National Institute for Health and Care Excellence

DIAGNOSTICS ASSESSMENT PROGRAMME

Evidence overview: MRI-based technologies for the assessment of non-alcoholic fatty liver disease

This overview summarises the main issues the diagnostics advisory committee needs to consider. It should be read together with the final scope and the diagnostics assessment report.

1 Aims and scope

The purpose of this assessment is to explore whether using MRI-based technologies to assess non-alcoholic fatty liver disease (NAFLD) is a cost-effective use of NHS resources. MRI could be an option for non-invasive assessment of liver disease to further evaluate the level of fibro-inflammation or fibrosis. Results from MRI assessment could help make decisions about whether a liver biopsy is needed and about the extent of future monitoring. Results may also allow targeted offering of lifestyle interventions or improve uptake and adherence to these interventions to reduce the likelihood of progression to more severe NAFLD.

Background

NAFLD is the term for a range of conditions caused by a build-up of fat in the liver. NAFLD develops in 4 stages:

- Simple fatty liver (steatosis): a largely harmless build-up of fat in the liver cells.
- Non-alcoholic steatohepatitis (NASH): a build-up of fat leading to inflammation.

- Fibrosis: persistent inflammation causing scar tissue to develop in the liver and nearby blood vessels, but the liver still functions normally.
- Cirrhosis: severe scarring from chronic inflammation, causing permanent damage.

People with NASH may have quicker disease progression to fibrosis than those with steatosis. People with liver fibrosis are at increased risk of death, with stage of fibrosis being the most influential predictor of all-cause or liver-related mortality. Once a diagnosis of NAFLD has been made, people are assessed for fibrosis to determine their risk of certain clinical outcomes.

Although there are several non-invasive tests available to assess the stage of liver disease, liver biopsy is considered the gold standard. NASH is diagnosed using biopsy. Approximately 7,000 to 8,000 people per year have a liver biopsy in the UK. Biopsy results are used to decide referral and treatment strategies for people with NAFLD. However, liver biopsy is an invasive procedure that is associated with well-recognised complications, including bleeding and death.

Treatment for NAFLD with no or minimal fibrosis consists of education on risk factors for advanced fibrosis and advice on weight management. According to the NICE guideline on NAFLD: assessment and management, people with advanced fibrosis may be offered pioglitazone or vitamin E, although clinical experts advise that this may not be done in practice. There are currently no treatments available specifically for NAFLD or NASH, but people with NASH or advanced fibrosis may be able to enter clinical trials for new treatments. People with cirrhosis are monitored for end-stage liver disease and liver cancer every 6 months, tested for varices, offered treatment for complications of cirrhosis (for example variceal band ligation), and potentially offered prophylactic treatment depending on comorbidities.

LiverMultiScan

LiverMultiScan is a standalone software application produced by Perspectum that provides quantitative multiparametric analysis of non-contrast MRI. LiverMultiScan is intended to help clinicians diagnose and stage liver disease by non-invasively imaging the liver. LiverMultiScan uses iron-corrected T1 (cT1), proton density fat fraction (PDFF) and T2* MRI protocols for its analyses. cT1 outputs are measured in milliseconds (ms), and correlate with liver fibro-inflammation. MRI PDFF is an MRI estimate of fat content and is expressed as a percentage. T2* is a measure correlated with the iron content of the liver and is used to produce the cT1 scan.

Perspectum has suggested that the normal reference range for MRI PDFF is less than 5.6% liver fat content. The diagnosis indicated by the cT1 output and the clinical recommendations are as follows:

- less than 800 ms: fatty liver
  - no inflammation present
  - reassess with MRI in 3 years
- 800 ms to 875 ms: NASH
  - recommend lifestyle modification
  - manage type 2 diabetes and cardiovascular disease
  - monitor disease status with MRI after 6 months
- more than 875 ms: high-risk NASH
  - reassess with MRI every 6 months
  - consider liver biopsy if cirrhosis is suspected
  - cancer surveillance
  - consider inclusion in NASH therapeutic trials.

Magnetic resonance elastography

Magnetic resonance elastography (MRE) combines MRI with low-frequency vibrations to create a 2D or 3D elastogram showing the stiffness of tissue. In addition to the MRI equipment needed, vibrations are created using an

external mechanical driver that passes vibrations through a flexible tube to a passive driver placed on a person's abdomen over the liver. The driver is manufactured by Resoundant. MRE is used for detecting and evaluating different stages of fibrosis and is usually added to a conventional abdominal MRI protocol. The commercially available version of the platform measures the magnitude of the complex shear modulus of propagating waves. Other investigational modes for MRE have been used in research (see section 5.3.1 of the diagnostics assessment report). MRE outputs are provided in kilopascals (kPa). Resoundant has suggested that MRE liver stiffness outputs can be used to stage liver fibrosis as follows:

- more than 2.9 kPa: any fibrosis
- more than 3.3 kPa: significant fibrosis
- more than 3.9 kPa: advanced fibrosis
- more than 4.8 kPa: cirrhosis.

Decision question

Are MRI-based technologies for assessing non-alcoholic fatty liver disease a cost-effective use of NHS resources?

Populations

People with NAFLD who have not had a diagnosis of advanced fibrosis or cirrhosis and

- have indeterminate results from fibrosis testing
- transient elastography or acoustic radiation force impulse (ARFI) elastography is unsuitable to assess fibrosis
- have discordant results from fibrosis testing.

If data permits, the following could be done:

- subgroup analysis based on which tests for fibrosis have been done before
- subgroup analyses in children or young people.

Interventions

- LiverMultiScan
- MRE.

Comparators

No further testing before a decision about whether to do a biopsy or any other aspect of care.

Healthcare setting

Secondary or tertiary care.

2 Clinical effectiveness evidence

The external assessment group (EAG) did a systematic review to identify evidence on the clinical effectiveness and diagnostic test accuracy of MRI-based technologies for assessing fibrosis, inflammation or steatosis in people with NAFLD who have not had a diagnosis of advanced fibrosis or cirrhosis. It used liver biopsy as the reference standard. It also looked for studies assessing the clinical impact of MRI-based technologies. For details of the inclusion and exclusion criteria, see table 2 of the diagnostics assessment report.

Find the full systematic review results on pages 37 to 66 of the diagnostics assessment report.

Overview of included studies

There were 13 studies reported in 15 publications in the diagnostic test accuracy review. Two studies evaluated LiverMultiScan alone, 10 studies evaluated MRE alone, and 1 study evaluated both technologies. For the clinical impact review, 11 studies reported in 14 publications were included (5 studies of LiverMultiScan and 6 of MRE). All studies in the diagnostic test accuracy review were prospective or retrospective cross-sectional studies. Four additional studies included in the clinical impact review (that were not

also included in the diagnostic test accuracy review) were a prospective cohort study, a randomised controlled trial (RADIcAL1), a retrospective cohort study and a qualitative study. Studies were located in the UK, US, Europe, Japan and South Korea.

More details of the studies can be found in tables 4 (page 43) and 8 (page 60) of the diagnostics assessment report.

RADIcAL1 was a Phase 4 open-label randomised controlled trial comparing LiverMultiScan (n=403) with local standard care (n=399) in people with suspected NAFLD. Suspicion of NAFLD was based on elevated liver function tests, imaging suggestive of fatty liver disease, or the presence of risk factors such as obesity or diabetes. It was done at 13 sites across 4 countries; 7 sites were in the UK (n=253). Before randomisation, 67% of UK patients had transient elastography, and 11% had liver biopsy. People in the trial were not blinded to their study arm. In the intervention arm, people were recommended to have further diagnostic evaluation (such as monitoring of liver enzymes, repeat LiverMultiScan at 6 to 12 months, assessment of liver stiffness, or assessment of response to lifestyle management activities) if the cT1 result was at least 800 ms, or if PDFF was at least 10% (Tonev et al. 2020). However, this was not mandatory and was left at the discretion of the clinician and patient. The primary endpoint was the proportion of people with suspected NAFLD who had liver-related hospital consultations or liver biopsies from the date of randomisation to end of study follow-up. 55 out of 802 people had liver biopsy to confirm diagnosis. In the intervention arm, only those with a high risk of NASH based on LiverMultiScan result had biopsy. Information about the study was from a published protocol (Tonev et al. 2020) and a clinical study report (CSR) provided by Perspectum. No publication or manuscript submitted or accepted for publication were available. For more information on RADIcAL1, see table 7 of the diagnostics assessment report.

No studies of LiverMultiScan in children or young people were identified. Two studies of MRE included children and young people (Trout et al. 2018 and Xanthakos et al. 2014).

Study quality

The studies included for the diagnostic test accuracy review were assessed for risk of bias using the QUADAS 2 tool. An overview of the QUADAS 2 assessment is shown in table 3 and appendix 8 of the diagnostics assessment report.

Xanthakos et al. (2014) was judged as having unclear risk of bias for patient selection because there was a lack of information about recruitment methods and the eligibility criteria used. Forsgren et al. (2020) was judged to have a high risk of bias in the index test domain. This study used cut-offs that were not prespecified and it was unclear whether the index test results were interpreted without knowledge of the results of the reference standard (liver biopsy). The 10 studies judged as having unclear risk of bias in the index test domain did not use prespecified thresholds (but interpreters of index test results were blinded to reference test results). Four studies were considered to have unclear risk of bias in the reference standard domain because of not providing details on whether the interpretation of the reference standard results occurred without knowledge of the index test results. Two studies were judged to have unclear risk of bias in the flow and timing domain. In Kim et al. 2013, the reference standard was done up to 1 year after the index test, and in Hoffman et al. not all people had a liver biopsy.

All the studies included in the reviews considered people with NAFLD who had not had a diagnosis of advanced fibrosis or cirrhosis. However, only 1 study (Eddowes et al. 2018) provided diagnostic test accuracy and clinical impact results for people with NAFLD who had ‘inconclusive’ results from fibrosis testing. The EAG considered this population to provide evidence for people with indeterminate or discordant results, but noted that it was unclear what the exact definition of ‘inconclusive’ used in the study meant. The EAG

also noted that people in this population were already scheduled for a biopsy so may not represent the full population defined in the scope.

The results of the Eddowes study were used to inform the EAG's economic model for the LiverMultiScan test. This was judged at low risk for all domains, except risk of bias related to the index test which was unclear. This was because prespecified thresholds were not used, so the index test results were interpreted without knowledge of the results of the reference standard.

Except for Eddowes et al., the populations of the other studies were considered unclear or high risk regarding applicability for patient selection. This was because of including people with liver disease not caused by NAFLD, or because it was not clear whether the populations included those specified in the scope (indeterminate or discordant results or for whom transient elastography or ARFI is unsuitable).

No studies of MRE explicitly included the scope population. Imajo et al. (2021) was used to provide accuracy estimates in the EAG's model for the MRE test for detecting fibrosis. This study was judged at low risk of bias for all domains except reference standard, which had unclear risk because of uncertainty about whether the reference standard was interpreted without knowledge of the index test result. Imajo et al. (2021) was rated as unclear for applicability concerns for patient selection, but low risk for applicability otherwise.

In 3 studies of MRE (Forsgren et al. 2020, Toguchi et al. 2017, Troelstra et al. 2021), there was a high risk of concern regarding the applicability of the index test. This was because the studies either assessed investigational MRE designs or techniques which may not be consistent with the commercially available device.

In addition to the studies used for the diagnostic test accuracy review, a further 2 cohort studies and RADIcAL1 were included in the clinical impact review. The cohort studies were quality assessed using the National Institute of Health tool (see table 24 in the diagnostics assessment report). The

randomised controlled trial was assessed separately (see appendix 11 in the diagnostics assessment report).

The EAG judged the overall risk of bias for RADIcAL1 as high because there were:

- Concerns about the randomisation process. This was because the trial was open-label and the authors did not present any patient characteristic data specifically for people with NAFLD who had LiverMultiScan and liver biopsy. Also, patient characteristics were not reported for the 2 treatment arms, only for the whole study population.
- A high level of missing data. This was because data on the number of unnecessary liver biopsies avoided was only available for 55 of the 802 people randomised.
- Lack of blinding of assessors.
- Deviation from intended interventions. This was because the trial was open-label. Also, there was limited information about the data analysis used to estimate the effect of assignment to intervention on the number of unnecessary liver biopsies avoided.

RADIcAL1 included people with NAFLD who had not had a diagnosis of advanced fibrosis or cirrhosis. However, it is unclear whether RADIcAL1 included people who had indeterminate results from fibrosis testing, for whom transient elastography or ARFI was unsuitable, or who had discordant results from fibrosis testing.

Diagnostic test accuracy results

Absolute numbers of true and false positive or negative results were not presented in any of the included studies, but data was provided by manufacturers or study authors, calculated by the EAG or obtained from systematic reviews (see table 5 in the diagnostics assessment report).

LiverMultiScan

For LiverMultiScan, cT1 test accuracy was available from 3 studies (figure 1). The EAG noted that Perspectum does not propose that LiverMultiScan is suitable for staging fibrosis. But it does consider that LiverMultiScan can stage NAFLD and distinguish between NASH and high-risk NASH. Perspectum provided diagnostic test accuracy data on the use of LiverMultiScan to detect 6 liver conditions:

- any fibrosis (at least F1 according to the NASH clinical research network [CRN] scoring system)
- significant fibrosis (at least F2)
- steatosis (Brunt grade at least 1)
- steatosis (Brunt grade at least 2)
- NASH (NAFLD activity score [NAS] at least 4, with at least 1 point in both hepatocyte ballooning and lobular inflammation domains)
- advanced NASH (NAS at least 4 with fibrosis at least F2).

In response to a request from the EAG, Perspectum also provided data for people with advanced fibrosis (at least F3). Available data used cT1 cut-off values of 800 ms or 875 ms to indicate a positive result (figure 1). Data for LiverMultiScan MRI PDFF can be found in section 5.3.3 and figure 3 of the diagnostics assessment report.

The EAG considered that the Eddowes et al. (2018) study of LiverMultiScan was the most relevant study for this assessment, and used data from this study in their economic modelling. For diagnosis of fibrosis, sensitivity and specificity were higher for cT1 than PDFF. For diagnosis of steatosis, sensitivity and specificity values were similar between the 2 measures. For the diagnosis of NASH and advanced NASH, sensitivity was estimated to be 64% for both cT1 and PDFF. There was some variation in the specificity estimates from this study for NASH (cT1 67%, PDFF 57%) and advanced NASH (cT1 63%, PDFF 54%). The highest estimate of sensitivity from this study was for use of LiverMultiScan cT1 to detect any level of fibrosis (at least F1) with

sensitivity of 88% (95% confidence interval [CI] 73% to 96%). Other sensitivity estimates ranged from 57% to 80%, and specificity estimates ranged from 0% to 75% (there were no true negatives for the population of 46 people with data for steatosis Brunt grade at least 1).

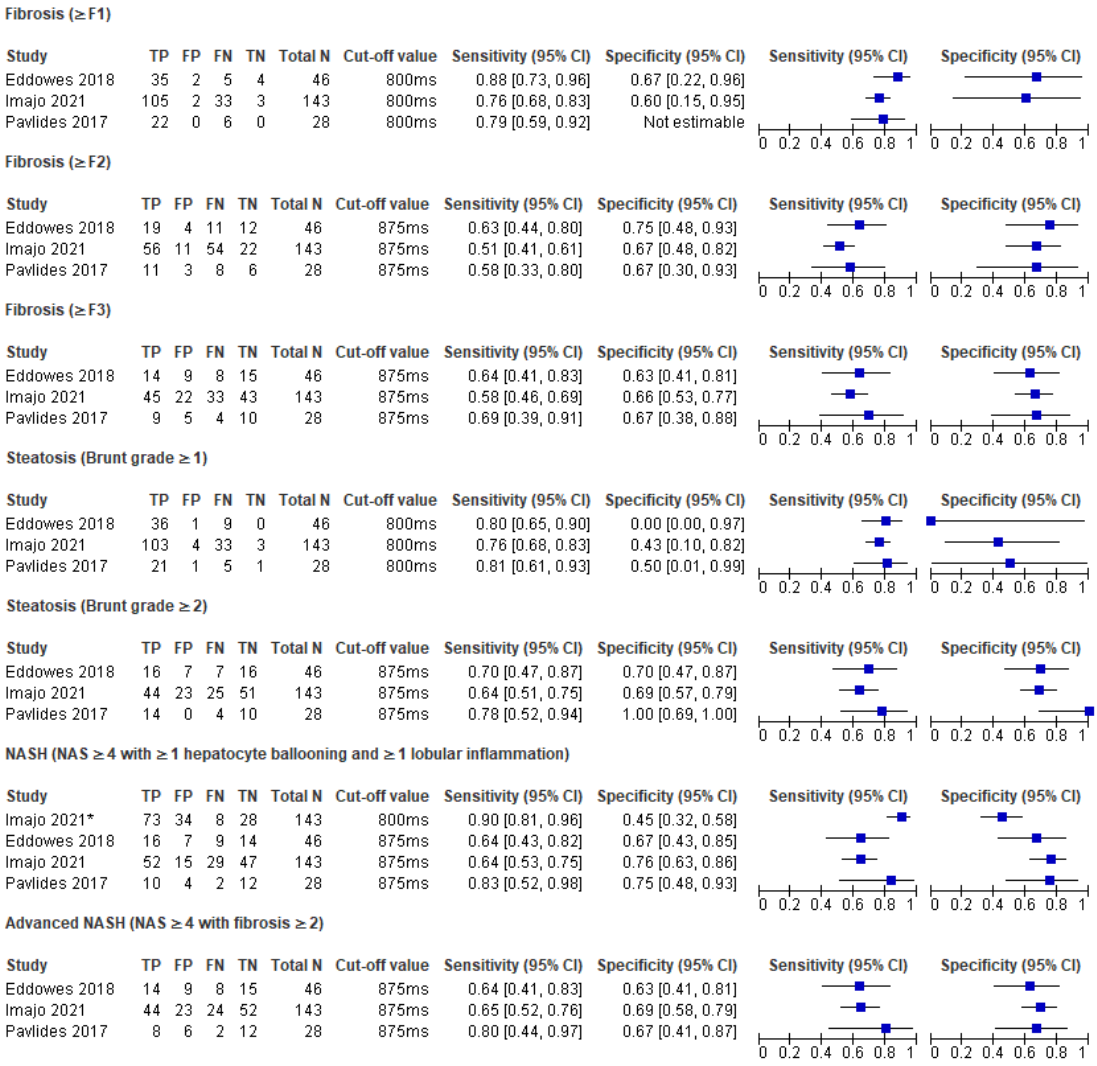

Figure 1 Forest plot showing sensitivity and specificity for LiverMultiScan cT1 for fibrosis, steatosis and stages of NASH at different cut-off values, reported from 3 studies

Data for NASH was available from the Imajo et al. 2021 study for 2 cut-off values, 800 ms and 875 ms. All other studies reporting data for NASH used the 875 ms cut-off value only. Abbreviations: CI, confidence interval; cT1, corrected longitudinal relaxation time; FN, false negative; FP, false positive;

NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; TN, true negative; TP, true positive

Where possible the EAG did meta analyses for outcomes from the included studies of LiverMultiScan (table 1). Meta analysis was only done if results were available from 3 or more studies. Results from the meta analyses suggested that the LiverMultiScan cT1 output is more sensitive and specific than the LiverMultiScan PDFF output, except for steatosis Brunt grade at least 2 (see diagnostics assessment report table 6).

Table 1 LiverMultiScan cT1 diagnostic accuracy meta analyses

Outcome	Definition	Cut-off value	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)
Fibrosis	≥F2	875 ms	54.1 (46.3 to 61.7)	69.0 (56.0 to 79.5)
Fibrosis	≥F3	875 ms	60.2 (50.9 to 68.8)	65.4 (55.8 to 73.9)
Steatosis	Brunt grade ≥1	800 ms	77.3 (71.1 to 82.5)	40.0 (15.8 to 70.3)
Steatosis	Brunt grade ≥2	875 ms	67.3 (58.0 to 75.4)	72.0 (62.7 to 79.6)
NASH	NAS ≥4 with ≥1 in ballooning and inflammation	875 ms	66.1 (57.1 to 74.1)	73.7 (64.2 to 81.5)
Advanced NASH	NAS ≥4 and ≥F2	875 ms	66.0 (56.2 to 74.6)	67.5 (58.5 to 75.4)

All meta analyses based on 3 studies. Abbreviations: CI, confidence interval; ms, milliseconds; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis

MRE

Data for MRE was available for 4 studies (figure 2). Diagnosis definitions were consistent between studies but cut-off values varied, making it difficult compare the results of the studies. No studies of MRE explicitly included people with indeterminate or discordant results from previous fibrosis testing.

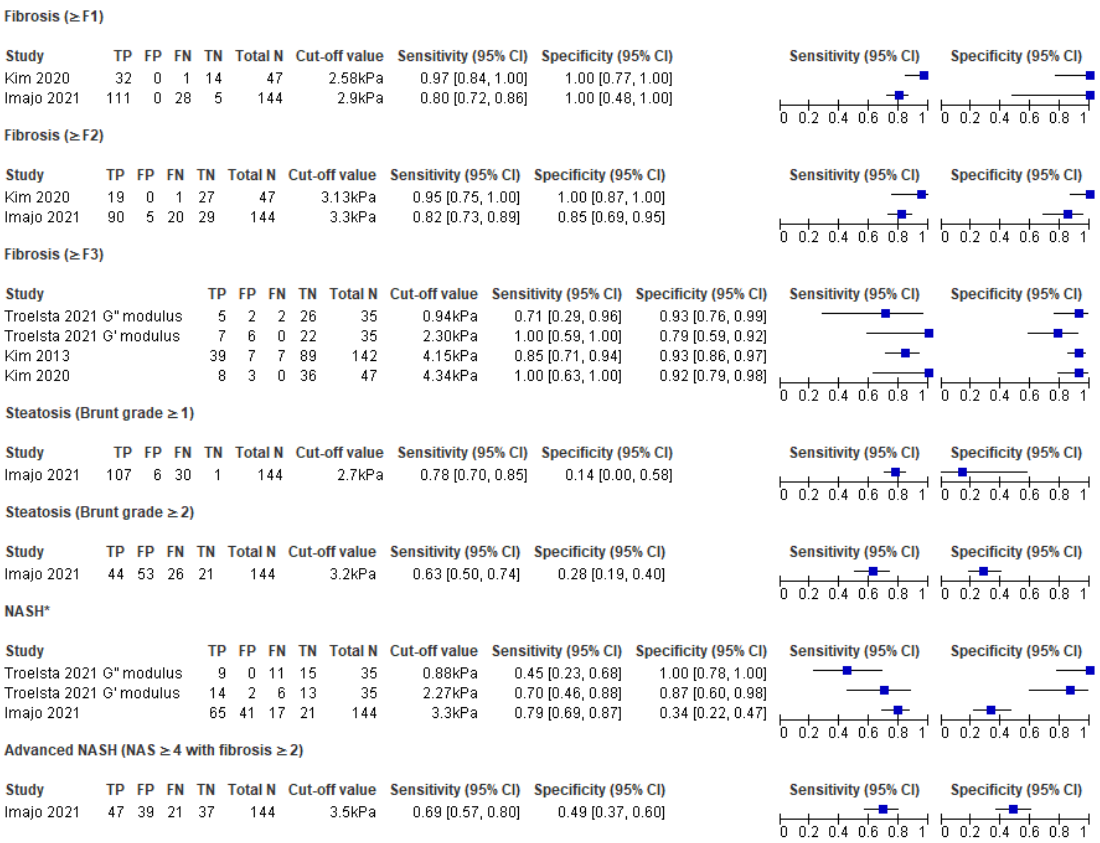

Figure 2 Forest plot displaying 2x2 data, sensitivity and specificity for MRE for fibrosis and steatosis at different cut-off values, reported from 3 studies

NASH was defined in the Imajo et al. (2021) study as NAS at least 4 with at least 1 hepatocyte ballooning and at least 1 lobular inflammation, and in the Troelstra 2021 study as at least 1 steatosis, at least 1 hepatocyte ballooning and at least 1 lobular inflammation. Abbreviations: CI, confidence interval; FN, false negative; FP, false positive; MRE, magnetic resonance elastography; TN, true negative; TP, true positive

Differences in accuracy estimates between studies of MRE could be attributed to the different cut-off values used (for any fibrosis or significant fibrosis, the cut-off values used in Imajo et al. are the same as those proposed by the company). The Kim et al. (2020) study calculated optimal cut-off values for fibrosis staging from receiver operator characteristic (ROC) curve analysis which were lower than those suggested by Resoundant.

Clinical advice to the EAG was that MRE G' shear modulus reported in Troelstra et al. (2021) is directly comparable to the MRE complex shear modulus used in the other included studies for advanced fibrosis (at least F3). Sensitivity and specificity did not vary systematically with changes in cut-off value, but this may be because of low participant numbers, or because of clinical or methodological differences between the studies. No study used the manufacturer-proposed cut-off for advanced fibrosis of 3.9 kPa.

For MRE, there was only 1 outcome (fibrosis at least F3) where at least 3 studies provided data. As cut-off values varied between the studies, a summary ROC curve was estimated (figure 3).

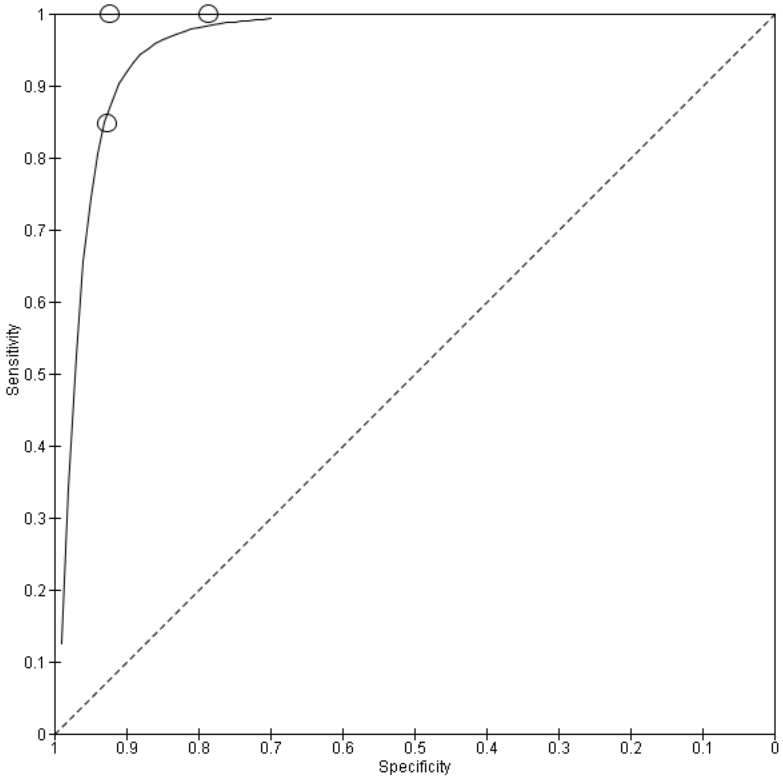

Figure 3 Summary ROC plot for fibrosis (at least F3) data from the MRE test
The solid line is the summary ROC curve. The dashed line indicates sensitivity=1-specificity (that is, an uninformative test). The circles represent individual study results

A more detailed summary of the diagnostic test accuracy results for both interventions is in section 5.3.3 of the diagnostics assessment report.

Clinical impact outcomes

Number of liver biopsies

The RADIcAL1 randomised controlled trial provided evidence for the number of liver biopsies avoided by using LiverMultiScan. Unnecessary biopsies were defined as those done for people who were not subsequently identified by biopsy as having NASH, or as having fibrosis because of conditions other than NAFLD. Biopsy data was available for 55 of the 802 people. A lower proportion of people with non-NAFLD and NAFLD had unnecessary biopsies to diagnose NASH and fibrosis unrelated to NAFLD for the LiverMultiScan arm (n=9 out of 22, 41%) compared with the standard care arm (n=16 out of 31, 52%, EAG calculated odds ratio [OR] 0.65, 95% CI 0.22 to 1.96). Also, compared with the standard care arm (n=13 out of 24, 54%), fewer people with no to mild fibrosis (F0 to F1) in the LiverMultiScan arm had unnecessary biopsies (n=9 out of 22, 41%, EAG calculated OR 0.59, 95% CI 0.18 to 1.89). The proportion of people who had unnecessary biopsies was similar whether or not they had transient elastography before biopsy.

Prognostic ability

One prospective cohort study (Jayaswal et al. 2020) assessed the prognostic ability of LiverMultiScan cT1 to predict clinical outcomes for a population of people indicated for liver biopsy or with a known diagnosis of cirrhosis. However, data was not provided for the subpopulation of people with NAFLD only. Results from LiverMultiScan cT1 predicted event-free survival (defined as survival without occurrence of ascites, variceal bleeding, hepatic encephalopathy, hepatocellular carcinoma, liver transplantation or mortality). The hazard ratio (HR 1.007, 95% CI 1.002 to 1.011, p=0.005) was equivalent to a 0.7% increased risk of a clinical event per 1 ms increase in cT1. When a predefined cut-off of cT1 more than 825 ms was applied, LiverMultiScan predicted event-free survival (p=0.006). All 11 clinical events that were recorded occurred among those who had a cT1 value of more than 825 ms.

A retrospective cohort study (Gidener et al. 2022) reviewed long-term data (at least 10 years) from 1,269 people to assess the ability of MRE results to predict clinical outcomes for people with chronic liver disease who had a single MRE between January 2007 to December 2009. The study population included 375 people with NAFLD. The study reported that people with non-cirrhotic NAFLD at baseline had a lower rate of cirrhosis development (HR 0.37, 95% CI 0.19 to 0.71; p=0.003) than people with other non-cirrhotic liver disease aetiologies, namely hepatitis C, hepatitis B, alcohol-related and primary sclerosing cholangitis. However, no other prognostic data were reported for the subpopulation of people with NAFLD only.

Test failure rate

Three studies reported test failure rate for LiverMultiScan and 6 studies reported test failure rate for MRE. The test failure rate of LiverMultiScan for people with all liver aetiologies ranged from 5.3% to 7.6% and the test failure rate of LiverMultiScan for people with NAFLD only was 5.6%.

The EAG did a fixed-effects meta analysis to obtain a pooled estimate of MRE test failure rate for people with NAFLD, and found a test failure rate of 4.2% (95% CI 2.5% to 6.2%). The reasons for test failure included technical failure, MRI scan cancellation, people being unable to tolerate the scan, refusing the test or experiencing claustrophobia, or being unable to fit in the scanner. For more information see section 5.4.3 and appendix 14 in the diagnostics assessment report.

Patient acceptability of different testing modalities

A study by McKay et al. (2021) collected feedback from people with liver disease (n=90) and from carers (n=11) after people with liver disease had had a LiverMultiScan. Most people in the study considered the MRI scan to be harmless and tolerable and highlighted the importance of the non-invasive nature of the procedure. Some people struggled with breath holding needed, particularly those with lung-related comorbidities.

Some people reported that they hoped that the LiverMultiScan results would mean that they could avoid liver biopsy. They reported that biopsy was very uncomfortable and caused psychological stress. People preferred MRI-based technologies over liver biopsy because they were non-invasive, short in duration and results could be delivered quickly. For more information, see section 5.4.3 of the diagnostics assessment report.

Other measures of clinical impact

The EAG did not identify any relevant studies that provided evidence for the clinical impact of MRI-based technologies for people with NAFLD for the remaining outcomes specified in the scope. Time to diagnosis (defined as time from randomisation to diagnosis by the physician, recorded at the final follow-up visit) was listed as a secondary endpoint in the RADIcAL1 trial protocol. However, no data was available for this at the time of writing of the diagnostics assessment report.

Clinical advice to NICE during scoping was that assessment of liver health by MRI-based technologies can be motivational for people with NAFLD. Therefore, results generated by MRI-based technologies might improve the uptake and maintenance of lifestyle modifications. The EAG identified 1 study (Zelber-Sagi et al. 2017) that assessed the relationships between people with NAFLD and their perceptions of disease consequences and treatment, self-efficacy (belief in ability to organise and do necessary tasks) and healthy lifestyle maintenance. This study did not assess the impact of MRI-based tests. The study reported that self-efficacy and understanding of the illness were factors associated with better nutritional habits. Emotional representation (the extent that people are afraid or concerned about having NAFLD) and perceptions of more severe illness were associated with poorer nutritional habits. No data was available to determine whether LiverMultiScan or MRE specifically affect people's understanding of NAFLD or emotional representation. Also, no data was available on whether LiverMultiScan or MRE impact levels of adherence to lifestyle modifications. More information

on the outcomes of clinical impact review can be found in section 5.4.3 of the diagnostics assessment report.

3 Cost-effectiveness evidence

The EAG did a systematic review to identify any published economic evaluations of MRI-based technologies as diagnostic tools for people with NAFLD who have not had a diagnosis of advanced fibrosis or cirrhosis in the 3 subpopulations specified in the scope. It constructed a de novo economic model to examine the cost effectiveness of 2 diagnostic pathways, LiverMultiScan plus liver biopsy compared with liver biopsy only.

Systematic review of cost-effectiveness evidence

The full inclusion and exclusion criteria for the systematic review can be found in section 6.1.2 of the diagnostics assessment report. One study was identified as being relevant. Although the RADIcAL1 trial protocol states that the cost effectiveness of the introduction of LiverMultiScan is its primary objective, no data was provided.

Overview of included studies

Eddowes et al. (2018) was the only study identified by the systematic review of economic evidence that the EAG considered further. The study used data from 50 people with NAFLD and inconclusive results from fibrosis testing. The EAG noted that all people considered in this analysis were scheduled for a biopsy. This means that the study sample does not represent all people with indeterminate or discordant results from previous fibrosis testing. Clinical advice to the EAG was that not all people with indeterminate or discordant results will have a biopsy. The quality of the study was assessed using the Drummond and CHEERS checklists (see table 10 and appendix 17 of the diagnostics assessment report).

The Eddowes et al. study based their economic analysis on a previous study done by Blake et al. (2016). The analysis used a decision tree to compare the

costs for different diagnostic pathways to distinguish between people at low or high risk for progressive liver disease using non-invasive techniques. This corresponds with the T8 diagnostic strategy described in the economic analysis below. The pathways assessed in Eddowes et al. were LiverMultiScan (2 cut-offs: 822 ms and 875 ms), transient elastography (2 cut-offs: 5.8 kPa and 7.0 kPa), enhanced liver fibrosis (ELF) test (2 cut-offs: 7.7 and 9.8), and transient elastography plus LiverMultiScan (4 combinations of cut-offs), compared with a pathway in which all people had biopsy. The perspective of the analysis was the UK NHS, and the time horizon was 2 weeks (LiverMultiScan and transient elastography were done within 2 weeks of biopsy). Costs were sourced from the NHS tariffs from 2016.

The model generated results for the number of biopsies avoided, total costs, cost saving compared with biopsy use, and total cost per correct diagnosis for a hypothetical cohort of 1,000 people. Health-related quality of life was not considered. All pathways reduced biopsy use and were cost saving compared with an approach of biopsy only for people with suspected NAFLD (the reduction in biopsy use drove cost savings). This assumed that for the comparison of biopsy only, everyone with suspected NAFLD had a biopsy (without having had transient elastography, ELF or MRI tests). The reduction in number of biopsies would be less if this assumption does not reflect clinical practice (as a point of reference, in the RADIcAL1 study, 55 of 802 people with suspected or confirmed liver disease had biopsy). Modelling also did not consider any further assessment done after an initial LiverMultiScan that was not followed by a biopsy (for example a further LiverMultiScan test) or impact of missed diagnosis after the initial testing. The assumed cost of LiverMultiScan used in this study (£143) was also much less than used by the EAG in their modelling (see table 4). This consisted of the cost of an MRI (£148.24) plus the cost of LiverMultiScan data analysis and reporting (£199; £347.24 in total).

Further information on the economic analysis reported in Eddowes et al. can be found in section 6.1 of the diagnostics assessment report.

Economic analysis

The EAG commented that the population modelled by Blake et al. did not have inconclusive results from previous fibrosis testing so the subsequent analyses by Eddowes et al. (which do not specify this population for their economic modelling) are not relevant to this assessment. Also, there was limited data describing the study methods and results so the study quality and generalisability of results were unclear.

The EAG built a decision tree model in Microsoft Excel to estimate the costs and quality-adjusted life-years (QALYs) associated with 2 diagnostic pathways, namely LiverMultiScan plus biopsy and liver biopsy only. The EAG did not model MRE as an intervention in their initial report as there were no diagnostic test accuracy data for MRE specifically in the scope population, but provided analysis in an addendum to the diagnostics assessment report. The modelled population does not cover the full population stated in the scope as diagnostic test accuracy data was not available for people who might have a LiverMultiScan but would not otherwise have a biopsy. Similarly, no data is available for people for whom transient elastography or ARFI are unsuitable.

Perspectum have suggested that LiverMultiScan results can be used by clinicians to help diagnose people with fatty liver, NASH and high-risk NASH. Clinical advice to the EAG was that LiverMultiScan (or MRE) results are unlikely to inform treatment plans as they do not provide the level of detailed information that may be needed to make treatment decisions. For example, identifying clinical features that suggest additional cofactors for liver injury is only available from a biopsy. However, the results could potentially be used to help identify people for whom a biopsy may not be appropriate or needed. So the primary clinical outcome from the EAG model is the number of biopsies avoided if LiverMultiScan were introduced into the diagnostic pathway.

Model structure

In the comparator pathway, people will go straight to biopsy (figure 4). In the EAG's model, a liver biopsy diagnosis is assumed to be 100% accurate.

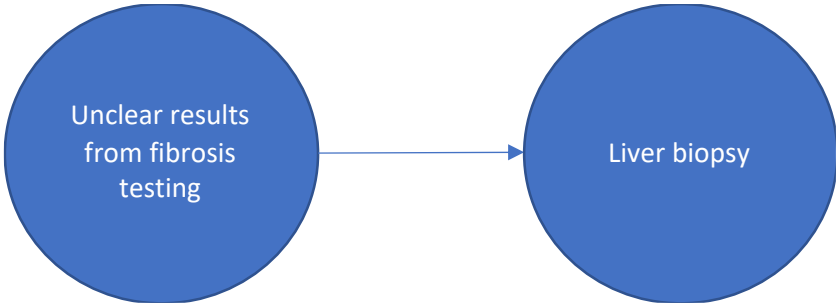

Figure 4 Comparator diagnostic pathway

For the intervention, 8 diagnostic test strategies were considered by the EAG, based on the accuracy of the test (using a cut-off of either 800 or 875 ms; see table 2) to detect the following conditions:

- T1: any fibrosis (at least F1)
- T2: significant fibrosis (at least F2)
- T3: advanced fibrosis (at least F3)
- T4: at least Brunt grade 1
- T5: at least Brunt grade 2
- T6: NASH (NAS at least 4, at least 1 for lobular inflammation and hepatocyte ballooning)
- T7: advanced NASH (NAS at least 4 plus at least F2)
- T8: high risk of progressive disease (NASH or more than F1).

For each of the 8 conditions (T1 to T8), if a person's LiverMultiScan result exceeds the specific cT1 or PDFF threshold associated with the condition, then the person is defined as having a positive result and will have a biopsy to confirm. The EAG's rationale for this was that the diagnostic accuracy data available for use in the model was only available for people who had a biopsy. The EAG stated that there is not sufficient information on the reasons that a person might not be referred for biopsy after a positive LiverMultiScan or MRE test result for them to make informed variations to the economic model on this parameter. In the RADIcAL1 trial, a high-risk LiverMultiScan result was recommended to be followed by further diagnostic evaluation, which could include biopsy. In the EAG's model, test failure always leads to liver biopsy.

For people with a negative LiverMultiScan result, the EAG assumed no biopsy or immediate further assessment is done. In the LiverMultiScan arm of the RADIcAL1 trial, only people who had high risk of NASH based on LiverMultiScan results were recommended to have further diagnostic evaluation. Those with low risk of NASH were recommended to have primary care management and follow-up within 12 months. In the EAG's model a further LiverMultiScan was assumed to be done after 6 months. Perspectum suggested that people will have a second LiverMultiScan after 6 months if their cT1 score is between 800 and 875 ms. However, the EAG assumed that people with cT1 scores less than 800 ms will also have a second LiverMultiScan. The EAG considers that this assumption is appropriate as it considers that all tests for this cohort have low specificity (high rates of false negatives). In an addendum, the EAG provided further analysis in which people who had a cT1 score of under 800 ms did not have a further LiverMultiScan (or subsequent biopsy) after 6 months. A disutility caused by false negative results was only applied for 6 months (a further scenario included no disutility for undiagnosed liver condition).

Further testing at 6 months was assumed to be 100% accurate. If the LiverMultiScan is negative, no further testing was assumed. If positive, a biopsy is done to confirm the result.

The model structure is shown in figure 5.

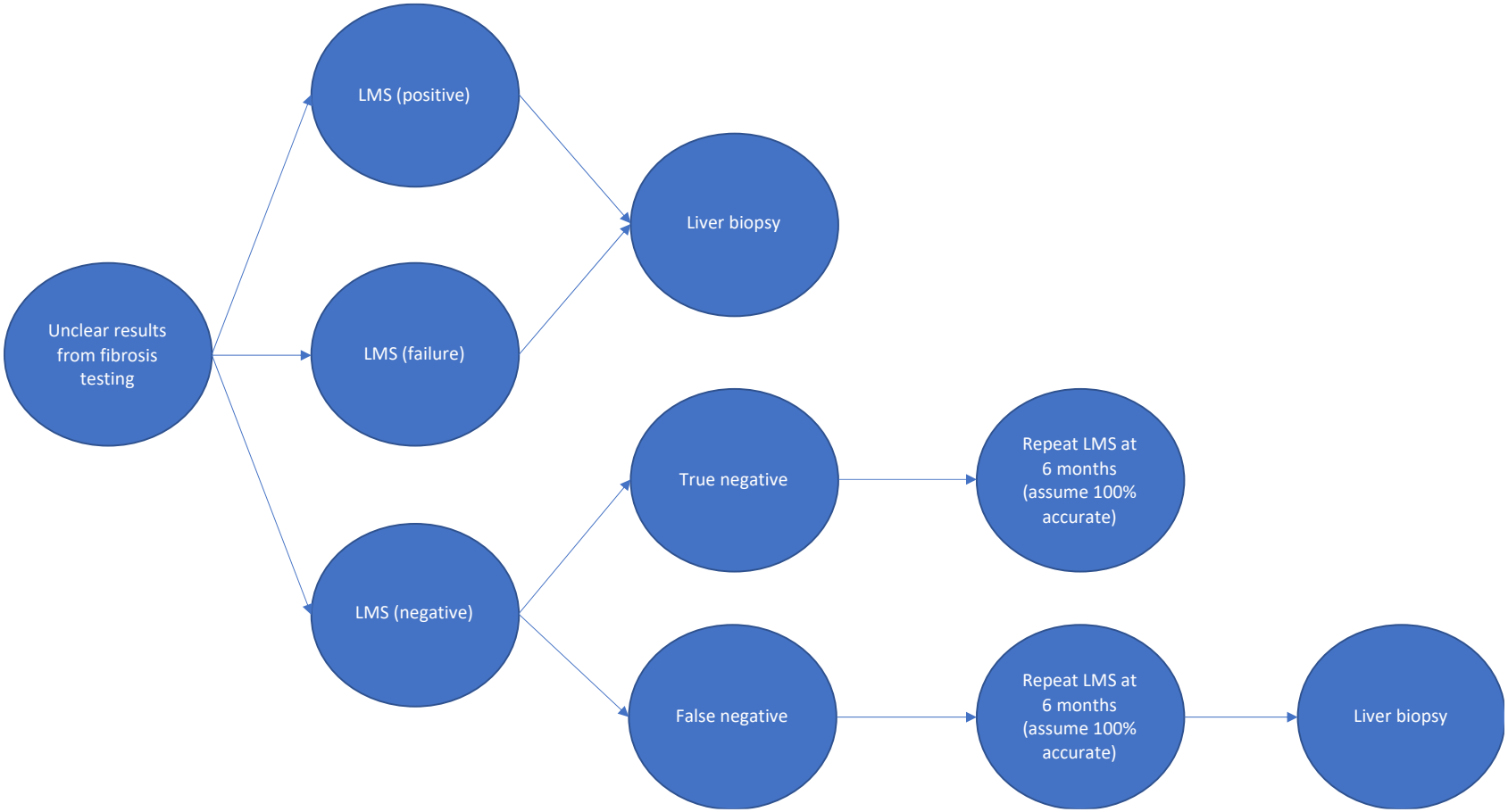

Figure 5 Intervention diagnostic pathway (LiverMultiScan [LMS] plus biopsy)

All people are assumed to have a correct diagnosis by 6 months. Benefits of the LiverMultiScan plus biopsy pathway arise from identifying people with true negative results and removing the costs and lost QALYs arising from unnecessary biopsies. These benefits are balanced against the LiverMultiScan plus biopsy pathway costs and the QALY loss associated with false negative results (that is, people whose liver disease is initially missed, but detected 6 months later). The EAG state that this is an optimistic assumption that favours the LiverMultiScan pathway (see section 6.2.2 in the diagnostics assessment report).

Cut-off values were proposed by Perspectum for the staging of fibro-inflammation, associated diagnoses and clinical management options (see section 1). When compared with diagnostic test accuracy based on PDFF values for the same test strategies from the same cohort of people (Eddowes et al. 2018), the cT1 scores generated the same or higher sensitivity and specificity except for sensitivity for steatosis of at least Brunt grade 2 (see section 2).

The EAG also modelled the MRE test. It used the same model structure and assumptions as for the LiverMultiScan model, including that all people with a negative result from a MRE are recalled at 6 months for a second MRE, at which point a correct diagnosis is made. Positive MRE results were also assumed to be confirmed by subsequent biopsy.

Population

The modelled population was people with inconclusive results from fibrosis testing who, without access to LiverMultiScan or MRE, would be scheduled for and would have a biopsy. Clinical advice to the EAG was that people with indeterminate results from previous fibrosis testing would likely be referred for biopsy unless there are clear reasons not to. The EAG stated that there is no population prevalence or test accuracy data for people with indeterminate results from previous fibrosis testing who would not be sent for a biopsy.

Therefore, this population was not considered in the model (see section 6.2.3 in the diagnostics assessment report).

Model parameters

Only costs and outcomes associated with LiverMultiScan or MRE and biopsy within the 6-month time horizon of the model were considered. No costs associated with previous testing or long-term outcomes were included.

Diagnostic test accuracy

Diagnostic test accuracy values and test failure rates for LiverMultiScan were estimated using values from Eddowes et al. (2018; table 2). Accuracy estimates for T8 were updated from those reported in the diagnostics assessment report in an addendum. This is because, after the report was submitted, the company indicated that values reported in Eddowes et al. used a previous version of the technology. Updated accuracy estimates from this study using the current version of the technology were provided and are in table 2.

Table 2 LiverMultiScan cT1 diagnostic test accuracy

Diagnostic test strategy	cT1 cut-off value	Population prevalence	Sensitivity	Specificity
T1: any fibrosis (≥F1)	800 ms	87.0%	0.88	0.67
T2: significant fibrosis (≥F2)	875 ms	65.2%	0.63	0.75
T3: advanced fibrosis (≥F3)	875 ms	47.8%	0.64	0.63
T4: Brunt grade ≥1	800 ms	97.8%	0.80	0.00
T5: Brunt grade ≥2	875 ms	50.0%	0.70	0.70
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875 ms	54.4%	0.64	0.67
T7: advanced NASH (NAS≥4 plus ≥F2)	875 ms	47.8%	0.64	0.62
T8: high risk of progressive disease (NASH or >F1)	875 ms	82.6%	0.58	0.88

Abbreviations: ms, milliseconds; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis

For MRE, accuracy estimates from Imajo et al. (2021) were used (table 3). The EAG commented that this population does not exactly match that specified in the scope (data for this population was not available). The proportion of failed MRE tests was assumed to be the same as LiverMultiScan.

Table 3 MRE diagnostic test accuracy

Diagnostic test strategy	Cut-off value	Population prevalence	Sensitivity	Specificity
T1: any fibrosis (≥F1)	2.9 kPa	87%	0.79	1.00
T2: significant fibrosis (≥F2)	3.3 kPa	65%	0.82	0.83
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3 kPa	54%	0.71	0.41
T7: advanced NASH (NAS≥4 plus ≥F2)	3.5 kPa	48%	0.69	0.50

Abbreviations: NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; kPa, kilopascals

Costs

Direct costs for biopsy were taken from NHS reference costs for 2019 to 2020 for transvascular and percutaneous liver biopsy, weighted according to usage. Biopsy complications were costed according to a study by Stevenson et al. (2012), weighted according to type of biopsy and inflated to 2019 to 2020 prices using the NHS cost inflation index. Costs for LiverMultiScan were based on the NHS reference cost for a single non-contrast MRI scan, added to the cost per scan for data analysis and reporting charged by Perspectum (table 4).

For MRE, the company indicated that the approximate cost of adding MRE to an existing MRI machine would be about £35,000. However, some MRI scanners in UK may already have MRE functionality. So, the EAG estimated 2 costs for MRE per scan. The first assumed the MRI scanner already has MRE capabilities (the cost of MRE is the same as the cost of MRI alone; £148.24). The second assumed that MRE would have to be installed onto the MRI

device (the cost of MRE is the cost of MRI alone plus an additional installation cost; £148.24 plus £59.50; table 4). To calculate the installation cost per scan done, the EAG had to make assumptions about the number of people per year having a scan, the number of MRI machines in the UK that would have MRE installed and the average lifespan of MRI machines. No additional cost for maintenance of the MRE is included. The EAG cautioned that this cost is built on several assumptions, some of which are not evidenced. Full details can be found in the diagnostics assessment report addendum.

Table 4 Costs used in EAG's economic model

Item	Cost	Description	Source
Transjugular biopsy	£1,513	YG10Z Percutaneous transvascular biopsy of lesion of liver	NHS Reference Costs 2019/20
Standard biopsy	£770	YG11A Percutaneous punch biopsy of lesion of liver, 19 years and over	NHS Reference Costs 2019/20
Average biopsy cost	£805	Weighted average of YG10Z and YG11A	NHS Reference Costs 2019/20
Treating biopsy complication	£8.54	Weighted average of costs for treating percutaneous and transjugular biopsy complications	Stevenson et al. (2012), inflated to 2019/20 using NHS cost inflation index
MRI	£148.24	RD01A Scan of one area, without contrast, 19 years and over	NHS Reference Costs 2019/20
LiverMultiScan	£199	Cost per scan for data analysis and reporting	Company
MRE (additional cost assuming MRE would have to be installed)	£59.50	Calculated cost per scan assuming MRE needs installing onto MRI scanner	Company and assumptions made by EAG

Abbreviations: MRE, magnetic resonance elastography; EAG, external assessment group

The cost of MRI used in the model does not reflect any change in infrastructure that would be needed if MRI scans were introduced in this population. The EAG commented that if MRI-based technologies were to be recommended by NICE, the implications for NHS service provision would be

significant. This would be because of the increased staffing levels and changes in infrastructure needed to accommodate the high demand for MRI scans for people with NAFLD.

Utility values

Utility values used in the model are the disutilities associated with having a biopsy, and the disutility of undiagnosed liver disease accrued during the 6-month period before a person with an initial false negative from an MRI test is assumed to get a correct diagnosis. Disutilities associated with biopsy may be because of direct pain and anxiety, serious adverse events, or death. Utility values for these conditions were obtained from Stevenson et al. (2012) and weighted according to the proportion of people having each type of biopsy in the NHS.

Disutility from undiagnosed liver disease was taken from the QALY loss associated with untreated NASH used in the NICE guideline for the assessment and management of NAFLD. The EAG commented that if people do not have symptoms during the 6 months before having the second test, the QALY loss should be interpreted as a loss in QALYs because of a delayed diagnosis. A delayed diagnosis means that the disease is more advanced at the time of diagnosis, which could mean reduced treatment options, more severe symptoms and potentially reduced life expectancy. In scenario analysis, any QALY loss from having a false negative result was removed from the model, which the EAG considered implausibly favourable to the MRI tests.

Table 5 Utility values

Source of QALY loss	Value of QALY loss	Source or justification
Liver biopsy: direct pain and anxiety	0.00453	Assumption based on clinical advice and EQ-5D-3L scoring
Live biopsy: serious adverse events	0.000147	Stevenson et al. (2012)
Liver biopsy: death	0.00141	Assumption based on risk of death from biopsy
Other: failure to treat advanced liver disease	0.03 per year	QALY loss from untreated NASH from NG49

Abbreviations: NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

Other parameters

The EAG did a scenario analysis in an addendum to the main report. In this analysis, people who had LiverMultiScan cT1 results of less than 800 ms did not have any further assessment. The EAG used data from Eddowes et al. to calculate the proportion of people with cT1 results under 800 ms (39.1%).

A full description of the assumptions and parameter sources used in the EAG's base case model can be found in sections 6.2.7 to 6.2.11 in the diagnostics assessment report.

Model results

LiverMultiScan

Base case

The EAG generated base case analysis cost-effectiveness results for a hypothetical cohort of 1,000 people according to the 8 test strategies previously described (table 6). For a full description of the EAG's base case, please see section 6.2.13 of the diagnostics assessment report.

There was wide variation in the number of biopsies avoided per 1,000 people between the 8 test strategies (minimum: T4 [n=0]; maximum: T5 [n=328.9]). For all strategies, introducing LiverMultiScan increased cost per person. While

biopsy procedure and complication costs were lower, the additional cost of doing the LiverMultiScan (between £411,556 and £511,311 per 1,000 people) was much higher. Full cost breakdown can be found in table 17 of the diagnostics assessment report and addendum. For all but 1 of the strategies (T5), QALY losses were greater for the LiverMultiScan pathway than the biopsy only pathway. QALY losses related to biopsy were lower for the LiverMultiScan arm, but the QALY loss associated with false negative LiverMultiScan results (that is, people who have undetected liver disease until further scans at 6 months) overcame this in most cases. For the full QALY breakdown see table 18 of the diagnostics assessment report and addendum.

Table 6 EAG base case analysis results for LiverMultiScan (per 1,000 people)

Diagnostic test strategy	cT1 cut-off value (ms)	Biopsies averted	Incremental cost	Incremental QALYs	ICER (£ per QALY gained)
T1: any fibrosis (≥F1)	800	82.2	£344,671	-0.96	LiverMultiScan dominated
T2: significant fibrosis (≥F2)	875	246.6	£310,655	-1.63	LiverMultiScan dominated
T3: advanced fibrosis (≥F3)	875	308.2	£260,617	-0.27	LiverMultiScan dominated
T4: steatosis (Brunt grade ≥1)	800	0.0	£411,556	-2.78	LiverMultiScan dominated
T5: steatosis (Brunt grade ≥2)	875	328.9	£243,770	0.19	£1,266,511
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	287.3	£277,597	-0.73	LiverMultiScan dominated
T7: advanced NASH (NAS≥4 and ≥F2)	875	308.1	£260,684	-0.27	LiverMultiScan dominated
T8: high risk of progressive disease (NASH or >F1)	875	143.8	£394,320	-3.90	LiverMultiScan dominated

Abbreviations: cT1, corrected T1; ICER, incremental cost-effectiveness ratio; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

The proportion of unnecessary biopsies decreased if LiverMultiScan was used (table 7).

Table 7 Proportion of unnecessary biopsies in the base case analysis for LiverMultiScan

Diagnostic test strategy	Proportion of unnecessary biopsies without LiverMultiScan	Proportion of unnecessary biopsies with LiverMultiScan	Change in proportion of unnecessary biopsies
T1: any fibrosis (≥F1)	13.0%	5.2%	-7.8%
T2: significant fibrosis (≥F2)	34.8%	13.5%	-21.3%
T3: advanced fibrosis (≥F3)	52.2%	30.9%	-21.3%
T4: steatosis (Brunt grade ≥1)	2.2%	2.2%	–
T5: steatosis (Brunt grade ≥2)	50.0%	25.5%	-24.5%
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	45.6%	23.7%	-21.9%
T7: advanced NASH (NAS≥4 and ≥F2)	52.2%	30.9%	-21.3%
T8: high risk of progressive disease (NASH or >F1)	17.4%	3.7%	-13.7%

Abbreviations: NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis

The EAG commented that cost-effectiveness results for the LiverMultiScan plus biopsy pathway are optimistic as they were generated using the assumption that people will get a correct diagnosis after a maximum of 2 LiverMultiScan tests.

The impact of uncertainty of the parameter values on the outcome of the model was investigated by the EAG in threshold and scenario analyses. For more detail, see sections 6.2.14 to 6.2.16 in the diagnostics assessment report.

Threshold analyses

Population prevalence

If the LiverMultiScan test was 100% accurate, ICERs would only fall below £20,000 per QALY gained if the prevalence of the condition being tested for was less than 40%. If the acceptability threshold was £30,000 per QALY gained, prevalence would need to be less than 46%. The EAG explained that this was because at lower prevalence use of LiverMultiScan increased the number of unnecessary biopsies avoided. In the Eddowes et al. dataset used in the model, the test strategy condition with the lowest prevalence was advanced NASH (NAS at least 4 and at least F2, 47.8%). However, the accuracy of LiverMultiScan to detect this condition was not close to 100% (sensitivity 0.64 and specificity 0.62; see figure 1).

The EAG noted that some studies included in the diagnostic test accuracy review for LiverMultiScan or MRE reported significantly different prevalence of the test strategy diagnoses (for example, significant fibrosis had a prevalence ranging from 43.6% in Kim et al. 2020 to 75% in Imajo et al. 2021). The EAG considered Eddowes et al. was the most appropriate study to provide estimates of condition prevalence for this assessment because of explicitly including people with inconclusive results from previous testing. However, it noted that disparity between the estimates in these studies highlighted that there may be uncertainty about the population prevalence. Also, other studies carried out in the same population may lead to substantially different population prevalence estimates.

In an addendum to the main diagnostic assessment report, the EAG provided a threshold analysis using the test accuracies used in the base case (rather than assuming 100% accuracy, as above). Analyses of population prevalence based on a maximum acceptable ICER of £20,000 and £30,000 per QALY gained are shown in table 8. As noted above, lower prevalence increases the numbers of biopsies avoided (compared with the base case in table 6) which is a major driver of cost effectiveness.

Table 8 Threshold analysis of prevalence: cost effectiveness assuming base case LiverMultiScan test accuracy

Diagnostic test strategy	Prevalence cost effective at maximum acceptable ICER of £20,000 per QALY	Number of biopsies averted (at prevalence in preceding column)	Prevalence cost effective at maximum acceptable ICER of £30,000 per QALY	Number of biopsies averted (at prevalence in preceding column)
T1: any fibrosis (≥F1)	8%	582	16%	531
T2: significant fibrosis (≥F2)	13%	617	18%	581
T3: advanced fibrosis (≥F3)	2%	579	8%	543
T4: steatosis (Brunt grade ≥1)	Never	–	Never	–
T5: steatosis (Brunt grade ≥2)	9%	599	15%	559
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	6%	592	12%	554
T7: advanced NASH (NAS≥4 and ≥F2)	2%	578	8%	543
T8: high risk of progressive disease (NASH or >F1)	21%	657	24%	625

Abbreviations: ICER, incremental cost-effectiveness ratio; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

QALY loss associated with biopsy

Use of LiverMultiScan reduced the number of biopsies, reducing the QALY loss caused by this procedure. The EAG also examined the extent of QALY loss because of biopsy that would be needed for the most cost-effective strategy (T5) to become cost effective at a threshold of £20,000 or £30,000

per QALY (table 9). They found that the QALY loss would have to increase between 340% and 514% for cost effectiveness at these thresholds to be achieved.

Table 9 Results of threshold analyses on QALY loss associated with biopsy for LiverMultiScan

Diagnostic test strategy	Original QALY loss	Threshold QALY loss	Increase from original
Brunt Grade ≥2 threshold: £20,000 per QALY	0.007	0.044	514%
Brunt Grade ≥2 threshold: £30,000 per QALY	0.007	0.031	340%

Abbreviation: QALY, quality-adjusted life year

Scenario analyses

Test failure rate

In the base case analysis, the failure rate of LiverMultiScan was set to 5.5% based on the reported rate of failure in Eddowes et al. The EAG examined scenarios in which the failure rate was set to 0% or 10%. This did not have a significant effect on the ICERs, with most strategies remaining dominated by the biopsy only pathway.

Removal of QALY loss associated with a delayed diagnosis

The EAG assessed a scenario in which there were no QALY losses associated with a delayed diagnosis of liver disease if missed by the initial LiverMultiScan (but detected at 6 months). It commented that this provides information on the importance of this QALY loss to overall cost-effectiveness results. But, the EAG noted this is not plausible because it would indicate no impact of a correct and an incorrect diagnosis, and consequently no point to testing. Even under this potentially unrealistic assumption, cost-effectiveness estimates remained over £100,000 per QALY (table 10). A threshold analysis based on prevalence of conditions needed for the test to be cost effective was also done by the EAG. Results can be found in the addendum to the diagnostics assessment report.

Table 10 Results of scenario analyses of LiverMultiScan with no QALY loss assumed for delayed diagnosis caused by initial false negative test result

Diagnostic test strategy	Scenario with no QALY loss for delayed diagnosis ICER (£ per QALY gained)
T1: any fibrosis (≥F1)	£587,405
T2: significant fibrosis (≥F2)	£176,491
T3: advanced fibrosis (≥F3)	£118,501
T4: steatosis (Brunt grade ≥1)	LiverMultiScan dominated
T5: steatosis (Brunt grade ≥2)	£103,861
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	£135,392
T7: advanced NASH (NAS≥4 and ≥F2)	£118,563
T8: high risk of progressive disease (NASH or >F1)	£384,204

Abbreviations: ICER, incremental cost-effectiveness ratio; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

No use of a second LiverMultiScan for people with a result under 800 ms

In the base case for LiverMultiScan, everyone with a negative result from an initial LiverMultiScan test had a second LiverMultiScan after 6 months. In a scenario analysis, this was removed for people with a result of less than 800 ms (table 11). Even when no disutility was included for false negatives (as above, which favours the cost effectiveness of LiverMultiScan), the ICER for the comparison of testing with LiverMultiScan compared with no LiverMultiScan remained above £48,000 per QALY gained. In some scenarios LiverMultiScan is no longer dominated when further LiverMultiScan testing at 6 months is removed because this also removes subsequent biopsies and associated disutility.

Table 11 Scenario analyses with people with cT1 less than 800 ms not sent for second LiverMultiScan at 6 months

Diagnostic test strategy	Base case result	Scenario 1: Patients with cT1<800ms not sent for second LiverMultiScan	Scenario 2: Same as scenario 1 with no QALY loss for false negative results
T1: any fibrosis (≥F1)	LiverMultiScan dominated	LiverMultiScan dominated	£587,405
T2: significant fibrosis (≥F2)	LiverMultiScan dominated	LiverMultiScan dominated	£73,054
T3: advanced fibrosis (≥F3)	LiverMultiScan dominated	£748,291	£54,248
T4: steatosis (Brunt grade ≥1)	LiverMultiScan dominated	LiverMultiScan dominated	LiverMultiScan dominated
T5: steatosis (Brunt grade ≥2)	£1,266,511	£225,729	£48,738
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	LiverMultiScan dominated	LiverMultiScan dominated	£60,194
T7: advanced NASH (NAS≥4 and ≥F2)	LiverMultiScan dominated	£754,729	£54,271
T8: high risk of progressive disease (NASH or >F1)	LiverMultiScan dominated	LiverMultiScan dominated	£116,081

Abbreviations: NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

Other uncertainty analyses

The EAG did not do probabilistic sensitivity analyses, stating that the model is a single-node decision tree so is linear. It stated that probabilistic sensitivity

analysis is not needed to explore the impact of non-linearity on cost-effectiveness results. But it provided further analysis of the model in an addendum to show that any impact of non-linearity would not be important for decision making. The EAG highlighted that the distributions around most of the model inputs are unknown. It also considered that deterministic one-way sensitivity analyses would not add further value. See section 6.2.16 in the diagnostics assessment report and the addendum to the diagnostics report for further details.

The EAG stated that extending the time horizon of the model beyond 6 months would further reduce the cost effectiveness of the LiverMultiScan pathway. This is because of increased QALY losses associated with missed diagnoses, and increased costs associated with further diagnostic tests.

MRE

Base case

The EAG provided cost-effectiveness estimates for the MRE test in an addendum to the main report. Because of uncertainties about the cost of testing (see costs), results were provided assuming no additional cost of MRE over an MRI scan (£148 per test; table 12) and with a further cost applied (£208 per test; table 13). For detecting significant fibrosis (at least F2), MRE has an ICER of about £230,000 per QALY gained if an additional cost for MRE (over the cost of doing an MRI) is included but dominates if this cost is not included. Full results can be found in the addendum to the diagnostics assessment report.

Table 12 EAC base case analysis results for MRE with cost of test set to MRI alone (per 1,000 people)

Diagnostic test strategy	Cut-off value (kPa)	Biopsies averted with MRE	Additional cost for MRE pathway	Incremental QALYs	ICER (£ per QALY gained)
T1: any fibrosis (≥F1)	2.9 kPa	122.9	£92,102	-1.71	MRE dominated
T2: significant fibrosis (≥F2)	3.3 kPa	273.0	-£16,916	0.28	MRE dominates
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.0 kPa	176.7	£52,797	-0.98	MRE dominated
T7: Advanced NASH (NAS≥4 and ≥F2)	3.5 kPa	246.6	£4,905	-0.34	MRE dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; NAS, NAFLD activity score; kPa, kilopascals; MRE, magnetic resonance elastography; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year

Threshold analyses

The EAG also did threshold analyses for MRE to determine at what prevalence and cost testing became cost effective. Full results can be found in the diagnostics assessment report addendum.

Disease prevalence

For detecting significant fibrosis (at least F2; T2), assuming an additional cost for MRE on top of the cost of doing an MRI scan (base case ICER of £229,967 per QALY gained; table 13), prevalence would need to decrease to 56% for the test to be cost effective at a maximum acceptable ICER of £20,000 per QALY. In the base case prevalence was 65%, based on

Eddowes et al. Prevalence of significant fibrosis in Imajo et al. (2021), the study used to provide accuracy estimates for MRE to detect fibrosis, was 76%.

Table 13 EAC base case analysis results for MRE with a cost of £59.50 on top of MRI cost (per 1,000 people)

Diagnostic test strategy	Cut-off value (kPa)	Biopsies averted with MRE	Additional cost for MRE pathway	Incremental QALYs	ICER (£ per QALY gained)
T1: any fibrosis (≥F1)	2.9kPa	122.9	£169,184	-1.71	MRE dominated
T2: significant fibrosis (≥F2)	3.3kPa	273.0	£65,424	0.28	£229,967
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.0kPa	176.7	£131,679	-0.98	MRE dominated
T7: advanced NASH (NAS≥4 and ≥F2)	3.5kPa	246.6	£87,412	-0.34	MRE dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; QALY, quality-adjusted life year.

Cost of MRE

For detecting significant fibrosis (at least F2; T2), MRE testing became cost effective at a maximum acceptable ICER of £20,000 and £30,000 per QALY at a cost per test of about £165. That is, the cost of doing an MRI scan (£148) plus an additional cost of £17 per scan for MRE.

For the other strategies, including any fibrosis (at least F1), the cost of doing MRE would need to be lower than the cost of doing an MRI scan.

Scenario analyses

Removal of QALY loss associated with a delayed diagnosis

As for LiverMultiScan, the EAG assessed a scenario in which there were no QALY losses associated with a delayed diagnosis of liver disease if missed by the MRE at initial scan (but detected at 6 months). When the cost of installing MRE was included in costs as well as doing an MRI scan alone, MRE testing no longer resulted in a loss of QALYs in any strategy, but all ICERs were over £30,000 per QALY gained. For full results, see the addendum to the diagnostics assessment report.

4 Summary

Clinical effectiveness

The EAG considered that 1 study (Eddowes et al. 2018) provided evidence for test accuracy of LiverMultiScan in people with discordant or indeterminate results from previous fibrosis testing. The diagnostic accuracy of cT1 and MRI PDFF outputs for a variety of NAFLD-related conditions was evaluated. For diagnosis of NASH and advanced NASH, cT1 had 64% sensitivity, and specificities of 67% for NASH and 63% for advanced NASH. Diagnostic accuracy values from Eddowes et al. (2018) were generally consistent with other studies of LiverMultiScan in broader populations, except in cases where low patient numbers resulted in extreme estimates of sensitivity or specificity (that is, where there were 0 people in the true negative or false positive categories). No studies were identified that provided evidence for the diagnostic test accuracy of MRI-based technologies for people with NAFLD for whom transient elastography or ARFI were unsuitable.

Diagnostic accuracy data for MRE was available from 4 studies in people with NAFLD who had not had a diagnosis of advanced fibrosis or cirrhosis. Although diagnostic definitions were consistent between these studies, MRE output cut-off values were not. Only 1 study (Imajo et al. 2021) used the cut-off values proposed by the company for any fibrosis (at least F1) or significant

fibrosis (at least F2). No study used the 3.9 kPa cut-off proposed for advanced fibrosis (at least F3). The sensitivity and specificity for advanced fibrosis observed in the 2 studies that used the Resoundant MRE platform that is commercially available ranged from 85% to 100% and from 92% to 93%, respectively. No studies of MRE were identified that specified whether the population had previous testing, or for whom transient elastography or ARFI were unsuitable.

There was very little data on how the tests influence decisions about care. Results from the RADIcAL1 randomised controlled trial of LiverMultiScan found that LiverMultiScan could reduce the number of unnecessary biopsies for people with non-NAFLD, NAFLD and no to mild fibrosis (F0 to F1) when compared with standard care but results were not statistically significant (EAG calculated OR 0.65, 95% CI 0.22 to 1.96). Numbers of biopsies done in study were low (55 of the 802 enrolled had a biopsy).

Acceptability of LiverMultiScan from patient feedback was generally positive.

No data was available to assess the impact of receiving MRI results on a person's motivation to modify their lifestyle, or to adhere to lifestyle advice and treatment.

Cost effectiveness

LiverMultiScan

Based on the EAG's modelling, LiverMultiScan would identify people for whom a biopsy is not necessary and reduce the proportion of people who have an unnecessary biopsy (up to about 330 biopsies averted per 1,000 people who would otherwise all have a biopsy). However, this reduction comes with a large increase in additional MRI scans and associated cost (between about £412,000 to £511,000 per 1,000 people). The cost of this may not have been fully captured in the EAG's cost estimates.

In the EAG's base case LiverMultiScan was dominated by biopsy only for most diagnostic strategies.

If QALY loss associated with delayed diagnosis of liver conditions caused by a false negative initial LiverMultiScan test was removed from the model, cost-effectiveness estimates remained over £100,000 per QALY.

Removal of subsequent LiverMultiScan for people with an initial result of under 800 ms from the base case improved cost effectiveness, but ICERs were still over £225,000 per QALY, or testing with LiverMultiScan remained dominated. Even if QALY loss associated with delayed diagnosis was removed from the model, ICERs were still over £48,000 per QALY gained.

If the accuracy of the test is assumed to be 100%, LiverMultiScan could become cost effective if the prevalence of the condition being tested for was less than 40% to 45%. Lower prevalence improves cost effectiveness because this increases the number of unnecessary biopsies done in the comparator (where everyone is assumed to get a biopsy) and the number of unnecessary biopsies that can be avoided by using MRI-based tests (a major driver of cost effectiveness). Using accuracy estimates for LiverMultiScan from the EAG's base case, prevalence would need to be much lower for the test to be cost effective. For example, for advanced NASH, prevalence would need to be 8% or lower for the test to be cost effective at £30,000 per QALY. Prevalence in Eddowes et al. was 48% and the lowest prevalence for advanced NASH identified in studies included in the EAG's systematic review was 36% (see figure 1). These analyses assume that if LiverMultiScan was not available everyone in this population would still have a biopsy.

MRE

If no additional cost for MRE was assumed (over the cost of doing an MRI scan), detection of significant fibrosis (at least F2) dominated standard care. If the EAG's calculated additional cost (over an MRI scan; an additional £59.50) was included, the ICER was about £230,000 per QALY gained. MRE was cost

effective using a maximum acceptable ICER of £20,000 and £30,000 per QALY if this additional cost for MRE was about £17 or less.

Lower prevalence made MRE testing more cost effective. For detection of significant fibrosis (at least F2), and including the EAG's calculated additional cost to doing an MRI for MRE, prevalence would need to be about 10 percentage points lower than the base case (65%) for the test to be cost effective at a maximum acceptable ICER of £20,000 per QALY.

5 Issues for consideration

Clinical effectiveness

The data identified by the EAG was largely on the accuracy of the tests. The EAG highlighted that little of this data was specifically in the population set out in the scope. There was very little evidence on how test results would impact decisions about care. At scoping, uncertainty about the impact of a diagnosis of NASH from the LiverMultiScan on care was highlighted, in terms of how this would change care for people with NAFLD. The only data on the impact of LiverMultiScan on biopsy use came from the RADIcAL1 trial. However, the number of people who had a biopsy in this trial was low. In the discussion section of the report on this study provided by the company, the authors state that this is likely because there are no current treatment options for NASH. So unless the clinician suspects advanced fibrosis, clinical management will be the same for NAFLD or NASH. They further stated that this will change if NASH therapeutics become available. Because there was uncertainty about how the outcome of LiverMultiScan informed the decision to do a biopsy in this study, the EAG stated that the clinical value of LiverMultiScan in avoiding biopsies remains uncertain.

There was no data identified on how MRE results impact on care, for example use of biopsy.

Only 1 study was identified that provided accuracy estimates for MRE using thresholds stated by the manufacturer (for fibrosis; Imajo et al. 2021).

Cost effectiveness

The cost of the MRI tests used in the model does not include any changes in infrastructure needed if MRI scans were introduced in this population. Introducing MRI scans into a population in which they do not routinely already occur has the potential to have a large impact on the NHS.

In the base case, it was assumed that all people entering the model would otherwise be referred for a biopsy. The EAG stated that there is not sufficient information to allow them to model people who, in current care, would not have a biopsy. Using MRI-based tests for this group would not reduce biopsy use (as none would be done in current care) but could potentially detect liver disease earlier. No data on this group was identified.

The EAG assumed that all positive results from the MRI tests would need to be confirmed by a biopsy. As noted above, very limited data was identified to show how test results impact on decisions about care.

There is uncertainty about the impact of a delayed diagnosis of liver disease (such as fibrosis or NASH) on health-related quality of life.

LiverMultiScan

Using LiverMultiScan resulted in a QALY loss in the base case compared with standard care in most of the strategies modelled by the EAG, and all strategies based on accuracy to detect NASH. This was caused by disutility from false negative results from LiverMultiScan for 6 months (after which the model assumes further testing correctly identifies liver disease) generally outweighing QALY gains from reductions in biopsy use. The EAG used a figure from the NICE guideline on NAFLD on the QALY loss associated with untreated NASH for all undetected liver conditions. If this is lower in reality, testing may not reduce health-related quality of life compared with standard

care. The EAG ran analyses in which this QALY loss was removed entirely, and the most cost-effective strategy had an ICER of £103,861 per QALY.

Reduction of biopsy use was a major driver of cost effectiveness. The only data on the impact of LiverMultiScan on biopsy use included in the systematic review was from the RADIcAL1 randomised controlled trial, which enrolled people with suspected NAFLD. Biopsy use in this study was low, with use lower (but not statistically significantly so) in the LiverMultiScan arm (22 out of 403 people; 5.5%) than control arm (31 out of 399 people; 7.8%); an approximately 30% decrease. A similar decrease in biopsy use was seen in the EAG's base case. For example, for T7, 691.9 biopsies were estimated to be done with LiverMultiScan, compared with 1,000 for standard care (see table 6). A lower proportion of people had unnecessary biopsies with LiverMultiScan compared with standard care in RADIcAL1 (41% compared with 52%). Again, this was similar to results from the EAG's model; for example, for T7 (30.9% compared with 52.2%; table 7).

Cost-effectiveness estimates were sensitive to the underlying prevalence of the conditions being tested for. LiverMultiScan had ICERs lower than £20,000 or £30,000 per QALY when condition prevalence was set to lower values (see the threshold analyses). The population prevalence values used in the base case model were calculated using data from the 46 people in the Eddowes et al. (2018) study (see table 2). The EAG noted that this is a small population and that the true population prevalence in NHS clinical practice is uncertain. All cost-effectiveness estimates using low condition prevalence assume that current practice (without LiverMultiScan use) is to offer everyone in this population a biopsy, even though fewer people are expected to have positive results from biopsy. Any widening of the population having testing is likely to have a greater infrastructure-related impact on the NHS.

MRE

Testing for advanced fibrosis using MRE was cost effective in the EAG's model when using MRE added no additional cost to doing an MRI scan. There

was considerable uncertainty about the extent of any additional cost incurred through MRE use and the EAG's cost estimates may have underestimated costs of increased MRI use in the NHS. Based on the EAG's threshold analysis, MRE testing for advanced fibrosis would not be cost effective if it cost more than about £17 in addition to the cost of doing an MRI. The EAG's assessment of this likely additional cost was £59.50.

Accuracy estimates for MRE using the company's specified thresholds were only available from 1 study (Imajo et al. 2021; n=144). This study had inclusion criteria broader than the scope population and EAG rated it unclear in terms of applicability for patient selection.

6 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

People with a South Asian family background may have a more centralised distribution of body fat, leading to a higher risk of associated chronic diseases such as NAFLD or NASH (De Silva et al. 2018; British Liver Trust, 2018). Criteria for suspected NAFLD or NASH may be different in the South Asian population than in the wider population. According to the British Liver Trust, a BMI above 23 kg per m^2 is considered to increase the risk of NAFLD in people with a South Asian family background. Similarly, a recommendation was made to reduce the healthy waist circumference range for men in this population, from 94 cm to 90 cm (British Liver Trust, 2018).

One of the major risk factors for NAFLD is obesity. Transient elastography or ARFI may not work in people with obesity because of fat or fluid overlying the liver. MRI techniques may be beneficial for people with obesity if they enable non-invasive characterisation of fibrosis where other techniques may not work. However, MRI techniques may not be suitable for people with a very high BMI because there are weight or size limits for some scanners.

7 Implementation

The EAG stated that implementation of MRI technologies for assessing NAFLD would have significant implications for the NHS. This would be because of increased staffing levels and changes in infrastructure needed to accommodate the increased demand for MRI scans.

The COVID-19 pandemic has caused increased wait times for many in-hospital tests such as MRI and ultrasound (NHS England, March 2022), Access to these services may still be restricted. Increasing the number of people referred to MRI for liver imaging may further increase wait times.

8 Authors

Jacob Grant
Topic lead

Thomas Walker
Technical adviser

9 Glossary

Acoustic radiation force impulse (ARFI) is an ultrasound technique that applies a shear wave laterally to the usual ultrasound pulse. It is used to measure liver stiffness.

The **Brunt scoring system** is a component of the NAS (see below) used to assess the level of steatosis (fat) in a liver biopsy sample. The score ranges from 0 (less than 5% fat) to 3 (more than 66% fat).

Cirrhosis is severe scarring of the liver, preventing normal liver function.

An intervention or procedure is considered **dominated** if it is both more costly and less effective than the comparator.

Iron corrected T1 (cT1) is an MRI technique that is part of the LiverMultiScan package. It uses a T2* scan to correct for hepatic iron on T1 scans.

Fibrosis is accumulation of scar tissue and may be caused by a variety of conditions.

Hepatocyte ballooning is when cells in the liver swell and enlarge, and is found particularly in steatohepatitis.

Lobular inflammation is a histological feature of liver disease in which aggregates of immune cells are observed.

Non-alcoholic fatty liver disease (NAFLD) is a term for a range of conditions caused by a build-up of fat in the liver.

Non-alcoholic steatohepatitis (NASH) is an advanced form of NAFLD in which accumulation of fat causes inflammation and changes the structure of liver cells. This can lead to fibrosis and eventually cirrhosis.

The **NASH clinical research network (CRN)** system uses the **NAFLD activity score (NAS)** to assess the histological stage of NAFLD from liver biopsy information (see table 1 of the diagnostics assessment report). The

NAS is the unweighted sum of the individual scores for steatosis, hepatocellular ballooning and lobular inflammation. A NAS of 5 or more indicates a diagnosis of NASH. The NASH-CRN system also includes a fibrosis staging system which is evaluated separately from the NAS. Typically, F1, F2, F3 are considered to represent minimal, significant and advanced fibrosis respectively, and F4 to represent cirrhosis.

Proton density fat fraction (PDFF) is an MRI technique that is offered as part of the LiverMultiScan package, but is also available using software provided by MRI manufacturers. It provides an estimate of steatosis (proportion of fat) in the liver.

A **receiver operating characteristic (ROC)** curve plots the sensitivity of a test against the false positive rate (or 1−specificity) at different diagnostic thresholds. The area under the curve can be used to provide an estimate of the accuracy of the test.

T1 and **T2*** are time constants (measured in milliseconds) describing the decay of a magnetic resonance signal.

Transient elastography evaluates liver stiffness by measuring the speed of a vibration generated on the skin using an ultrasound probe. It can also be used to measure liver fat.