosis progression occurs in the first decade of hepatitis C infection, with cirrhosis developing after 20–30 years of infection[25].

Mortality from chronic liver disease (not including NAFLD) has been relatively stable in Scotland over the past three years at approximately 15 deaths per 100,000 population per year[26].

6. UK guidelines

Four UK guidelines were identified that make recommendations on tests for liver fibrosis and cirrhosis in patients with chronic liver disease[10, 20, 27, 28]. The relevant recommendations are summarised in table 4.

Table 4: UK guideline recommendations for liver fibrosis testing

Guideline	Liver disease	Summary of recommendations
British Society of Gastroenterology (2017)[10]	NAFLD	Following abnormal LFT results where NAFLD is suggested by ultrasound and/or a negative liver screen: ■ First-line testing should use the FIB4 or NAFLD fibrosis score in primary care. Patients with low FIB4 (<1.30) or NAFLD fibrosis score (<-1.455) may be managed in primary care. Patients with elevated FIB4 (>3.25) or

		NAFLD fibrosis score (>0.675) should be referred to a specialist clinic. ■ For patients with a FIB4 score of 1.30–3.25 or an NAFLD fibrosis score of -1.455–0.675, second-line tests that could be considered are the ELF test™, Fibroscan® or ARFI elastography. Patients with an ELF score >9.5 or Fibroscan® reading >7.8kPa should be referred to a specialist clinic.
	Alcohol-related liver disease	Following abnormal LFT results patients who are harmful drinkers should have a clinical assessment and either Fibroscan® or ARFI elastography. Patients with evidence of cirrhosis, or hypertension and/or a Fibroscan® reading >16kPa, should be referred to a specialist clinic.
	Hepatitis B or hepatitis C	Following abnormal LFT results refer all hepatitis B and C patients to specialist clinics in accordance with local processes.
NICE (2016)[27]	NAFLD	Consider using the ELF test™ to detect advanced liver fibrosis. Refer patients with NAFLD and advanced fibrosis (ELF score ≥10.51) to a hepatology specialist.
NICE (2016)[20]	Cirrhosis	Offer transient elastography to diagnose cirrhosis in people consuming >50 units of alcohol/week (male) or >35 units of alcohol/week (female); people with hepatitis C; and people diagnosed with alcohol-related liver disease. Offer either transient elastography or ARFI elastography (whichever is available) to diagnose cirrhosis in people with NAFLD and advanced liver fibrosis (ELF score ≥10.51). Consider liver biopsy in people for whom transient elastography is not suitable. Do not offer tests for cirrhosis to people who are obese or have type 2 diabetes, unless they have NAFLD and advanced liver fibrosis.
NICE (2017)[28]	Hepatitis B	Offer transient elastography to patients newly referred for assessment. Consider liver biopsy to confirm fibrosis stage in patients who have transient elastography readings between 6kPa and 10kPa or patients who have a transient elastography reading less than 6kPA, are under 30 years old, have abnormal ALT levels on two consecutive tests and have high hepatitis B viral DNA levels.

7. Clinical effectiveness

A health technology assessment (HTA) was identified that evaluated the clinical and cost effectiveness of numerous non-invasive liver fibrosis tests, including those of interest to this evidence note[29]. This HTA informed the cost effectiveness section of the evidence note and the diagnostic accuracy systematic review contained within the HTA is described in section 7.1. As this HTA did not compare fibrosis tests with each other, and test accuracy could not be directly compared between studies due to different patient populations, an additional search was carried out to identify studies that compared direct fibrosis tests with each other or with indirect fibrosis tests using liver biopsy as the reference standard. A summary of the number of studies reporting each test comparison is provided in table 5. In total, one HTA, two meta-analyses and six prospective diagnostic cohort studies (additional to the secondary literature) provided evidence for the clinical effectiveness section of this review.

Table 5: summary of number of studies for each liver fibrosis test comparison

	Direct tests				Indirect tests			
	HA	**P3NP**	**Fibroscan®**	**ARFI**	**FIB4**	**APRI**	**AST:ALT**	**NAFLD**
ELF test™	-	-	5	-	1	1	-	-
HA		-	1	-	-	-	-	-
P3NP			-	-	-	-	-	-
Fibroscan®				13	1	2	1	2
ARFI					2	2	1	-

HA = hyaluronic acid. Some cells include studies from within a systematic review, for example ELF test™ versus Fibroscan®.

Studies included in this evidence note may not match the sequence of testing in the typical pathway in Scotland. Some studies reported multiple comparisons of interest and therefore the same study may be discussed in more than one section of the evidence note. The majority of studies used METAVIR scoring to define fibrosis stages (table 1); where studies used an alternative classification system this is reported in the description of the individual study. The thresholds used to diagnose stages of fibrosis and cirrhosis varied between studies for some tests.

For any diagnostic test there are four possible outcomes: true positive, false positive, true negative and false negative. Diagnostic tests can be evaluated based on their ability to minimise false positives or false negatives depending on which outcome carries the higher risk for patients. In the case of liver fibrosis, greater risk would appear to be associated with false negative results as patients with liver fibrosis who are wrongly given a negative test result would not receive the appropriate treatment and could potentially progress to irreversible cirrhosis before being correctly diagnosed. Therefore tests for liver fibrosis with high sensitivity are desirable.

Many fibrosis tests were originally validated in patients with a specific liver disease, yet were applied in other patient populations in the included studies[30]. In clinical practice patients often present without a specific diagnosis or with comorbid liver diseases and therefore tests are applied in a real-world patient population with liver disease of different aetiologies who are at risk of liver fibrosis (Dr A MacGilchrist, Consultant Hepatologist, NHS Lothian. Personal communication, 11 December 2017). Four included studies reported performance of fibrosis tests in mixed patient populations and may therefore be generalisable to the Scottish clinical setting. Where studies recruited patients with specific liver diseases the diagnostic accuracy of the tests evaluated was assumed – on the basis of consultation with a clinical advisor – to be

generalisable to other liver disease patient populations unless explicitly stated otherwise in the literature (Dr A MacGilchrist, Consultant Hepatologist, NHS Lothian. Personal communication, 11 December 2017). No studies were identified that compared relevant tests in patients with alcohol-related liver disease.

7.1. Diagnostic accuracy of direct fibrosis tests using liver biopsy as a reference standard

A systematic review and meta-analysis conducted as part of a large HTA evaluated non-invasive tests (direct and indirect) for the assessment of liver fibrosis in patients with chronic liver disease[29]. The HTA determined the diagnostic accuracy of each test using liver biopsy as the reference standard. Where included studies used a fibrosis staging system other than METAVIR, the HTA authors converted the fibrosis scores to METAVIR for the meta-analysis. Meta-analyses were conducted separately for each liver disease aetiology: hepatitis C, hepatitis B, alcohol-related liver disease, and NAFLD. An additional analysis assessed non-invasive tests for detecting cirrhosis irrespective of liver disease aetiology.

The HTA included 302 studies; not all of which included tests of interest for this evidence note. Very few studies (five out of 302) were considered to be high quality based on QUADAS-2 appraisal criteria. Few studies reported the cut-off value(s) used or obtained acceptable quality liver biopsy samples for the reference standard. Due to the low quality of included studies and uncertainty about the thresholds used in many studies, the HTA authors recommend the results of the meta-analyses be treated with caution as the diagnostic accuracy of the tests evaluated may have been overestimated.

Hepatitis C

All five direct fibrosis tests of interest were evaluated in one or more studies in patients with hepatitis C (table 6). There was variability between studies in the thresholds (cut-offs) used, particularly in studies assessing hyaluronic acid or Fibroscan®. The five direct fibrosis tests had good sensitivity (≥70%) for detecting significant fibrosis, advanced fibrosis and cirrhosis in this population. Specificity across the tests was also ≥70%, except for the ELF test™ and the biomarker P3NP for diagnosing advanced fibrosis.

Table 6: diagnostic accuracy of direct tests for liver fibrosis in patients with chronic hepatitis C

	Biomarkers			Imaging tests	
	ELF test™	Hyaluronic acid	P3NP	Fibroscan®	ARFI elastography
Significant fibrosis (F≥2)					
No. of studies	1	8	2	37	3
Cut-off	8.75	34–110ng/ml	8.3–9.1	5.2–10.1kPa	1.21–1.34m/s
Summary sensitivity (95% confidence interval (CI))	0.84 (0.69 to 0.92)	0.75 (0.64 to 0.83)	0.78 (0.63 to 0.87)	0.79 (0.74 to 0.84)	0.79 (0.75 to 0.83)
Summary specificity (95% CI)	0.70 (0.52 to 0.83)	0.75 (0.68 to 0.82)	0.76 (0.54 to 0.90)	0.83 (0.77 to 0.88)	0.89 (0.84 to 0.93)
Advanced fibrosis (F≥3)					
No. of studies	1	4	2	19	4
Cut-off	9.59	20–85ng/ml	8.0–9.1	8.6–15.4kPa	1.49–2.11m/s
Summary sensitivity (95% CI)	0.85 (0.77 to 0.90)	0.79 (0.52 to 0.93)	0.71 (0.58 to 0.81)	0.88 (0.82 to 0.92)	0.85 (0.69 to 0.94)
Summary specificity (95% CI)	0.63 (0.57 to 0.70)	0.72 (0.65 to 0.78)	0.63 (0.54 to 0.71)	0.90 (0.85 to 0.93)	0.89 (0.72 to 0.97)
Cirrhosis (F=4)					
No. of studies	1	7	3	36	4
Cut-off	9.4	78–237ng/ml	0.8–1*	9.2–17.3kPa	1.6–2.3m/s
Summary sensitivity (95% CI)	0.93 (0.69 to 0.99)	0.80 (0.61 to 0.91)	0.70 (0.42 to 0.89)	0.89 (0.84 to 0.92)	0.84 (0.72 to 0.91)
Summary specificity (95% CI)	0.79 (0.67 to 0.88)	0.88 (0.78 to 0.94)	0.84 (0.74 to 0.90)	0.91 (0.89 to 0.93)	0.77 (0.50 to 0.92)

**An error in the cut-off values for P3NP for cirrhosis appears to have been reported in the original study[29]*

Hepatitis B

Studies were only identified for three direct fibrosis tests of interest in patients with hepatitis B: hyaluronic acid, Fibroscan® and ARFI elastography. Data on hyaluronic acid were only available for detecting significant fibrosis and cirrhosis; for ARFI elastography data were only reported for significant fibrosis (table 7).

Table 7: diagnostic accuracy of direct tests for liver fibrosis in patients with chronic hepatitis B

	Biomarker	Imaging tests	
	Hyaluronic acid	Fibroscan®	ARFI elastography
Significant fibrosis (F≥2)			
No. of studies	1	13	1
Cut-off	185.3ng/ml	6.3–8.9kPa	1.33m/s
Summary sensitivity (95% CI)	0.84 (0.73 to 0.91)	0.71 (0.62 to 0.78)	0.71 (0.59 to 0.80)
Summary specificity (95% CI)	0.83 (0.66 to 0.93)	0.84 (0.74 to 0.91)	0.67 (0.30 to 0.90)
Advanced fibrosis (F≥3)			
No. of studies	0	13	0
Cut-off	-	7.3–10.7kPa	-
Summary sensitivity (95% CI)	-	0.69 (0.58 to 0.78)	-
Summary specificity (95% CI)	-	0.84 (0.79 to 0.89)	-
Cirrhosis (F≥4)			
No. of studies	1	19	0
Cut-off	77ng/ml	9.4–16.0kPa	-
Summary sensitivity (95% CI)	0.82 (0.52 to 0.95)	0.86 (0.79 to 0.91)	-
Summary specificity (95% CI)	0.88 (0.79 to 0.93)	0.85 (0.78 to 0.89)	-

Alcohol-related liver disease

Diagnostic accuracy studies were only available for Fibroscan® imaging in patients with alcohol-related liver disease. In one study, using a cut-off of 7.8kPa, sensitivity was 0.81 (95% CI 0.70 to 0.88) and specificity was 0.92 (95% CI 0.76 to 0.98) for detecting significant fibrosis. In a meta-analysis of four studies summary sensitivity was 0.87 (95% CI 0.64 to 0.96) and specificity was 0.82 (95% CI 0.67 to 0.91) for detecting advanced fibrosis. In a meta-analysis of six studies sensitivity was 0.86 (95% CI 0.76 to 0.92) and specificity was 0.83 (95% CI 0.74 to 0.89) for the detection of cirrhosis.

NAFLD

Diagnostic accuracy of liver fibrosis tests was reported for a sub-set of the NAFLD population who had non-alcoholic steatohepatitis (NASH). No studies evaluated the P3NP biomarker in patients with NAFLD and studies assessing ARFI elastography were only available for the detection of advanced fibrosis in this patient group. For detection of cirrhosis, results were only available for Fibroscan® imaging. Both sensitivity and specificity were ≥70% for the detection of significant and advanced fibrosis using the ELF test™, serum hyaluronic acid or Fibroscan® imaging (table 8).

Table 8: diagnostic accuracy of direct tests for detection of liver fibrosis in patients with NASH

	Biomarkers		Imaging tests	
	ELF test™	Hyaluronic acid	Fibroscan®	ARFI elastography
Significant fibrosis (F≥2)				
No. of studies	1	1	7	0
Cut-off	9.9	218ng/ml	6.8–10.0kPa	-
Summary sensitivity (95% CI)	0.70 (0.59 to 0.79)	0.78 (0.45 to 0.94)	0.79 (0.72 to 0.85)	-
Summary specificity (95% CI)	0.80 (0.72 to 0.86)	0.89 (0.67 to 0.97)	0.76 (0.71 to 0.80)	-
Advanced fibrosis (F≥3)				
No. of studies	1	4	8	1
Cut-off	10.35	46–50ng/ml	7.5–10.4kPa	4.2m/s
Summary sensitivity (95% CI)	0.80 (0.65 to 0.89)	0.88 (0.58 to 0.97)	0.82 (0.74 to 0.88)	0.90 (0.77 to 0.96)
Summary specificity (95% CI)	0.90 (0.84 to 0.94)	0.82 (0.75 to 0.87)	0.84 (0.78 to 0.89)	0.90 (0.82 to 0.94)
Cirrhosis (F≥4)				
No. of studies	0	0	4	0
Cut-off	-	-	10.3–17.5kPa	-
Summary sensitivity (95% CI)	-	-	0.96 (0.83 to 0.99)	-
Summary specificity (95% CI)	-	-	0.89 (0.85 to 0.92)	-

Cirrhosis irrespective of aetiology

In this analysis the ability of tests to detect cirrhosis was evaluated regardless of liver disease aetiology. One or more studies reported diagnostic accuracy of all five direct tests of interest for detecting cirrhosis (table 9). Sensitivity for detecting cirrhosis was lowest for the serum biomarker P3NP and highest for the ELF test™.

Table 9: diagnostic accuracy of direct tests for detection of cirrhosis irrespective of liver disease aetiology

	Biomarkers			Imaging tests	
	ELF test™	Hyaluronic acid	P3NP	Fibroscan®	ARFI elastography
No. of studies	1	8	3	65	4
Cut-off	9.4	78–237ng/ml	0.8–1.0	9.2–26.5kPa	1.59–2.00m/s
Summary sensitivity (95% CI)	0.93 (0.69 to 0.99)	0.81 (0.65 to 0.90)	0.70 (0.48 to 0.86)	0.89 (0.86 to 0.91)	0.84 (0.72 to 0.91)
Summary specificity (95% CI)	0.79 (0.67 to 0.88)	0.88 (0.80 to 0.94)	0.79 (0.34 to 0.96)	0.89 (0.87 to 0.91)	0.77 (0.50 to 0.92)

Summary:

- Sensitivity and specificity of fibrosis tests could not be directly compared in the HTA as tests were evaluated in separate studies with different participants.
- Sensitivity was ≥70% for all five direct tests of interest for the detection of fibrosis and cirrhosis in patients with hepatitis C. Specificity was ≥70% for all five tests for the detection of significant fibrosis and cirrhosis.
- Sensitivity and specificity were both ≥70% for the ELF test™, hyaluronic acid and Fibroscan® imaging for detecting significant and advanced fibrosis in patients with NAFLD.

7.2. Direct serum biomarkers compared with indirect fibrosis tests

A prospective diagnostic cohort study compared the performance of the ELF test™ with the indirect APRI and FIB4 scores in 107 consecutively recruited patients with hepatitis C[31]. All participants underwent liver biopsy, 106 patients had APRI and FIB4 scores, and 68 participants received the ELF test™. It is unclear how much time elapsed between tests or why some patients did not receive all four tests. Liver biopsies were assessed by a single expert to reduce inter-observer bias. Very few study participants had cirrhosis on liver biopsy: F0 n=8, F1 n=43, F2 n=31, F3 n=23 and F4 n=2. The diagnostic accuracy of the ELF test™ compared with the indirect APRI and FIB4 scores is reported in table 10. The ELF test™ had high sensitivity (≥83%) for detecting significant fibrosis, advanced fibrosis and cirrhosis. The ELF test™ had higher sensitivity and lower specificity than the APRI and FIB4 scores for detecting significant fibrosis. High sensitivity and specificity (100%) for detecting cirrhosis was reported for all three tests. Due to the small number of study participants with cirrhosis (n=2) these results should be treated with caution.

Table 10: diagnostic accuracy of the ELF test™ compared with the indirect APRI and FIB4 scores in patients with hepatitis C[31]

	ELF test™ (direct biomarker)	APRI (indirect test)	FIB4 (indirect test)
Significant fibrosis (F≥2)			
Cut-off	8.98	0.67	1.29
Sensitivity (95% CI)	0.83 (0.64 to 0.94)	0.48 (0.35 to 0.62)	0.68 (0.54 to 0.80)
Specificity (95% CI)	0.56 (0.40 to 0.72)	0.86 (0.73 to 0.94)	0.76 (0.62 to 0.87)
Advanced fibrosis (F≥3)			
Cut-off	9.47	0.67	1.22
Sensitivity (95% CI)	0.83 (0.52 to 0.98)	0.60 (0.39 to 0.79)	0.84 (0.64 to 0.95)
Specificity (95% CI)	0.70 (0.56 to 0.81)	0.77 (0.66 to 0.85)	0.60 (0.49 to 0.71)
Cirrhosis (F=4)			
Cut-off	11.00	4.30	6.51
Sensitivity (95% CI)	1.00* (NR)	1.00 (0.16 to 1.00)	1.00 (0.16 to 1.00)
Specificity (95% CI)	1.00 (0.93 to 1.00)	1.00 (0.97 to 1.00)	1.00 (0.97 to 1.00)

**An error in the specificity estimate appears to have been reported in the original study. A confidence interval was not reported for the sensitivity estimate for this test[31].*

Summary:

- In a prospective diagnostic cohort study (n=107) the direct ELF test™ had high sensitivity (83%) for detecting significant and advanced fibrosis in patients with hepatitis C. The ELF test™ was more sensitive, but less specific, than the indirect APRI and FIB4 scores for detecting significant fibrosis in this patient population.

7.3. Direct serum biomarkers compared with Fibroscan® imaging

One meta-analysis and two additional prospective diagnostic cohort studies reported comparisons between direct serum biomarkers and Fibroscan® imaging in patients with chronic liver disease[13, 31, 32]. One of the additional primary studies was described in section 7.2[31].

The meta-analysis incorporated four studies comparing the accuracy of the ELF™ test with Fibroscan®[13]. The included studies recruited patients with viral hepatitis (three studies, n=510) or liver disease of varied aetiologies (one study, n=102). The included studies were rated as being of overall good quality using the QUADAS appraisal tool, although results were not reported by study. High heterogeneity in the analysis was attributed by the authors to threshold (cut-off) bias. Pooled diagnostic accuracy was only reported for the detection of cirrhosis which was present on biopsy for 20% of participants: F0 n=62, F1 n=151, F2 n=80, F3 n=108 and F4 n=126. Sensitivity and specificity were both lower for the ELF test™ compared with Fibroscan® for the detection of cirrhosis (table 11). The cut-off values, sensitivity and specificity for the ELF test™ were lower in studies included in the

meta-analysis than in the study by Ragazzo et al (2017)[31] reported in section 7.2. This may be due to the larger number of patients with cirrhosis in the meta-analysis (n=126).

Table 11: pooled diagnostic accuracy of the ELF test™ compared with Fibroscan® imaging for detecting cirrhosis in patients with chronic liver disease[13]

	ELF test™ (direct biomarker)	Fibroscan® (direct imaging test)
Cut-off*	9.50–10.31	10.0–17.6kPa
Pooled sensitivity (95% CI)	0.78 (0.70 to 0.85)	0.82 (0.75 to 0.88)
Pooled specificity (95% CI)	0.64 (0.59 to 0.69)	0.89 (0.86 to 0.92)

**Range of thresholds used in studies included in the meta-analysis*

The study by Ragazzo et al (2017) compared the ELF test™ with Fibroscan® in patients with hepatitis C (n=107)[31]. The ELF test™ had high sensitivity (83%) and Fibroscan® had high specificity (92%) for detecting significant fibrosis (table 12). Fibroscan® had good sensitivity and specificity for the detection of advanced fibrosis (80% and 79% respectively). The ELF test™ had similar sensitivity and lower specificity than Fibroscan® imaging for the detection of advanced fibrosis.

Table 12: diagnostic accuracy of the ELF test™ compared with Fibroscan® imaging in patients with hepatitis C[31]

	ELF test™ (direct biomarker)	Fibroscan® (direct imaging test)
Significant fibrosis (F≥2)		
Cut-off	8.98	6.5kPa
Sensitivity (95% CI)	0.83 (0.64 to 0.94)	0.71 (0.58 to 0.83)
Specificity (95% CI)	0.56 (0.40 to 0.72)	0.92 (0.81 to 0.98)
Advanced fibrosis (F≥3)		
Cut-off	9.47	7.1kPa
Sensitivity (95% CI)	0.83 (0.52 to 0.98)	0.80 (0.59 to 0.93)
Specificity (95% CI)	0.70 (0.56 to 0.81)	0.79 (0.69 to 0.87)
Cirrhosis (F=4)		
Cut-off	11.0	27kPa
Sensitivity (95% CI)	1.00* (NR)	1.00 (0.16 to 1.00)
Specificity (95% CI)	1.00 (0.93 to 1.00)	0.99 (0.95 to 1.00)

**An error in the specificity estimate appears to have been reported in the original study. A confidence interval was not reported for the sensitivity estimate for this test[31]*

The second prospective diagnostic cohort study compared the serum biomarker hyaluronic acid with Fibroscan® imaging in patients with liver disease of varied aetiologies (n=404)[32]. Participants with a

clinical diagnosis of cirrhosis did not undergo a liver biopsy, therefore not all study participants received the reference standard. This may have allowed patients with more severe disease to be included in the analysis. Clinicians interpreting tests were blinded to Fibroscan® imaging results. This study only evaluated the diagnostic accuracy of tests for detecting cirrhosis which was biopsy proven or clinically defined in 31% of participants: F0 n=104, F1 n=83, F2 n=49, F3 n=44 and F4 n=124. In this study sensitivity and specificity were statistically significantly lower ($p<0.05$) for hyaluronic acid compared with Fibroscan® imaging for the detection of cirrhosis (table 13). The threshold used to define cirrhosis on Fibroscan® imaging was much lower than that reported in the study by Ragazzo et al (2017)[31] (table 12).

Table 13: diagnostic accuracy of serum hyaluronic acid compared with Fibroscan® imaging for the detection of cirrhosis in patients with chronic liver disease[32]

	Hyaluronic acid (direct biomarker)	Fibroscan® (direct imaging test)
Cut-off	>150ng/ml	>12kPa
Sensitivity	0.72	0.92
Specificity	0.79	0.88

Summary:

- In a meta-analysis of four studies (n=612) the ELF test™ had lower sensitivity and specificity than Fibroscan® imaging for detecting cirrhosis.
- In a prospective diagnostic cohort study (n=107) the ELF test™ had higher sensitivity and lower specificity for detection of significant fibrosis, and similar sensitivity with lower specificity for the detection of advanced fibrosis, compared with Fibroscan® imaging.
- In a prospective diagnostic cohort study (n=404) the serum biomarker hyaluronic acid had statistically significantly lower sensitivity and specificity compared with Fibroscan® imaging for the detection of cirrhosis.

7.4. Direct serum biomarkers compared with ARFI elastography

No studies were identified that compared direct serum biomarkers with ARFI elastography for detection of fibrosis and cirrhosis.

7.5. Fibroscan® imaging compared with indirect fibrosis tests

Three diagnostic accuracy studies compared Fibroscan® imaging with indirect fibrosis tests in patients with chronic liver disease[32-34]. One primary study was previously described in section 7.3[32].

In the study by Malik et al (2010) described in section 7.3, Fibroscan® imaging was compared with the APRI score and AST:ALT ratio (indirect tests) for detecting cirrhosis in patients with chronic liver disease of varying aetiologies (n=404)[32]. Fibroscan® imaging had statistically significantly higher ($p<0.05$) sensitivity and specificity than the indirect tests for detecting cirrhosis (table 14).

Table 14: diagnostic accuracy of Fibroscan® imaging compared with the APRI score and AST:ALT ratio for detecting cirrhosis in patients with chronic liver disease[32]

	Fibroscan® (direct imaging test)	APRI (indirect test)	AST:ALT ratio (indirect test)
Cut-off	>12kPa	>1	>1
Sensitivity	0.92	0.73	0.64
Specificity	0.88	0.54	0.59

A second prospective diagnostic cohort study compared Fibroscan® with the NAFLD fibrosis score (n=88)[33]. This study used the NAFLD Activity Score to define fibrosis stages: significant fibrosis (F2-4), severe fibrosis (F3-4) and cirrhosis (F4). While these categories appear similar to the METAVIR system it is unclear from the study report if this is the case. A single operator conducted all Fibroscan® tests to reduce inter-observer bias. No information is provided on the time interval between tests or the cut-off values used, and only nine participants had cirrhosis: F0 n=23, F1 n=21, F2 n=17, F3 n=18 and F4 n=9. Fibroscan® imaging had higher sensitivity and specificity compared with the NAFLD fibrosis score for detection of significant fibrosis, severe fibrosis and cirrhosis (table 15). The results for cirrhosis should be treated with caution due to the small number of study participants in this category (n=9).

Table 15: diagnostic accuracy of Fibroscan® imaging compared with the NAFLD fibrosis score[33]

	Fibroscan® (direct imaging test)	NAFLD fibrosis score (indirect test)
Significant fibrosis (F2-4)		
Sensitivity	0.75	0.52
Specificity	0.93	0.89
Advanced fibrosis (F3-4)		
Sensitivity	0.96	0.63
Specificity	0.90	0.82
Cirrhosis (F4)		
Sensitivity	1.00	0.67
Specificity	0.76	0.72

The third study compared Fibroscan® imaging with the NAFLD fibrosis, APRI and FIB4 scores in 452 consecutively recruited patients with biopsy-proven NAFLD[34]. Clinicians were blinded to patient data and other test results. Up to three months elapsed between the Fibroscan® examination and the liver biopsy or blood tests. The authors used the NAFLD Activity Score to stage fibrosis, describing stages as significant fibrosis (F≥2), advanced fibrosis (F≥3) and cirrhosis (F4). Sensitivity and specificity were only reported for detection of advanced fibrosis; a quarter of study participants had this stage of fibrosis on biopsy: F0 n=39, F1 n=123, F2 n=118, F3 n=114 and F4 n=58. The study authors calculated the diagnostic accuracy of the tests for an optimal cut-off that maximised

sensitivity and specificity for each test. In this study Fibroscan® imaging had higher sensitivity than the NAFLD fibrosis, APRI and FIB4 scores for detection of advanced fibrosis (table 16). Specificity was lower for Fibroscan® imaging compared with the FIB4 and APRI scores for detection of advanced fibrosis.

Table 16: diagnostic accuracy of Fibroscan® imaging compared with the NAFLD fibrosis, APRI and FIB4 scores for detection of advanced fibrosis (F≥3) in patients with NAFLD[34]

	Fibroscan® (direct imaging test)	NAFLD fibrosis score (indirect test)	APRI (indirect test)	FIB4 (indirect test)
Cut-off	8.7kPa	-1.04	0.56	1.52
Sensitivity	0.88	0.77	0.61	0.76
Specificity	0.63	0.60	0.76	0.67

Summary:

- In one prospective diagnostic cohort study (n=404) Fibroscan® imaging had statistically significantly higher sensitivity and specificity than the indirect APRI and FIB4 scores for detecting cirrhosis.
- In a prospective diagnostic cohort study (n=88) Fibroscan® imaging had higher sensitivity and specificity compared with the indirect NAFLD fibrosis score for detection of significant fibrosis, severe fibrosis and cirrhosis.
- In one prospective study (n=452) Fibroscan® imaging had higher sensitivity for the detection of advanced fibrosis than the indirect NAFLD fibrosis, APRI and FIB4 scores. In this study Fibroscan® imaging had lower specificity than the FIB4 and APRI scores for detection of advanced fibrosis. These results should be treated with caution as up to three months may have elapsed between tests.

7.6. Fibroscan® imaging compared with ARFI elastography

A meta-analysis compared the two direct imaging tests of interest – Fibroscan® and ARFI elastography[35]. Thirteen studies with 1,163 participants with chronic liver disease were included in the meta-analyses. The included studies were judged to have low overall risk of bias using the QUADAS-2 appraisal tool. The two imaging tests were compared for detection of significant fibrosis and cirrhosis (table 17). No substantial differences in sensitivity or specificity were found between Fibroscan® and ARFI elastography for detection of significant fibrosis or cirrhosis.

Table 17: summary diagnostic accuracy for Fibroscan® imaging compared with ARFI elastography in patients with liver disease of varied aetiologies[35]

	Fibroscan® (direct imaging test)	ARFI elastography (direct imaging test)
Significant fibrosis (F≥2)		
No. of studies	10	10
Cut-off*	6.1–9.1kPa	1.20–1.43m/s
Summary sensitivity (95% CI)	0.74 (0.66 to 0.80)	0.78 (0.72 to 0.83)
Summary specificity (95% CI)	0.83 (0.75 to 0.89)	0.84 (0.75 to 0.90)
Cirrhosis (F=4)		
No. of studies	13	13
Cut-off*	9.1–16.5kPa	1.45–2.05m/s
Summary sensitivity (95% CI)	0.87 (0.79 to 0.92)	0.89 (0.80 to 0.94)
Summary specificity (95% CI)	0.87 (0.81 to 0.91)	0.87 (0.82 to 0.91)

**Range of thresholds used in studies included in the meta-analysis*

Summary:

- In a meta-analysis of 13 studies (n=1,163) sensitivity and specificity were similar for Fibroscan® imaging and ARFI elastography for the detection of significant fibrosis and cirrhosis in patients with chronic liver disease.

7.7. ARFI elastography compared with indirect fibrosis tests

Two prospective diagnostic cohort studies compared ARFI elastography with indirect fibrosis tests in patients with chronic liver disease[36, 37].

The first study compared ARFI elastography with the indirect APRI score, FIB4 score and AST:ALT ratio in 171 patients with liver disease of varying aetiologies[36]. This study used mini-laparoscopic liver biopsy as the reference standard and defined fibrosis stages using the Ishak system: significant fibrosis F2-3, advanced fibrosis F4-5, cirrhosis F=6. Only twelve study participants had cirrhosis on biopsy: F0 n=42, F1 n=53, F2 n=20, F3 n=16, F4 n=14, F5 n=14 and F6 n=12. All tests were performed within six days of the biopsy. Results were reported for detection of compensated cirrhosis and 'high grade' fibrosis – it is not specified in the study what this means. For detection of compensated cirrhosis ARFI elastography had lower sensitivity than the FIB4 score, and higher sensitivity than the APRI score and AST:ALT ratio (table 18). For detection of high-grade fibrosis ARFI elastography had lower sensitivity than the FIB4 score, similar sensitivity to the APRI score and higher sensitivity than the AST:ALT ratio. Specificity of ARFI elastography for detection of high-grade fibrosis was greater than any of the indirect tests evaluated.

Table 18: diagnostic accuracy of ARFI elastography compared with the AST:ALT ratio, APRI and FIB4 scores in patients with chronic liver disease[36]

	ARFI elastography (direct imaging test)	FIB4 (indirect test)	APRI (indirect test)	AST:ALT (indirect test)
High-grade fibrosis				
Cut-off	1.85m/s	1.83	0.75	0.86
Sensitivity	0.66	0.74	0.66	0.57
Specificity	0.85	0.77	0.79	0.73
Compensated cirrhosis				
Cut-off	1.94m/s	1.98	0.94	0.88
Sensitivity	0.82	0.89	0.66	0.63
Specificity	0.83	0.77	0.82	0.71

The second prospective diagnostic study compared ARFI elastography with the indirect APRI and FIB4 scores in 51 consecutively recruited patients with hepatitis C[37]. Blood tests were conducted in the same week as the reference standard, but ARFI elastography was conducted up to six months later. Results for all three tests were not reported for every stage of fibrosis. Clinicians reviewing test results were blinded to the outcome of other tests and clinical data, and nine participants had cirrhosis on biopsy: F0 n=8, F1 n=15, F2 n=10, F3 n=9 and F4 n=9. For detection of significant fibrosis, advanced fibrosis and cirrhosis, ARFI elastography had higher sensitivity and similar specificity compared with the indirect fibrosis test reported for that stage of fibrosis (table 19). The results for detection of cirrhosis should be treated with caution due to the small number of patients in this category.

Table 19: diagnostic accuracy of ARFI elastography compared with the APRI and FIB4 scores in patients with hepatitis C[37]

	ARFI elastography (direct imaging test)	APRI (indirect test)	FIB4 (indirect test)
Significant fibrosis (F≥2)			
Cut-off	1.31m/s	0.86	-
Sensitivity	0.89	0.68	-
Specificity	0.87	0.87	-
Advanced fibrosis (F≥3)			
Cut-off	1.68m/s	-	3.25
Sensitivity	0.94	-	0.61
Specificity	0.91	-	0.94
Cirrhosis (F=4)			

Cut-off	1.95m/s	1.71	-
Sensitivity	1.00	0.67	-
Specificity	0.95	0.93	-

Summary:

- In a prospective diagnostic cohort study (n=171) ARFI elastography had lower sensitivity and higher specificity than the FIB4 score for detection of high-grade fibrosis and compensated cirrhosis. Both sensitivity and specificity were higher for ARFI elastography compared with the AST:ALT ratio for detection of high-grade fibrosis and compensated cirrhosis.
- In a second prospective diagnostic cohort study (n=51) with a high risk of bias due to the delay between tests, ARFI elastography had higher sensitivity and similar specificity compared with the indirect APRI and FIB4 scores for detection of significant fibrosis, advanced fibrosis and cirrhosis.

8. Safety

In the studies considered in the clinical and cost-effectiveness sections of this review, no adverse events were identified relating to the use of non-invasive liver fibrosis tests. The consequences of false positive and false negative findings were discussed at the beginning of the clinical effectiveness section (section 7).

9. Cost effectiveness

The HTA described in section 7.1 evaluated the cost-effectiveness of non-invasive liver tests and test strategies for liver fibrosis and cirrhosis in patients with hepatitis C, hepatitis B, alcohol-related liver disease or NAFLD[29]. The diagnostic accuracy of each test was derived from the systematic review and meta-analysis in the HTA.

The results for non-invasive fibrosis tests of interest to this evidence note (figure 1) – or any testing strategy consisting exclusively of these tests[a] – were extracted from the HTA and incremental cost-effectiveness ratios (ICERs) were recalculated. **These tests may not have been the best option indicated in the original HTA analysis which included a wider range of fibrosis tests.** Literature was not available to inform cost-effectiveness analyses for all tests of interest in all liver disease patient groups, for example hyaluronic acid was the only serum biomarker of interest that was included in the hepatitis B analyses.

Decision-analytic models were developed in the HTA to assess the cost-effectiveness of non-invasive fibrosis tests for each liver disease aetiology. Long-term consequences following diagnosis were estimated through a series of Markov models and included in the analysis as the diagnosis is expected to affect future treatment decisions or behaviour change in patients with chronic liver disease. Health outcomes were measured using quality-adjusted-life-years (QALYs). Costs were

[a]Not all possible combinations of the tests of interest in figure 1 were reported in the HTA. Due to the large number of tests included in the HTA the authors included only the two most cost-effective single tests in each category (indirect, direct, imaging) in the sequential strategies.

estimated from an NHS perspective and used 2012 prices (GBP, £). Unless otherwise stated, the analysis was conducted over a lifetime horizon and both costs and QALYs were discounted at an annual rate of 3.5%. Transition probabilities between health states in the model were derived from UK data.

9.1. Hepatitis B

HBeAg-positive and HBeAg-negative patients with hepatitis B and suspected fibrosis or cirrhosis who would normally have a liver biopsy to assess eligibility for antiviral treatment were included in this analysis. A METAVIR score of F≥2 equated to a positive test result and treatment with antiviral agents would commence at this stage. For patients with a METAVIR score F<2, 10% were assumed to receive treatment for necrosis and inflammation of the liver and 90% go on to watchful waiting. The watchful waiting option incorporated a re-test after two years with the same test.

The model assumed that:

- the two-year re-test for patients on watchful waiting had perfect accuracy,
- patients could not regress to an earlier health state (less severe disease) following antiviral therapy,
- patients with no fibrosis or mild fibrosis (F0-1) received the same benefit from treatment as patients with more advanced fibrosis or cirrhosis,
- HBeAg-positive patients developed disease at age 31 and were 70% male; HBeAg-negative patients developed disease at age 40 and were 90% male, and
- the average prevalence of fibrosis (F≥2) in the population was 54%.

The diagnostic accuracy of the tests of interest from the HTA meta-analysis on hepatitis B are reported in appendix 3 table I. Separate economic models were constructed for HBeAg-positive and HBeAg-negative patient cohorts as the natural history differs and therefore starting age, transition probabilities and relative risks (RRs) from treatment differed.

HBeAg-positive patients

Results from the base case analyses for the fibrosis tests of interest, and strategies made up exclusively of the tests of interest, are presented in table 20. There were three possible cost-effective strategies in this patient group: test and treat nobody, test with direct serum biomarker hyaluronic acid and treat patients according to test outcome, and treat everyone without testing. Not testing or treating anyone resulted in 9.64 QALYs at a cost of £37,831. Hyaluronic acid was associated with an increase in QALYs and costs, at an ICER of £20,422 per QALY – indicating it was potentially cost effective. Treating everyone with antiviral therapy without testing also provide additional QALYs, but at an ICER of £43,077 per QALY.Applying commonly accepted cost effectiveness thresholds would suggest that hyaluronic acid is the preferred test strategy.

Table 20: base case results for non-invasive tests for detection of fibrosis (F≥2) in HBeAg-positive patients (all incremental costs, QALYs and ICERs were re-calculated based on the original HTA analysis)

Test strategy	Test type	Cost, £	QALY	Incremental cost, £	Incremental QALY	ICER, £
Treat no-one	-	37,831	9.64	-	-	-
FIB4 (high cut-off)	Indirect	73,028	11.33			Extendedly dominated
APRI (high cut-off)	Indirect	75,139	11.45			Extendedly dominated
Fibroscan®	Direct imaging	79,004	11.61			Extendedly dominated
Hyaluronic acid	Direct biomarker	79,084	11.66	41,253	2.02	20,422
FIB4 (low cut-off)	Indirect	81,347	11.66			Dominated
ARFI elastography	Direct imaging	83,487	11.71			Extendedly dominated
APRI (low cut-off)	Indirect	83,770	11.75			Extendedly dominated
Treat all	-	101,484	12.18	101,484	0.52	43,077

HBeAg-negative patients

In this patient group the base case result suggests that none of the non-invasive fibrosis test of interest are cost-effective (table 21). The optimal strategy would be either treat no-one at cost-effectiveness thresholds lower than £28,204, or treat everyone without testing. This is consistent with results from the original HTA analysis in which the probability of a 'treat all' strategy being optimum at a £30,000 threshold was 38%[29].

Table 21: base case results for non-invasive tests for detection of fibrosis (F≥2) in HBeAg-negative patients (all incremental costs, QALYs and ICERs were re-calculated based on the original HTA analysis)

Test strategy	Test type	Cost, £	QALY	Incremental cost, £	Incremental QALY	ICER, £
Treat no-one	-	37,579	8.83	-	-	-
FIB4 (high cut-off)	Indirect	67,267	9.56			Extendedly dominated
APRI (high cut-off)	Indirect	69,428	9.70			Extendedly dominated
Fibroscan®	Direct imaging	73,007	9.93			Extendedly dominated
Hyaluronic acid	Direct biomarker	73,448	9.96			Extendedly dominated
FIB4 (low cut-off)	Indirect	75,648	10.01			Extendedly dominated

ARFI elastography	Direct imaging	77,512	10.10			Extendedly dominated
APRI (low cut-off)	Indirect	78,083	10.13			Extendedly dominated
Treat all	-	96,525	10.92	58,946	2.09	28,204

9.2. Hepatitis C

For hepatitis C, a similar methodology was used to the hepatitis B analysis. The diagnostic accuracy reported in the meta-analysis for non-invasive fibrosis tests of interest is described in appendix 3 table II. The average fibrosis prevalence (F≥2) estimate used in the model was 53%.

The base case results for the hepatitis C population across the range of testing strategies are displayed in table 22. Based on these results there are two possible cost-effective strategies: ARFI elastography or treat everyone without testing. With an ICER of £8,573 the treat all strategy is the optimal strategy if commonly accepted cost-effectiveness thresholds are applied. The original HTA analysis concluded that the probability of the 'treat all' strategy being the optimal strategy given a threshold of £20,000 was 45%.

Sensitivity analyses showed the cost-effectiveness of the 'treat all' strategy to be sensitive to the reduction of treatment effectiveness in patients with mild fibrosis.

Table 22: base case results for non-invasive tests for detection of fibrosis (F≥2) in patients with hepatitis C (all incremental costs, QALYs and ICERs were re-calculated based on the original HTA analysis)

Test strategy	Test type	Cost, £	QALY	Incremental cost, £	Incremental QALY	ICER, £
Treat no one	-	54,878	12.45			Dominated
APRI (high cut-off)	Indirect	47,525	14.14			Dominated
FIB4	Indirect	47,900	14.15			Dominated
ELF test™ (high cut-off)	Direct biomarker	47,846	14.17			Dominated
Hyaluronic acid (high cut-off)	Direct biomarker	48,969	14.18			Dominated
ARFI elastography	Direct imaging	47,126	14.25	-	-	-
AST:ALT ratio	Indirect	48,629	14.26			Dominated
FIB4 (high cut-off)	Indirect	48,158	14.27			Dominated
APRI	Indirect	47,522	14.28			Dominated
Fibroscan®	Direct imaging	47,449	14.28			Extendedly dominated
Hyaluronic acid	Direct biomarker	48,013	14.30			Dominated

P3NP	Direct biomarker	47,921	14.30			Extendedly dominated
Hyaluronic acid (low cut-off)	Direct biomarker	48,824	14.34			Dominated
ELF test™	Direct biomarker	48,232	14.34			Extendedly dominated
APRI (low cut-off)	Indirect	48,713	14.40			Extendedly dominated
ELF test™ (low cut-off)	Direct biomarker	49,041	14.44			Extendedly dominated
FIB4 (low cut-off)	Indirect	49,407	14.48			Extendedly dominated
Treat all	-	51,241	14.73	4,115	0.48	8,573

9.3. Alcohol-related liver disease

The alcohol-related liver disease analysis was based on people suspected of having alcohol-related steatohepatitis with cirrhosis. Only two tests of interest were included in the systematic review for detecting liver fibrosis and cirrhosis in alcohol-related liver disease: the indirect APRI score (high cut-off) and Fibroscan® imaging.

Current management of alcohol-related cirrhosis focuses on alcohol abstinence and prevention of cirrhosis complications. As UK and European guidelines do not recommend a specific treatment for patients with alcohol-related liver disease and cirrhosis, the focus of this model was on the health economic impact of diagnosis and resultant abstinence, assuming that abstinence increases following diagnosis.

In the model, patients receiving a positive test result were assumed to receive monitoring for hepatocellular carcinoma (HCC), varices, ascites and hepatic encephalopathy, plus lifestyle advice. Patients with a negative test result received lifestyle advice only, including recommendations to abstain from or reduce alcohol consumption. The model also assumed that:

- 31% of patients with a negative liver biopsy become abstinent and 62% of those with a positive result become abstinent,
- abstinence rates following diagnosis with a non-invasive fibrosis test would be 10% lower than after liver biopsy, and
- 20% of patients who continue to drink after diagnosis develop cirrhosis.

The diagnostic accuracy of the tests of interest in this analysis did not converge in the bivariate random-effects model (appendix 3 table III). The average prevalence of cirrhosis (F=4) was estimated as 37%.

The base case results are reported in table 23. Fibroscan® was potentially cost-effective offering 9.02 QALYs at a cost of £20,009. The next best alternative, that offered a slight increase in QALYs but at a higher cost, was treating all patients as having cirrhosis with an ICER of £22,906 per QALY.

Within sensitivity analyses the base case results were sensitive to changes in the probability of developing cirrhosis and increases in abstinence rates following testing.

Table 23: base case results for non-invasive tests for detection of cirrhosis (F=4) in patients with alcohol-related liver disease (all incremental costs, QALYs and ICERs were re-calculated based on the original HTA analysis)

Test strategy	Test type	Cost, £	QALY	Incremental cost, £	Incremental QALY	ICER, £
APRI (high cut-off)	Indirect	22,463	8.87			Dominated
Fibroscan®	Direct imaging	20,009	9.02	-	-	-
Treat all (HCC screening)	-	31,004	9.50	10,995	0.48	22,906

9.4. Non-alcoholic fatty liver disease

The population considered for this analysis were patients suspected of having NASH with fibrosis or cirrhosis. The lack of published studies with relevant data on treatments, clinical effectiveness, costs and QALY estimates for patients with NASH and fibrosis limited the modelling approach in this analysis. A probabilistic decision model was constructed to assess the cost per correct diagnosis of non-invasive fibrosis tests compared with liver biopsy. It was assumed that liver biopsy had perfect sensitivity and specificity. Using a hypothetical cohort of 1,000 patients with NASH and suspected liver fibrosis, the incremental cost per correct diagnosis associated with each test compared with the next best alternative was estimated.

The diagnostic accuracy for the tests of interest in patients with NASH and fibrosis (F≥3) is summarised in appendix 3 table IV. Fibrosis prevalence (F≥3) was estimated at 19% of people tested.

The cost per correct diagnosis for the tests of interest are reported in table 24. Not taking into account liver fibrosis stage, imaging using ARFI elastography provided the highest rate of true positives (TP) with an incremental cost per additional correct positive case detected of £8,600 compared with the serum biomarker hyaluronic acid. The NAFLD fibrosis score (high cut-off) offered the highest rate of true negatives (TN) at an incremental cost per additional correct negative case detected of £183 compared with the FIB4 score (high cut-off).

Table 24: results of cost per correct diagnosis analysis in patients with NASH in a cohort of n=1,000
(all incremental costs and incremental costs per correct diagnosis were re-calculated based on the original HTA analysis)

Test	Test type	No. of true positives	Test unit cost, £	No. of incremental correct diagnoses	Incremental cost (test only), £	Incremental cost per correct diagnosis (£/correct diagnosis gained)
Incremental cost per true positive (TP) diagnosis						
FIB4 (high cut-off)	Indirect	71	4.40			Dominated
NAFLD fibrosis score (high cut-off)	Indirect	75	4.95			Dominated
APRI	Indirect	76	4.05			Dominated
AST:ALT ratio (high cut-off)	Indirect	88	0.90			Dominated
NAFLD fibrosis score	Indirect	134	20.85			Dominated
AST-ALT (low cut-off)	Indirect	149	0.90	-	-	-
FIB4	Indirect	149	21.09			Dominated
ELF test™	Direct biomarker	151	108.00			Dominated
NAFLD fibrosis score (low cut-off)	Indirect	151	4.95			Dominated
FIB4 (low cut-off)	Indirect	159	4.40	10	3,500	350
Hyaluronic acid	Direct biomarker	165	8.00	6	3,600	600
ARFI elastography	Direct imaging	170	51.00	5	43,000	8,600
Liver biopsy	Reference standard	189	956.61	19	905,610	47,664
Incremental cost per true negative (TN) diagnosis						
NAFLD fibrosis score (low cut-off)	Indirect	535	4.95			Dominated
AST:ALT ratio (low cut-off)	Indirect	568	0.90			Dominated
FIB4 (low cut-off)	Indirect	603	4.40			Dominated
Hyaluronic acid	Direct biomarker	666	8.00			Dominated
APRI	Indirect	668	4.05			Dominated

ARFI elastography	Direct imaging	726	51.00			Dominated
ELF test™	Direct biomarker	730	108.00			Dominated
AST:ALT ratio (high cut-off)	Indirect	740	0.90	-	-	-
FIB4	Indirect	754	21.09			Dominated
NAFLD fibrosis score	Indirect	780	20.85			Dominated
FIB4 (high cut-off)	Indirect	783	4.40	43	3,500	80
NAFLD fibrosis score (high cut-off)	Indirect	786	4.95	3	550	183
Liver biopsy	Reference standard	811	956.61	25	952,000	38,080

9.5. Cirrhosis irrespective of aetiology

In one analysis the population of interest was people suspected of having cirrhosis, irrespective of aetiology. The cost-effectiveness of non-invasive fibrosis tests for diagnosing cirrhosis was assessed using a decision-tree model and adopted a time horizon of four years. The model was based on the recommended management of patients with cirrhosis: screening for oesophageal varices and ascites, and monitoring for HCC.

The diagnostic accuracy of the tests of interest for detecting cirrhosis irrespective of aetiology is reported in appendix 3 table V. The average cirrhosis prevalence used in the model was 20%.

Base case results for the fibrosis tests of interest are presented in table 25. The option offering the highest health benefit was the ELF test™ with an ICER of £10,333 per QALY.

Table 25: base case results for non-invasive tests for detection of cirrhosis (F=4) in patients with liver disease of any aetiology (all incremental costs, QALYs and ICERs were re-calculated based on the original HTA analysis)

Test strategy	Test type	Cost, £	QALY	Incremental cost, £	Incremental QALY	ICER, £
FIB4 (high cut-off)	Indirect	24,921	2.10			Dominated
APRI (high cut-off)	Indirect	24,920	2.10	-	-	-
AST:ALT ratio	Indirect	24,929	2.11			Extendedly dominated
ELF test™ (high cut-off)	Direct biomarker	25,024	2.11			Dominated

APRI (combined cut-off) + Fibroscan®	Indirect plus direct imaging	24,942	2.13			Extendedly dominated
P3NP	Direct biomarker	24,976	2.135			Dominated
APRI (low cut-off)	Indirect	24,958	2.14			Dominated
FIB4 (combined cut-off) + Fibroscan®	Indirect plus direct imaging	24,951	2.15			Dominated
APRI	Indirect	24,973	2.15			Dominated
FIB4	Indirect	24,960	2.15			Dominated
Hyaluronic acid	Direct biomarker	24,946	2.15	26	0.052	500
FIB4 (low cut-off)	Indirect	24,972	2.15			Extendedly dominated
ARFI elastography	Direct imaging	25,001	2.16			Dominated
ELF test™ (combined cut-off) + Fibroscan®	Direct biomarker plus direct imaging	25,059	2.16			Dominated
Fibroscan®	Direct imaging	24,988	2.16	42	0.012	3,500
ELF test™ (low cut-off)	Direct biomarker	25,070	2.16			Dominated
ELF test™	Direct biomarker	25,050	2.17	62	0.006	10,333

10. Conclusion

The best quality, and greatest quantity, of evidence on the diagnostic accuracy of non-invasive liver fibrosis tests related to direct imaging tests – Fibroscan® and ARFI elastography. There was variation and lack of reporting of threshold values used to define stages of fibrosis in many of the identified studies and all the included studies evaluated the performance of non-invasive direct and/or indirect fibrosis tests in a secondary or tertiary care setting.

Direct fibrosis tests (serum biomarkers and imaging) had high sensitivity and moderate–high specificity for the detection of fibrosis and cirrhosis in patients with hepatitis C or NAFLD. Evidence on the diagnostic accuracy of direct fibrosis tests in patients with hepatitis B or alcohol-related liver disease was limited, suggesting a need for further research in these populations.

There were few studies reporting direct comparisons between direct fibrosis tests. Evidence on the diagnostic accuracy of Fibroscan® compared with ARFI elastography and the ELF test™ came from

two meta-analyses with low risk of bias. For the detection of cirrhosis, Fibroscan® had similar diagnostic accuracy compared with ARFI elastography and greater accuracy than the ELF test™. In two primary studies providing additional evidence, Fibroscan® had higher specificity than the ELF test™ (sensitivity was similar or lower for Fibroscan®) for the detection of fibrosis, and statistically significantly higher sensitivity and specificity for detection of cirrhosis compared with serum hyaluronic acid.

Comparisons between direct and indirect fibrosis tests were limited to a small number of primary studies. Overall, Fibroscan® and the ELF test™ appear to have improved diagnostic accuracy compared with indirect FIB4 and APRI scores for the detection of fibrosis and cirrhosis.

No studies were identified that compared serum biomarkers with ARFI elastography, P3NP with any other non-invasive liver fibrosis test, or serum biomarkers with each other, therefore no conclusions could be reached about the relative accuracy of these tests.

Cost-effectiveness data on liver fibrosis tests of interest was extracted from a large UK-based HTA for each liver disease patient population separately. The analyses were subject to uncertainty and sensitive to changes in parameters and model assumptions. Not all tests of interest were included in each cost-effectiveness analysis and the tests of interest may not have been the most cost-effective option in the full HTA which covered a wider range of fibrosis tests.

In patients with HBeAg-positive hepatitis B the most cost effective strategy was to test for fibrosis using the direct serum biomarker hyaluronic acid. For patients with hepatitis C or HBeAg-negative hepatitis B the most cost-effective strategy was to treat all patients without testing for fibrosis. In patients with alcohol-related liver disease Fibroscan® offered 9.02 QALYs on average at a cost of £20,009; a strategy which treats all patients with alcohol-related liver disease as having cirrhosis offered a higher number of QALYs at a higher cost. The analysis in patients with NAFLD was limited to cost per correct diagnosis; in this case ARFI elastography provided the highest rate of true positives and the indirect NAFLD fibrosis score (high cut-off) offered the highest rate of true negatives detected. In an analysis that assessed cost-effectiveness of non-invasive tests for detecting cirrhosis irrespective of liver disease aetiology, the ELF test™ was the most cost-effective option over a four year time horizon.

11. Identified research gaps

Future research on non-invasive tests for liver fibrosis and cirrhosis should include prospective diagnostic cohort studies that:

- include patients with chronic liver disease of different aetiologies to reflect clinical practice,
- directly compare the diagnostic accuracy of serum biomarkers (ELF test™, hyaluronic acid, P3NP), imaging tests (Fibroscan®, ARFI elastography) and indirect fibrosis tests,
- measure patient outcomes such as change in treatment plan, regression or progression of fibrosis,
- explore patient reported outcome measures (PROMS) and patient experiences of liver fibrosis testing, including liver biopsy, or
- are conducted in a primary/community care setting.

Equality and diversity

Healthcare Improvement Scotland is committed to equality and diversity in respect of the nine equality groups defined by age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion, sex, and sexual orientation.

The process for producing evidence notes has been assessed and no adverse impact across any of these groups is expected. The completed equality and diversity checklist is available on www.healthcareimprovementscotland.org

About evidence notes

Evidence Notes are produced to inform a decision at a particular point in time and are therefore not routinely updated. They will however be considered for review if requested by stakeholders, based upon the availability of new published evidence which is likely to materially change the advice given. For further information about the evidence note process see:

www.healthcareimprovementscotland.org/our_work/clinical__cost_effectiveness/shtg/standard_operating_procedures.aspx

To propose a topic for an evidence note, email shtg.hcis@nhs.net

References can be accessed via the internet (where addresses are provided), via the NHS Knowledge Network www.knowledge.scot.nhs.uk, or by contacting your local library and information service.

A glossary of commonly used terms in Health Technology Assessment is available from htaglossary.net.

Acknowledgements

Healthcare Improvement Scotland and SHTG invited the following individuals and organisations to peer review the draft evidence note.

- Sara Jenks, Consultant in Metabolic Medicine, NHS Lothian
- Alastair MacGilchrist, Consultant Hepatologist, Royal Infirmary of Edinburgh
- Judi Rhys, Chief Executive, British Liver Trust
- Prof Peter Hayes, Professor of Hepatology, University of Edinburgh
- Ian Godber, Lead Clinician, Scottish Clinical Biochemistry Managed Diagnostic Network
- Sarah Smith, Co-Founder, Liver4Life
- Prof John F Dillon, Consultant Hepatologist, Professor of Hepatology and Gastroenterology, University of Dundee
- Prof Philip Newsome, Consultant Hepatologist, University of Birmingham
- Miriam Brazzelli, Senior Research Fellow, Health Services Research Unit, University of Aberdeen

Declarations of interest were sought from all peer reviewers. All contributions from peer reviewers were considered by the group. However the peer reviewers had no role in authorship or editorial control and the views expressed are those of Healthcare Improvement Scotland

Healthcare Improvement Scotland development team:

- Jenny Harbour, Health Services Researcher
- Paul Herbert and Iain Stewart, Information Scientists
- Lucian Gaianu, Health Economist
- Jess Kandulu, Project Officer
- Anne O'Connor, Team Administrator
- Members of the SHTG evidence review committee

© Healthcare Improvement Scotland 2018

References

1. National Institutes for Health. NCI dictionary of cancer terms: biomarker. c2018 [cited 2018 Jan 09]; Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45618.
2. Motola DM, Caravan P, Chung RT, Fuchs BC. Noninvasive biomarkers of liver fibrosis: clincial applications and future directions. Curr Pathobiol Rep. 2014;2(4):245-56.
3. DynaMed. Nonalcoholic fatty liver disease (NAFLD). Ipswich, MA: EBSCO Publishing; 2016.
4. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.
5. Joo SK, Kim JH, Oh S, Kim BG, Lee KL, Kim HY*, et al.* Prospective comparison of noninvasive fibrosis assessment to predict advanced fibrosis or cirrhosis in Asian patients with hepatitis C. J Clin Gastroenterol. 2015;49(8):697-704.
6. Brener S. Transient elastography for assessment of liver fibrosis and steatosis: an evidence-based analysis. 2015 [cited 2018 Mar 23]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4664937/pdf/ohtas-15-1.pdf.
7. Irvine KM, Wockner LF, Shanker M, Fagan KJ, Horsfall LU, Fletcher LM*, et al.* The enhanced liver fibrosis score is associated with clinical outcomes and disease progression in patients with chronic liver disease. Liver Int. 2016;36(3):370-7.
8. Lee MH, Cheong JY, Um SH, Seo YS, Kim DJ, Hwang SG*, et al.* Comparison of surrogate serum markers and transient elastography (Fibroscan) for assessing cirrhosis in patients with chronic viral hepatitis. Dig Dis Sci. 2010;55(12):3552-60.
9. Parkes J, Guha IN, Harris S, Rosenberg WM, Roderick PJ. Systematic review of the diagnostic performance of serum markers of liver fibrosis in alcoholic liver disease. Comp Hepatol. 2012;11(1):5.
10. Newsome PN, Cramb R, Davidson SM, Dillon JF, Foulerton M, Godfrey EM*, et al.* Guidelines on the management of abnormal liver blood tests. Gut. 2018;67(1):6-19.
11. Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. Hepatol. 2017;66(5):1486-501.
12. Stibbe KJM, Verveer C, Francke J, Hansen BE, Zondervan PE, Kuipers EJ*, et al.* Comparison of non-invasive assessment to diagnose liver fibrosis in chronic hepatitis B and C patients. Scand J Gastroenterol. 2011;46(7-8):962-72.
13. Xie Q, Zhou X, Huang P, Wei J, Wang W, Zheng S. The performance of enhanced liver fibrosis (ELF) test for the staging of liver fibrosis: a meta-analysis. PLoS One. 2014;9(4):e92772.
14. NHS Fife. Laboratory services handbook. [cited 2018 May 14]; Available from: http://www.nhsresearchscotland.mtcserver14.com/uploads/tinymce/nhs_fife_reference_ranges_all.pdf.
15. Liu H, Fu J, Hong R, Liu L, Li F. Acoustic radiation force impulse elastography for the non-invasive evaluation of hepatic fibrosis in non-alcoholic fatty liver disease patients: a systematic review & meta-analysis. PLoS One. 2015;10(7):e0127782.
16. MacGilchrist A. Survey of liver services in Scotland. 2013 [cited 2018 Mar 21]; Available from: http://ssg.rcpe.ac.uk/sites/ssg/files/documents/survey%20of%20liver%20services%20in%20Scotland.pdf.
17. BMJ Best Practice. Hepatic steatosis. 2017 [cited 2018 Jan 29]; Available from: http://bestpractice.bmj.com/topics/en-gb/796/pdf/796.pdf.
18. Scottish Public Health Observatory. Hepatitis C. 2017 [cited 2018 Mar 27]; Available from: http://www.scotpho.org.uk/health-wellbeing-and-disease/hepatitis-c/key-points/.
19. Jacobson IM, Davis GL, El-Serag H, Negro F, Trepo C. Prevalence and challenges of liver diseases in patients with chronic hepatitis C virus infection. Clin Gastroenterol Hepatol. 2010;8(11):924-33.
20. National Institute for Health and Care Excellence (NICE). Cirrhosis in over 16s: assessment and management. 2016 [cited 2018 Mar 21]; Available from: https://www.nice.org.uk/guidance/ng50.

21. Scottish Public Health Observatory. Infections: blood-bourne. 2017 [cited 2018 Mar 27]; Available from: http://www.scotpho.org.uk/health-wellbeing-and-disease/infections/data/blood-borne-infections/.
22. Gobel T, Erhardt A, Herwig M, Poremba C, Baldus SE, Sagir A*, et al.* High prevalence of significant liver fibrosis and cirrhosis in chronic hepatitis B patients with normal ALT in central Europe. J Med Virol. 2011;83(6):968-73.
23. ISD Scotland. Alcohol-related hospital statistics: Scotland 2016/17. 2017 [cited 2018 Mar 27]; Available from: http://www.isdscotland.org/Health-Topics/Drugs-and-Alcohol-Misuse/Publications/2017-11-21/2017-11-21-ARHS-Report.pdf.
24. Mathurin P, Bataller R. Trends in the management and burden of alcoholic liver disease. J Hepatol. 2015;62(1 Suppl):S38-46.
25. Sharma NK, Sherker AH. Epidemiology, risk factors and natural history of chronic hepatitis C. 2009 [cited 2018 Jan 29]; Available from: https://pdfs.semanticscholar.org/e779/742b617fa01671270e05a0a588ebefca3e92.pdf.
26. Scottish Public Health Observatory. Chronic liver disease. 2016 [cited 2018 Jan 29]; Available from: http://www.scotpho.org.uk/health-wellbeing-and-disease/chronic-liver-disease/key-points/.
27. National Institute for Health and Care Excellence (NICE). Non-alcoholic fatty liver disease (NAFLD): assessment and management. 2016 [cited 2017 Nov 16]; Available from: https://www.nice.org.uk/guidance/ng49.
28. National Institute for Health and Care Excellence (NICE). Hepatitis B (chronic): diagnosis and management. 2017 [cited 2018 Mar 21]; Available from: https://www.nice.org.uk/guidance/cg165.
29. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodriguez-Peralvarez M*, et al.* Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. Health Technol Assess. 2015;19(9):1-409.
30. Pavlov CS, Casazza G, Nikolova D, Tsochatzis E, Gluud C. Systematic review with meta-analysis: diagnostic accuracy of transient elastography for staging of fibrosis in people with alcoholic liver disease. Aliment Pharmacol Ther. 2016;43(5):575-85.
31. Ragazzo TG, Paranagua-Vezozzo D, Lima FR, Mazo DFD, Pessoa MG, Oliveira CP*, et al.* Accuracy of transient elastography-FibroScan (R), acoustic radiation force impulse (ARFI) imaging, the enhanced liver fibrosis (ELF) test, APRI, and the FIB-4 index compared with liver biopsy in patients with chronic hepatitis C. Clinics. 2017;72(9):516-25.
32. Malik R, Lai M, Sadiq A, Farnan R, Mehta S, Nasser I*, et al.* Comparison of transient elastography, serum markers and clinical signs for the diagnosis of compensated cirrhosis. J Gastroenterol Hepatol. 2010;25(9):1562-8.
33. Aykut UE, Akyuz U, Yesil A, Eren F, Gerin F, Ergelen R*, et al.* A comparison of FibroMeter (TM) NAFLD Score, NAFLD fibrosis score, and transient elastography as noninvasive diagnostic tools for hepatic fibrosis in patients with biopsy-proven non-alcoholic fatty liver disease. Scand J Gastroenterol. 2014;49(11):1343-8.
34. Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B*, et al.* Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. J Hepatol. 2016;65(3):570-8.
35. Bota S, Herkner H, Sporea I, Salzl P, Sirli R, Neghina AM*, et al.* Meta-analysis: ARFI elastograpy versus transient elastography for the evaluation of liver fibrosis. Liver Int. 2013;33(8):1138-47.
36. Pfeifer L, Adler W, Zopf S, Siebler J, Wildner D, Goertz RS*, et al.* Acoustic radiation force impulse elastography: comparison and combination with other noninvasive tests for the diagnosis of compensated liver cirrhosis. Eur J Gastroenterol Hepatol. 2017;29(5):524-30.
37. Silva RG, Schmillevitch J, Nascimento MDA, Miranda MLQ, Brant P, Schulz PO*, et al.* Acoustic radiation force impulse elastography and serum fibrosis markers in chronic hepatitis C. Scand J Gastroenterol. 2014;49(8):986-92.
38. Last J. A dictionary of epidemiology. 4th ed. New York: Oxford University Press; 2001.

Appendix 1: Abbreviations

ALT	alanine transaminase
APRI	aspartate transaminase: platelet ratio index
ARFI	acoustic radiation force impulse
ARLD	alcohol-related liver disease
AST:ALT	aspartate transaminase: alanine transaminase
CI	confidence interval
ELF	enhanced liver fibrosis
FIB4	fibrosis 4
FN	false negative
FP	false positive
HA	hyaluronic acid
HBeAg	hepatitis B e antigen
HCC	hepatocellular carcinoma
HTA	health technology assessment
ICER	incremental cost effectiveness ratio
kPa	kilopascals
LFT	liver function tests
NAFLD	non-alcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NICE	National Institute for Health and Care Excellence
NLR	negative likelihood ratio
NR	not reported
P3NP/PIIINP	N-terminal propeptide type III procollagen
PLR	positive likelihood ratio

QALY	quality adjusted life year
QUADAS	quality assessment of diagnostic accuracy studies
RR	relative risk
TIMP1	tissue inhibitor of metalloproteinases-1
TN	true negative
TP	true positive

Appendix 2: Definitions of diagnostic accuracy terms

Sensitivity: the probability that a person having a disease will be correctly identified by a clinical test, that is the number of true positive results divided by the total number with the disease[38].

Specificity: the probability that a person not having a disease will be correctly identified by a clinical test, that is the number of true negative results divided by the total number of those without the disease[38].

Positive likelihood ratio: the probability that a positive test result will occur in a person with the target condition divided by the probability of a positive test result occurring in a person without the disease, that is the sensitivity divided by one minus specificity[38].

Negative likelihood ratio: the probability that a negative test result will occur in a person with the target condition divided by the probability of a negative test result occurring in a person without the disease, that is the 1-sensitivity divided by specificity[38].

Receiver operating characteristic (ROC) curve: a graph used to assess the ability of a diagnostic test to discriminate between people with or without the target condition. For most diagnostic test data the ROC curve plots sensitivity against 1-specificity for different cut-off values[38]. Area under the ROC curve (AUROC) can be used to compare the diagnostic accuracy of tests when multiple ROC curves are plotted on the same graph.

Appendix 3: Diagnostic accuracy for cost-effectiveness analysis

Diagnostic accuracy measures for direct and indirect fibrosis tests of interest were extracted from an HTA (tables I-V) to inform the cost-effectiveness section.

Table I: diagnostic accuracy of non-invasive tests for detection of fibrosis F≥2 in patients with chronic hepatitis B

Test	No. of studies	Cut-off	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Analysis method used
INDIRECT TESTS					
APRI (low cut-off)	8	0.4–0.6	0.80 (0.68 to 0.88)	0.65 (0.52 to 0.77)	Bivariate random-effects model with correlation between sensitivity and specificity
APRI (high cut-off)	6	1.5	0.37 (0.22 to 0.55)	0.93 (0.85 to 0.97)	Bivariate random-effects model with correlation between sensitivity and specificity
AST:ALT ratio	1	0.67	0.57 (0.51 to 0.64)	0.59 (0.54 to 0.63)	Single study
FIB4 (low cut-off)	4	1.1–1.7	0.68 (0.60 to 0.75)	0.73 (0.67 to 0.79)	Bivariate random-effects model with correlation between sensitivity and specificity
FIB4 (high cut-off)	1	3.25	0.58 *(0.04 to 0.17)**	0.99 (0.96 to 1.00)	Single study
DIRECT SERUM BIOMARKER TESTS					
Hyaluronic acid	1	185.3ng/ml	0.84 (0.73 to 0.91)	0.83 (0.66 to 0.93)	Single study
DIRECT IMAGING TESTS					
ARFI elastography	1	1.33m/s	0.71 (0.59 to 0.80)	0.67 (0.30 to 0.90)	Single study
Fibroscan®	13	6.3–8.9kPa	0.71 (0.62 to 0.78)	0.84 (0.74 to 0.91)	Bivariate random-effects model with correlation between sensitivity and specificity

**An error in confidence interval appears to have been reported in the original HTA[29]*

Table II: diagnostic accuracy of non-invasive tests for detection of fibrosis F≥2 in patients with chronic hepatitis C

Test	No. of studies	Cut-off	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Analysis method used
INDIRECT TESTS					
APRI (low cut-off)	47	0.4–0.7	0.82 (0.77 to 0.86)	0.57 (0.49 to 0.65)	Bivariate random-effects model with correlation between sensitivity and specificity
APRI (high cut-off)	36	1.5	0.39 (0.32 to 0.47)	0.92 (0.89 to 0.95)	Bivariate random-effects model with correlation between sensitivity and specificity
AST:ALT ratio	7	0.6–1	0.44 (0.27 to 0.63)	0.71 (0.62 to 0.78)	Bivariate random-effects model with correlation between sensitivity and specificity
FIB4 (low cut-off)	11	0.6–1.45	0.89 (0.79 to 0.95)	0.42 (0.25 to 0.61)	Random-effects model for sensitivity and specificity without correlation
FIB4 (high cut-off)	9	1–3.25	0.59 (0.43 to 0.73)	0.74 (0.56 to 0.87)	Bivariate random-effects model with correlation between sensitivity and specificity
DIRECT SERUM BIOMARKER TESTS					
Hyaluronic acid	8	34–110ng/ml	0.75 (0.64 to 0.83)	0.75 (0.68 to 0.82)	Bivariate random-effects model with correlation between sensitivity and specificity
ELF test™	1	8.75	0.84 (0.69 to 0.92)	0.70 (0.52 to 0.83)	Single study
ELF test™ (low cut-off)	1	9.55	0.90 (0.85 to 0.93)	0.52 (0.43 to 0.61)	Single study
ELF test™ (high cut-off)	1	11.07	0.47 (0.41 to 0.54)	0.90 (0.83 to 0.94)	Single study
DIRECT IMAGING TESTS					
ARFI elastography	3	1.21–1.34m/s	0.79 (0.75 to 0.83)	0.89 (0.84 to 0.93)	Fixed-effects model for sensitivity and specificity without correlation
Fibroscan®	37	5.2–10.1kPa	0.79 (0.74 to 0.84)	0.83 (0.77 to 0.88)	Bivariate random-effects model with correlation between sensitivity and specificity

Table III: diagnostic accuracy of non-invasive tests for detection of cirrhosis (F=4) in patients with alcohol-related liver disease

Test	No. of studies	Cut-off	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Analysis method used
APRI (high cut-off) [INDIRECT TEST]	1	2	0.40 (0.22 to 0.61)	0.62 (0.41 to 0.79)	Single study
Fibroscan® [DIRECT IMAGING TEST]	6	11.4–25.8kPa	0.86 (0.76 to 0.92)	0.83 (0.74 to 0.89)	Random-effects model for sensitivity and specificity without correlation

Table IV: diagnostic accuracy of non-invasive tests for detection of fibrosis (F≥3) in patients with NASH

Test	No. of studies	Cut-off	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Analysis method used
INDIRECT TESTS					
APRI	4	0.5–1.0	0.40 (0.07 to 0.86)	0.82 *(0.78 to 0.60)**	Random-effects model for sensitivity and fixed-effect model for specificity without correlation
AST:ALT ratio (low cut-off)	4	0.8	0.79 (0.51 to 0.91)	0.70 (0.55 to 0.82)	Bivariate random-effects model with correlation between sensitivity and specificity
AST:ALT ratio (high cut-off)	3	1.0	0.46 (0.29 to 0.65)	0.91 (0.85 to 0.95)	Random-effects model for sensitivity and fixed-effect model for specificity without correlation
FIB4 (low cut-off)	4	1.3–1.92	0.84 (0.75 to 0.90)	0.74 (0.64 to 0.83)	Bivariate random-effects model with correlation between sensitivity and specificity
FIB4 (high cut-off)	2	3.25	0.38 (0.22 to 0.57)	0.97 (0.92 to 0.99)	Bivariate random-effects model with correlation between sensitivity and specificity
NAFLD fibrosis score (low cut-off)	10	–1.455	0.80 (0.67 to 0.89)	0.66 (0.57 to 0.74)	Bivariate random-effects model with correlation between sensitivity and specificity

NAFLD fibrosis score (high cut-off)	9	0.676	0.40 (0.20 to 0.64)	0.97 (0.94 to 0.98)	Bivariate random-effects model with correlation between sensitivity and specificity
DIRECT SERUM BIOMARKER TESTS					
Hyaluronic acid	4	46–50ng/ml	0.88 (0.58 to 0.97)	0.82 (0.75 to 0.87)	Bivariate random-effects model with correlation between sensitivity and specificity
ELF test™	1	10.35	0.80 (0.65 to 0.89)	0.90 (0.84 to 0.94)	Bivariate random-effects model with correlation between sensitivity and specificity
DIRECT IMAGING TESTS					
ARFI elastography	1	4.2m/s	0.90 (0.77 to 0.96)	0.90 (0.82 to 0.94)	Single study
Fibroscan®	8	7.5–10.4kPa	0.82 (0.74 to 0.88)	0.84 (0.78 to 0.89)	Bivariate random-effects model with correlation between sensitivity and specificity

**An error in confidence interval appears to have been reported in the original HTA[29]*

Table V: diagnostic accuracy of non-invasive liver tests for detection of cirrhosis (F=4) irrespective of liver disease aetiology

Test	No. of studies	Cut-off	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Analysis method used
INDIRECT TESTS					
APRI (low cut-off)	27	0.75–1	0.75 (0.71 to 0.8)	0.78 (0.75 to 0.81)	Bivariate random-effects model with correlation between sensitivity and specificity
APRI (high cut-off)	23	2	0.45 (0.37 to 0.52)	0.93 (0.9 to 0.95)	Bivariate random-effects model with correlation between sensitivity and specificity
AST:ALT ratio	13	1	0.49 (0.39 to 0.59)	0.87 (0.75 to 0.94)	Bivariate random-effects model with correlation between sensitivity and specificity
FIB4 (low cut-off)	5	1.45–1.92	0.84 (0.76 to 0.89)	0.71 (0.62 to 0.79)	Bivariate random-effects model with correlation between sensitivity and specificity
FIB4 (high cut-off)	4	3.25–4.44	0.42 (0.2 to 0.69)	0.92 (0.58 to 0.99)	Random-effects model for sensitivity and

					specificity without correlation
DIRECT SERUM BIOMARKER TESTS					
Hyaluronic acid	8	78–237ng/ml	0.81 (0.65 to 0.9)	0.88 (0.8 to 0.94)	Bivariate random-effects model with correlation between sensitivity and specificity
ELF test™	1	9.4	0.93 (0.69 to 0.99)	0.79 (0.67 to 0.88)	Single study
ELF test™ (low cut-off)	1	-	0.90 (0.84 to 0.94)	0.53 (0.46 to 0.59)	Single study
ELF test™ (high cut-off)	1	-	0.52 (0.43 to 0.6)	0.90 (0.85 to 0.93)	Single study
DIRECT IMAGING TESTS					
ARFI elastography	4	1.59–2m/s	0.84 (0.72 to 0.91)	0.77 (0.5 to 0.92)	Random-effects model for sensitivity and specificity without correlation
Fibroscan®	65	9.2–26.5kPa	0.89 (0.86 to 0.91)	0.89 (0.87 to 0.91)	Bivariate random-effects model with correlation between sensitivity and specificity