FINAL REPORT

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

MRI-based technologies for the assessment of patients with non-alcoholic fatty liver disease [DAP59]

Confidential until published

This Diagnostics Assessment Report protocol was commissioned by the NIHR Evidence Synthesis Programme as project number 135067

Completed 23 March 2022

[redacted] REPRESENTS REDACTED CONFIDENTIAL INFORMATION

UNIVERSITY OF LIVERPOOL

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP

A MEMBER OF THE RUSSELL GROUP

Title: MRI-based technologies for the assessment of patients with non-alcoholic fatty liver disease [DAP59]

Produced by: Liverpool Reviews & Implementation Group (LR*i*G)

Authors: Rui Duarte, Deputy Director, LR*i*G, University of Liverpool

Rebecca Bresnahan, Research Associate (Clinical Effectiveness), LR*i*G, University of Liverpool

James Mahon, Director, Coldingham Analytical Services, Berwickshire

Sophie Beale, Director, Hare Research, North Yorkshire

Marty Chaplin, Research Associate (Medical Statistician), LR*i*G, University of Liverpool

Devarshi Bhattacharyya, Health Economic Modeller, LR*i*G, University of Liverpool

Rachel Houten, Health Economic Modeller, LR*i*G, University of Liverpool

Katherine Edwards, Research Associate (Clinical Effectiveness), LR*i*G, University of Liverpool

Sarah Nevitt, Research Associate (Medical Statistician), LR*i*G, University of Liverpool

Michelle Maden, Information Specialist, LR*i*G, University of Liverpool

Angela Boland, Director, LR*i*G, University of Liverpool

Correspondence to: Rebecca Bresnahan, Liverpool Reviews and Implementation Group, [redacted]

Date completed: 23 March 2022

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number 135067

Rider on responsibility for report: The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

Declared competing interests of the authors: The authors report no conflicts of interest.

This report should be referenced as follows: Duarte R, Bresnahan R, Mahon J, Beale S, Chaplin M, Bhattacharyya D, Houten R, Edwards K, Nevitt S, Maden M and Boland A. MRI-based technologies for the assessment of patients with non-alcoholic fatty liver disease [DAP59]. Liverpool Reviews and Implementation Group (LR*i*G), University of Liverpool, 2022.

ABSTRACT

Background

Magnetic resonance imaging (MRI) based technologies are non-invasive diagnostic tests that can be used to assess non-alcoholic fatty liver disease (NAFLD) and potentially identify which patients should be referred for liver biopsy.

Objectives

The objectives of this study were to assess the diagnostic test accuracy (DTA), clinical impact and cost effectiveness of two MRI-based technologies, LiverMultiScan and magnetic resonance elastography (MRE), for patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed and who have indeterminate results from fibrosis testing, or for whom transient elastography (TE) or acoustic radiation force impulse (ARFI) is unsuitable, or who have discordant results from fibrosis testing.

Data sources

The data sources searched were MEDLINE, MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Embase, Cochrane Database of Systematic Reviews, Cochrane Central Database of Controlled Trials, Database of Abstracts of Reviews of Effects and the Health Technology Assessment.

Methods

The systematic review methods followed published guidance. Two reviewers screened the search results (database inception to October 2021), extracted data and assessed the quality of the included studies. Summary DTA estimates were calculated using bivariate models, and a summary receiver operating characteristic (ROC) curve was calculated using a hierarchical model. An economic model was developed to estimate the cost effectiveness of MRI-based technologies.

Results

Thirteen studies (15 publications) were identified for inclusion in the DTA review and 11 studies (14 publications) were identified for inclusion in the clinical impact review. However, the evidence for patients who have indeterminate or discordant results from fibrosis testing was limited to one study that evaluated the DTA and clinical impact of LiverMultiScan and performed an economic evaluation. No studies were identified for patients for whom TE or ARFI were unsuitable.

The LiverMultiScan sensitivity and specificity values for the diagnosis of fibrosis that were reported in the study that evaluated DTA for patients with NAFLD who had indeterminate or discordant results from fibrosis testing ranged from 50% to 88% and from 42% to 75% respectively.

Considering additional evidence provided by Perspectum Ltd for advanced fibrosis (≥F3), the pooled sensitivity and specificity values for LiverMultiScan iron corrected longitudinal relaxation time (cT1) were 60.2% (95% confidence interval [CI]: 50.9% to 68.8%) and 65.4% (95% CI: 55.8% to 73.9%) respectively. The summary ROC for MRE for advanced fibrosis (≥F3) indicated high DTA but not all observed study results lay close to the curve.

Acceptability of LiverMultiScan was reported in one study and was generally positive.

The base case incremental cost effectiveness ratio (ICER) per quality adjusted life year (QALY) gained results for six diagnostic test strategies showed that the LiverMultiScan plus biopsy pathway was dominated by the biopsy only pathway and, for the other two diagnostic test strategies, the ICERs per QALY gained were £749,886 and £1,266,511. The results from EAG threshold and scenario analyses demonstrated that these results were robust to plausible variations in the magnitude of key parameters.

Limitations

DTA, clinical impact and cost effectiveness data for MRI-based technologies for the population that is the focus of this assessment were limited.

Conclusions

MRI-based technologies may be useful to identify patients who may benefit from additional testing in the form of liver biopsy and those for whom this additional testing may not be necessary. However, there is a paucity of DTA and clinical impact data for patients who have indeterminate results from fibrosis testing, for whom TE or ARFI are unsuitable or patients who have discordant results from fibrosis testing.

The use of MRI-based technologies for assessing NAFLD for patients with inconclusive results from previous fibrosis testing is unlikely to be a cost effective use of NHS resources compared with liver biopsy only, given the assumptions used in the EAG cost effectiveness model.

Study registration

This study is registered as PROSPERO CRD42021286891.

Funding

Funding for this study was provided by the Evidence Synthesis Programme of the National Institute for Health Research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Cost effectiveness analysis	An economic analysis that converts effects into health terms and describes the costs per additional health gain
Decision modelling	A theoretical construct that allows the comparison of the relationship between costs and outcomes of alternative healthcare interventions
Decision tree	A model of a series of related choices and their possible outcomes
False negative	An incorrect negative test result – an affected individual with a negative test result
False positive	An incorrect positive test result – an unaffected individual with a positive test result
Incremental cost effectiveness ratio	The difference in the mean costs of two interventions in the population of interest divided by the difference in the mean outcomes in the population of interest
Index test	The test whose performance is being evaluated
Meta-analysis	A statistical technique used to combine the results of two or more studies and obtain a combined estimate of effect
Negative predictive value	The probability that people with a negative test result truly do not have the disease
Positive predictive value	Probability that people with a positive test result truly have the disease
Receiver operating characteristic curve	A graph which illustrates the trade-offs between sensitivity and specificity that result from varying the diagnostic threshold
Reference standard	The best currently available diagnostic test against which the index test is compared
Sensitivity	The proportion of people with the target disorder who have a positive test result
Specificity	The proportion of people without the target disorder who have a negative test result
True negative	A correct negative test result – an unaffected individual with a negative test result
True positive	A correct positive test result – an affected individual with a positive test result

LIST OF ABBREVIATIONS

ARFI	Acoustic radiation force impulse
AUROC	Area under the receiver operating characteristic curve
BMI	Body mass index
BSG	British Society of Gastroenterology
CASP	Critical Appraisal Skills Programme
CCG	Clinical Commissioning Group
CD	Cannot determine
CDSR	Cochrane Database of Systematic Reviews
CEA	Cost effectiveness analysis
CENTRAL	Cochrane Central Database of Controlled Trials
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CI	Confidence interval
CRD	Centre for Reviews and Dissemination
CRN	Clinical Research Network
CSR	Clinical study report
cT1	Iron-corrected T1
DAP	Diagnostics Assessment Programme
DARE	Database of Abstracts of Reviews of Effects
DTA	Diagnostic test accuracy
EAG	External Assessment Group
EASL	European Association for the Study of the Liver
ELF	Enhanced liver fibrosis
FIB-4	Fibrosis-4 index
FN	False negative
FP	False positive
fhROI	Free hand region of interest
GLP1	Glucagon-like peptide 1
HR	Hazard ratio
HTA	Health technology assessment
LIF	Liver inflammation and fibrosis
MRE	Magnetic resonance elastography
MRI	Magnetic resonance imaging
MRR	Mortality rate ratio
MRS	Magnetic resonance spectroscopy
NAFLD	Non-alcoholic fatty liver disease
NAS	NAFLD activity score
NASH	Non-alcoholic steatohepatitis
NFS	NAFLD fibrosis score
NG	NICE guideline
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIH	National Institute of Health

OR	Odds ratio
PDFF	Proton density fat fraction
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality adjusted life year
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
ROI	Region of interest
SoC	Standard of care
SGLT2	Sodium-glucose co-transporter 2
srROI	Small round regions of interest per slice
T_1	Longitudinal relaxation time
TE	Transient elastography
TN	True negative
TP	True positive
2x2 data	Numbers of true positive, false positive, true negative and false negative test results

1 PLAIN ENGLISH SUMMARY

Non-alcoholic fatty liver disease (NAFLD) includes a range of conditions that are caused by a build-up of fat in the liver, and not by alcohol consumption. This build-up of fat can cause inflammation. Persistent inflammation can cause scar tissue (fibrosis) to develop. Severe fibrosis can cause permanent liver damage (cirrhosis), which can lead to liver failure and liver cancer.

In the NHS, patients with NAFLD undergo tests to determine whether they have fibrosis. The test results are not always accurate and multiple tests can give conflicting results. Some of the tests may not be suitable for patients who are obese or who have a very high body mass index.

In the NHS, a liver biopsy may be offered to patients with inconclusive or conflicting test results or to those patients for whom other tests are unsuitable. However, liver biopsy is expensive, and is associated with side-effects such as pain and bleeding. Magnetic resonance imaging (MRI)-based testing could be used as an extra test to help clinicians assess NAFLD and identify patients who may need a liver biopsy.

We assessed two MRI-based diagnostic tests, LiverMultiScan and magnetic resonance elastography (MRE). LiverMultiScan is imaging software that is used alongside MRI to measure markers of liver disease. MRE is used in some NHS centres to assess liver fibrosis; however, MRE requires more equipment than just an MRI scanner.

We reviewed all studies examining how well LiverMultiScan and MRE assess patients with NAFLD. We also built an economic model to estimate the costs and benefits of using LiverMultiScan to identify patients who should be sent for a biopsy. Results from the model showed that LiverMultiScan may not provide good value for money to the NHS.

2 SCIENTIFIC SUMMARY

Background

Non-alcoholic fatty liver disease (NAFLD) is an umbrella term for a range of conditions caused by a build-up of fat in the liver that has not been caused by alcohol consumption. NAFLD covers a spectrum of histological lesions ranging from steatosis (simple fatty liver) to complex patterns of hepatocyte injury, inflammation and fibrosis.

In the current NHS diagnostic pathway for staging fibrosis (based on guidelines and expert advice to NICE), patients with NAFLD (confirmed by ultrasound and liver aetiology screen) are referred for the fibrosis-4 (FIB-4), NAFLD fibrosis score (NFS) or enhanced liver fibrosis (ELF) test as first-line testing. Patients with an indeterminate result from first-line testing are referred for second-line testing using transient elastography (TE), acoustic radiation force impulse (ARFI) or the ELF test, if it had not already been used as a first-line test. Patients with indeterminate or discordant results from fibrosis testing and patients with high risk of advanced fibrosis are considered for liver biopsy. Magnetic resonance imaging (MRI)-based testing could be used as an additional, non-invasive, diagnostic test to help clinicians stage NAFLD and potentially identify which patients should be referred for liver biopsy. Liver biopsy is expensive and is an invasive procedure that is associated with complications.

Objectives

The objectives of this study were to assess the diagnostic test accuracy (DTA), the clinical impact and the cost effectiveness of two non-invasive MRI-based technologies, namely LiverMultiScan and magnetic resonance elastography (MRE), for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable, or who had discordant results from fibrosis testing. To achieve the study objectives, the External Assessment Group (EAG):

1. conducted a systematic literature review to evaluate the (1) DTA of MRI-based technologies for the assessment of fibrosis, inflammation, and steatosis for a patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed, using liver biopsy as the reference standard, and (2) the clinical impact of MRI-based technologies
2. conducted a systematic literature review to explore the cost effectiveness of MRI-based technologies as diagnostic tools and built a de novo economic model to assess the cost effectiveness of two diagnostic pathways, namely MRI-based technologies plus biopsy and liver biopsy.

Methods: assessment of diagnostic test accuracy and clinical impact

Electronic databases (MEDLINE, MEDLINE Epub Ahead of Print In-Process & Other Non-Indexed Citations, Embase, Cochrane Databases of Systematic Reviews, Cochrane Central Database of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database) were searched from inception to 4th October 2021. Eligible studies assessed the DTA or clinical impact of LiverMultiScan or MRE for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed (who have indeterminate results from fibrosis testing, for whom TE or ARFI is unsuitable, or who have discordant results from fibrosis testing).

Two reviewers independently screened the titles and abstracts of all reports identified through electronic database searches and of all full-text articles subsequently obtained for assessment. Data extraction and quality assessment were conducted by one reviewer and checked for agreement by a second reviewer. The methodological quality of the included DTA studies was assessed using the QUality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. The methodological quality of randomised controlled trials (RCTs) evaluating the clinical impact of MRI-based technologies were assessed using the Cochrane Risk of Bias 2.0 tool. The National Institute of Health study quality assessment tools for cohort studies, case-control studies and before-after (pre-post) studies with no control group were used to assess risk of bias of included non-randomised studies. Qualitative studies were assessed using the Critical Appraisal Skills Programme (CASP) qualitative studies checklist.

The sensitivity and specificity of each index test were summarised in forest plots. Where at least three studies provided both sensitivity and specificity data for a specific combination of index test, diagnosis of interest, and cut-off value, a bivariate random-effects meta-analysis to provide pooled estimates of sensitivity and specificity was considered. We did not perform bivariate meta-analyses where statistical heterogeneity between the studies (assessed by visually examining forest plots) was so great that pooled estimates of sensitivity and specificity would have been meaningless. Where at least three studies provided both sensitivity and specificity data for a specific combination of index test and diagnosis of interest, but used different cut-off values for the index test, we used a hierarchical model to estimate a summary receiver operating characteristic (ROC) curve.

Methods: assessment of cost effectiveness

The External Assessment Group (EAG) appended an economic evaluation-specific search filter to the clinical search strategies to identify published cost effectiveness studies. In addition, two databases of economic publications (EconLit [EBSCO] and the Cost-Effectiveness Analysis [CEA] registry) were searched from inception until 4th October 2021.

The EAG developed a simple, flexible de novo model to estimate the cost effectiveness of an MRI-based technologies plus biopsy pathway versus liver biopsy only pathway.

Results

The EAG searches of the electronic databases, and reference lists of relevant studies and systematic reviews identified 4489 records (3331 unique records). Although all the identified studies for inclusion in the DTA, clinical impact and cost effectiveness reviews included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed, only one study provided results for patients with NAFLD who had indeterminate or discordant results from fibrosis testing. No studies were identified that considered patients for whom TE or ARFI was unsuitable.

2.1.1 Diagnostic test accuracy

The EAG identified 13 studies (15 publications). Two studies (four publications) were evaluations of LiverMultiScan, 10 studies (10 publications) were evaluations of MRE, and one study (one publication) was an evaluation of LiverMultiScan and MRE.

MRI-based technology: LiverMultiScan

For the LiverMultiScan proton density fat fraction (PDFF) and LiverMultiScan cT1 outputs, 2x2 data were available from three studies. The EAG considers that the Eddowes 2018 study is the most relevant study to this assessment. Eddowes 2018 recruited patients who were scheduled for non-targeted liver biopsy to stage fibrosis after inconclusive non-invasive assessment of fibrosis or to make a diagnosis after a range of non-invasive tests had not confirmed a diagnosis. For diagnosis of fibrosis, estimates from Eddowes 2018 ranged from 50% to 88% for sensitivity and from 42% to 75% for specificity. Sensitivity and specificity values for fibrosis testing in Eddowes 2018 were consistently higher for LiverMultiScan cT1 than for LiverMultiScan PDFF.

Data from three studies were included in the meta-analyses for LiverMultiScan. For advanced fibrosis (≥F3), the pooled sensitivity and specificity values were higher for LiverMultiScan cT1 (sensitivity=60.2%, 95% confidence interval [CI]: 50.9% to 68.8%; specificity=65.4%, 95% CI: 55.8% to 73.9%) than for LiverMultiScan PDFF (sensitivity=38.6%, 95% CI: 23.8% to 56.0%; specificity=43.6%, 95% CI: 30.7% to 57.5%).

MRI-based technology: MRE

For the MRE test, 2x2 data were available from four studies. Estimates of sensitivity and specificity for advanced fibrosis (≥F3) were high and ranged from 71% to 100% and 79% to 93%, respectively. However, the cut-off values used to indicate a positive result from the index

test varied between studies therefore a summary ROC curve was estimated. The summary ROC curve indicates high DTA. However, observed study results do not all lie close to the summary ROC curve which could be due to small sample sizes and/or clinical and methodological heterogeneity between the included studies.

2.1.2 Clinical impact

Eleven studies (14 publications) were included in the clinical impact review. Five studies (eight publications) were evaluations of LiverMultiScan and six studies (six publications) were evaluations of MRE.

MRI-based technology: LiverMultiScan

Two studies reported on the prognostic ability of LiverMultiScan cT1. However, neither study reported results specifically for the subpopulation of patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. One study reported that LiverMultiScan cT1 and LiverMultiScan PDFF could [redacted] the number of unnecessary biopsies for patients with [redacted] and [redacted] to diagnose [redacted] and [redacted] (EAG calculated odds ratio [OR]=[redacted], 95%CI: [redacted] to [redacted]) and for patients with [redacted] to diagnose [redacted] EAG calculated OR=[redacted], 95% CI: [redacted] to [redacted]) when compared to standard of care. Three studies reported the test failure rate of LiverMultiScan for patients with all liver aetiologies. The test failure rate ranged from 5.3% to 7.6%. One study reported the test failure rate for LiverMultiScan for patients with NAFLD (5.6%). Acceptability of LiverMultiScan was reported in a qualitative study and was generally positive.

MRI-based technology: MRE

Six studies reported the test failure rate of MRE for patients with all liver aetiologies. The test failure rate ranged from 0.0% to 7.6%. Three studies reported the test failure rate for MRE specifically for patients with NAFLD. The EAG performed a fixed-effects meta-analysis to obtain a pooled estimate of 4.2% (95% CI: 2.5% to 6.2%) test failure rate for patients with NAFLD.

Despite conducting additional targeted searches, the EAG did not identify any relevant studies that provided evidence of the clinical impact of MRI-based technologies for patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed, for the remaining clinical impact outcomes listed in the final scope issued by NICE.

2.1.3 Cost effectiveness

The EAG base case ICER per QALY gained results for six of the eight diagnostic test strategies considered, showed that the LiverMultiScan plus biopsy pathway was dominated by the biopsy only pathway. For the other two diagnostic test strategies, the ICERs per QALY gained were £749,886 and £1,266,511. Results from the EAG threshold and scenario analyses demonstrated that these results were robust to plausible variations in the magnitude of key parameters.

Conclusions

The DTA, clinical impact and cost effectiveness data for MRI-based technologies are limited for patients who have indeterminate results from fibrosis testing, for whom TE or ARFI are unsuitable or patients who have discordant results from fibrosis testing.

Only one small LiverMultiScan study provided DTA and population prevalence data for patients described in the final scope issued by NICE. It is unclear whether sensitivity and specificity estimates reported by this small study will give clinicians sufficient confidence to use LiverMultiScan test results to triage patients with inconclusive results from previous fibrosis testing to biopsy. Cost effectiveness results from the EAG model are only informative if clinicians have confidence in LiverMultiScan DTA data. Using the available DTA and population prevalence data, EAG cost effectiveness results showed that LiverMultiScan is unlikely to be cost effective at current prices when used to triage patients with inconclusive results from previous fibrosis testing to biopsy.

LiverMultiScan data are not available for patients for whom TE or ARFI were unsuitable. Further, no MRE DTA data were available for the population described in the final scope issued by NICE. The EAG was unable to generate cost effectiveness results for this technology; however, even if MRE was 100% accurate, due to high population prevalence estimates, it is unlikely that MRE would be cost effective at current prices.

Study registration

This study is registered as PROSPERO CRD42021286891.

Funding

Funding for this study was provided by the Evidence Synthesis Programme of the National Institute for Health Research.

3 BACKGROUND

3.1 Purpose of the assessment

The purpose of this assessment is to explore whether two non-invasive magnetic resonance imaging (MRI) based technologies, specifically LiverMultiScan and magnetic resonance elastography (MRE), can be used to assess non-alcoholic fatty liver disease (NAFLD), and whether use of these technologies represents a cost effective use of NHS resources compared to a diagnostic pathway that does not include them.

In the current NHS diagnostic pathway, patients with NAFLD who have indeterminate results from fibrosis testing, for whom transient elastography (TE) or acoustic radiation force impulse (ARFI) is unsuitable, or who have discordant results from fibrosis testing, are considered for liver biopsy. However, liver biopsy is expensive and is an invasive procedure that is associated with well-recognised complications. Additional non-invasive tests results may help to determine which patients should be referred for liver biopsy.

3.2 Target condition

NAFLD is an umbrella term for a range of conditions caused by a build-up of fat in the liver that has not been caused by alcohol consumption.[1] NAFLD covers a spectrum of histological lesions ranging from steatosis (simple fatty liver) to complex patterns of hepatocyte injury, inflammation and fibrosis.[2] Liver biopsy is the only diagnostic procedure that can reliably assess these various patterns.[2] Approximately 7000 to 8000 patients per year undergo liver biopsy in the UK.[3] Biopsy results are required to determine appropriate referral and treatment strategies for patients with NAFLD.[4] However, liver biopsy is an invasive procedure that is associated with well-recognised complications, including minor bleeding (1 in 500), severe intraperitoneal bleeding (1 in 2,500 to 1 in 10,000) and death (1 in 10,000 to 1 in 12,000).[5] Liver biopsy complications lead to hospitalisation for 1% to 3% of patients, most commonly because of pain or hypotension.[5]

It is estimated that between 20%[1] to 33%[6] of people in the UK have early stage NAFLD (simple fatty liver). Risk factors for NAFLD include type 2 diabetes, high blood pressure or high cholesterol, underactive thyroid, smoking and being overweight or obese.[7] The prevalence of NAFLD increases with age and is most prevalent in men aged 40 to 65 years.[8] However, the prevalence of NAFLD is increasing in younger people due to rising levels of obesity among children (aged 1 to under 16 years) and young people (aged 16 to under 18 years).[9] Studies have reported that 34% to 38% of children with obesity have biopsy-proven NAFLD.[10]

The four main stages of NAFLD are:[6]

1. simple fatty liver (steatosis) - a largely harmless build-up of fat in liver cells. Approximately 20% of patients with NAFLD develop non-alcoholic steatohepatitis (NASH)
2. NASH - the build-up of fat in the liver leads to inflammation. Approximately 25% to 40% of patients with NASH develop liver fibrosis and approximately 20% to 30% of patients with NASH develop cirrhosis.[11] It is estimated that 3.3 million people in the UK have NASH,[6] and that approximately 80% of these people have undiagnosed NASH because early-stage NASH is usually asymptomatic.[12,13] It is widely accepted that liver fibrosis develops as a result of liver damage that is secondary to NASH.[14]
3. fibrosis - persistent inflammation develops in response to the build-up of fat and causes scar tissue formation in the liver and blood vessels. Approximately 21% to 28% of patients with fibrosis develop cirrhosis.[15]
4. cirrhosis - chronic inflammation in the liver produces severe and irreversible scarring causing liver damage. Cirrhosis can lead to liver failure and liver cancer.[16]

The NASH Clinical Research Network (CRN) system uses the NAFLD Activity Score (NAS) to assess the histological stage of NAFLD from liver biopsy information (Table 1).[17] The NAS is the unweighted sum of the individual scores for steatosis, hepatocellular ballooning and lobular inflammation. A NAS of ≥5 indicates a diagnosis of NASH.[17] The NASH CRN system also includes a fibrosis staging system which is evaluated separately from the NAS.[17] Typically, F1, F2, F3 are considered to represent minimal, significant and advanced fibrosis, respectively and F4 to represent cirrhosis. Compared to patients with minimal to significant fibrosis (F1 to F2), patients with advanced fibrosis to cirrhosis (F3 to F4) are at increased risk of liver events (hazard ratio [HR]=5.58, 95% confidence intervals [CI] 3.70 to 8.40) including liver failure, gastroesophageal varices, ascites, encephalopathy, hepatopulmonary syndrome, hepatocellular carcinoma.[14]

Table 1 NASH Clinical Research Network histological scoring system

NAFLD activity score (NAS)					
Steatosis (brunt grade)		Hepatocyte ballooning		Lobular inflammation (foci per 200x field)	
Score	Definition	Score	Definition	Score	Definition
0	<5%	0	None	0	None
1	5% to 33%	1	Few	1	<2
2	34% to 66%	2	Many	2	2 to 4
3	>66%	-	-	3	>4

Fibrosis level			
Stage	Definition		
F0	No fibrosis		
F1	Perisinusoidal or periportal fibrosis	F1A	Mild, zone 3, perisinusoidal
		F1B	Moderate, zone 3, perisinusoidal
		F1C	Portal/periportal
F2	Perisinusoidal and portal/periportal fibrosis		
F3	Bridging fibrosis (across lobules, between portal areas, or between portal areas and central veins)		
F4	Cirrhosis		

NAFLD=non-alcoholic fatty liver disease; NAS=NAFLD Activity Score
Source: Kleiner et al 2005[17]

Compared to patients with NAFLD with no fibrosis (F0), the risk of liver-related mortality in patients with NAFLD with fibrosis (F1 to F4) increases exponentially with each stage of fibrosis (F1, mortality rate ratio [MRR]=1.41, 95% CI 0.17 to 11.95; F2, MRR=9.57, 95% CI 1.67 to 54.93; F3, MRR=16.69, 95% CI 2.92 to 95.36; and F4, MRR=42.30, 95% CI 3.51 to 510.34).[18] The risk of liver-related mortality in patients with NAFLD who have a fibrosis level ≥F2 is statistically significantly greater ($p<0.02$) than in patients with NAFLD who do not have fibrosis (F0).[18]

3.3 Current NHS diagnostic practice

The National Institute for Health and Care Excellence (NICE) guideline[9] (Non-alcoholic fatty liver disease: assessment and management, NG49) includes a summary of current best practice for the diagnosis and management of NAFLD.

In NG49,[9] it is recommended that clinicians should:

- suspect NAFLD in patients with type 2 diabetes or metabolic syndrome
- take an alcohol-related history from patients presenting with symptoms of NAFLD to rule out alcohol-related liver disease
- not use routine liver blood tests to rule out NAFLD.

For adults, NAFLD is most often suspected following abnormal liver function test results in the primary care setting,[19] or following an incidental ultrasound finding.[9,20] Clinical advice to the

External Assessment Group (EAG) is that NAFLD is a diagnosis of exclusion, meaning that clinicians exclude other liver disease aetiologies based on liver aetiology screen results, and then use the patient's clinical history to confirm a diagnosis of NAFLD. Clinical advice to the EAG is that NAFLD is confirmed in the primary or secondary care setting before referral for advanced fibrosis testing in the secondary care setting (Figure 1).

Figure 1 presents an overview of the current diagnostic pathway for the assessment of fibrosis in the NHS based on guidelines[8,9,21,22] and expert advice to NICE.[23]

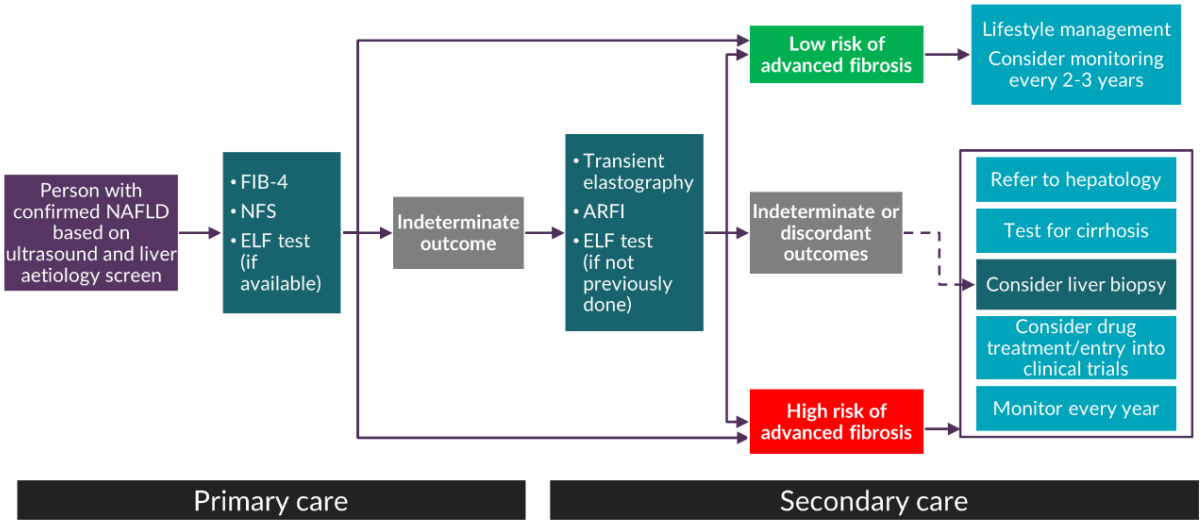

Figure 1 Overview of current diagnostic pathway for assessment of fibrosis in the NHS, based on guidelines and expert advice

ARFI=acoustic radiation force impulse; ELF=enhanced liver fibrosis; FIB-4=fibrosis-4; MRI=magnetic resonance imaging; NALFD=non-alcoholic fatty liver disease; NFS=NAFLD fibrosis score
Source: Final scope[23] issued by NICE

NG49[9] includes a diagnostic test accuracy (DTA) review. Results from the review were used to identify the most accurate assessment tool for diagnosing NAFLD in adults, young people and children, and for identifying the severity or stage of NAFLD. In NG49[9], it is considered that liver biopsy is the 'gold standard' for diagnosis and staging of NAFLD. However, in NG49[9], it is reported that it is not feasible to perform liver biopsy in large numbers of at risk patients because biopsy is invasive and expensive. The recommendations for non-invasive tests are as follows:

- offer testing for advanced liver fibrosis to patients with NAFLD and consider using the enhanced liver fibrosis (ELF) test
- patients with NAFLD and an ELF score ≥10.51 should be diagnosed with advanced liver fibrosis
- patients with NAFLD and an ELF score <10.51 are unlikely to have advanced liver fibrosis and should be reassessed regularly (adults every 3 years, and children and young people annually)

- offer a liver ultrasound to test children and young people for NAFLD if they have type 2 diabetes or metabolic syndrome and do not misuse alcohol. Children and young people are diagnosed with NAFLD if a fatty liver is detected on ultrasound. If the ultrasound is normal, then offer to retest with liver ultrasound for NAFLD every 3 years.

In the British Society of Gastroenterology (BSG) national guidelines,[21] the recommendations are that liver biopsy should not be used as first-line testing for NAFLD and disease staging. According to the BSG national guidelines,[21] only patients with high risk of advanced liver disease or with suspected concomitant secondary liver disease should be referred for liver biopsy. The BSG national guidelines[21] and the Lancet Commission into liver disease in the UK[24] recommendations are that the Fibrosis-4 (FIB-4) test and the NAFLD fibrosis score (NFS) test should be used as first-line testing to assess the stage of fibrosis. The FIB-4 and NFS tests have high negative predictive value and therefore can accurately exclude patients who do not have advanced fibrosis.[24]

However, Byrne 2018[22] recommends that ultrasound should be used as first-line testing to diagnose hepatic steatosis and to exclude other liver pathology and that ELF and TE should be used to investigate for liver fibrosis in patients with confirmed hepatic steatosis.

The BSG national guidelines,[21] state that:

- a FIB-4 score <1.30 or a NFS <-1.455 demonstrates that patients have low risk of advanced fibrosis
- patients with low risk of advanced fibrosis can be managed in primary care and advised on lifestyle modifications
- patients with an indeterminate FIB-4 score (1.3 to 3.25) or NFS (-1.455 to 0.672) should undergo second-line testing using the ELF test, TE or ARFI
- patients with FIB-4 score >3.25 or NFS >0.672 should be considered to have high risk of advanced fibrosis and should be referred to a specialist clinic irrespective of second-line tests
- if the non-invasive tests are not able to exclude advanced fibrosis, then a liver biopsy should be considered to assess NAFLD and to rule out other concomitant liver diseases.

In the UK, the tests used to diagnose advanced liver fibrosis vary by NHS centre, depending on availability.[25] In NG49,[9] there is a list of alternative diagnostic tools that have been used in NHS clinical practice to diagnose and assess advanced fibrosis and cirrhosis. These tools include TE, ARFI, MRI, MRI proton density fat fraction (PDFF), magnetic resonance spectroscopy (MRS), MRE, shear wave elastography and liver biopsy. The use of liver biopsy in current NHS diagnostic practice is described in Section 3.9.1.

Findings from a cross-sectional survey[25] of liver disease management, conducted from June to October 2020 indicated that only 25% (40/159) of UK Clinical Commissioning Groups

(CCGs) used TE and only 16% (26/159) used the ELF test to assess liver fibrosis. Approximately two-fifths of UK CCGs (44%, 70/159) followed the BSG national guidelines[21] and used FIB-4 and NFS to assess liver fibrosis.

3.4 Treatment options

There are currently no pharmacological treatments licensed specifically for the treatment of NAFLD, although there are weak recommendations (NG49[9]) for the off-licence use of vitamin E and pioglitazone for NAFLD. Current clinical management of NAFLD relies on lifestyle advice and modifications.[21] However, novel therapies are in clinical development, such as glucagon-like peptide 1 (GLP1) agonists and sodium-glucose co-transporter 2 (SGLT2) inhibitors.[26]

NG49[9] recommendations for lifestyle modifications for patients diagnosed with NAFLD are as follows:

- offer advice on physical activity and diet to patients with NAFLD who are overweight or obese and explain that exercise may reduce liver fat content
- consider the lifestyle interventions detailed in NICE's obesity guideline[27] for patients with NAFLD, regardless of their body mass index (BMI)
- explain the importance of adhering to the national recommended limits for alcohol consumption.

NG49[9] pharmacological therapy recommendations are as follows:

- pharmacological therapy may be considered in secondary or tertiary care settings only
- consider pioglitazone or vitamin E for adults with advanced liver fibrosis, whether they have diabetes or not
- consider vitamin E for children with advanced liver fibrosis, whether they have diabetes or not (only in tertiary care settings)
- consider vitamin E for young people with advanced liver fibrosis, whether they have diabetes or not
- offer to retest patients with advanced liver fibrosis 2 years after they start a new pharmacological therapy to assess whether treatment is effective
- consider using the ELF test to assess whether pharmacological therapy is effective
- if an adult's ELF test score has risen, stop either vitamin E or pioglitazone and consider switching to the other pharmacological therapy
- if a child or young person's ELF test score has risen, stop vitamin E.

Although pioglitazone or vitamin E may be offered to patients with advanced liver fibrosis,[9] clinical advice to NICE[23] is that this may not be current NHS practice. Patients with advanced fibrosis may be considered for entry into clinical trials of novel therapies for NAFLD.

3.5 Population

In line with the final scope[23] issued by NICE, the population of interest is patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed. This population consists of:

- patients who have indeterminate results from fibrosis testing
- patients for whom TE or ARFI is unsuitable
- patients who have discordant results from fibrosis testing.

If data permitted, additional subgroup analyses were to be considered (for example, based on prior tests for fibrosis, children or young people).

3.5.1 Patients who have indeterminate results from fibrosis testing

Results from TE, ARFI and ELF tests may indicate that some level of fibrosis is present but may not be able to confirm the presence of advanced fibrosis (F3) or cirrhosis (F4). Where results show that some level of fibrosis is present, but the level of fibrosis cannot be confirmed, these results are referred to as indeterminate results. The range of values used to define indeterminate results and the language used to describe indeterminate results varies across guidelines and clinical studies (e.g., 'grey zone',[28] 'intermediate risk'[21] and 'inconclusive results'[29]).

In the BSG guidelines,[21] it is recommended that clinicians should consider liver biopsy for patients with a TE score between 7.9kPa and 9.6kPa (intermediate risk of advanced fibrosis), and for patients with a TE score >9.6kPa (high risk of advanced fibrosis). In the European Association for the Study of the Liver (EASL) guidelines,[30] it is recommended that a TE score <8kPa rules out advanced fibrosis and that a TE score ≥8kPa represents an intermediate to high risk of advanced fibrosis. Clinical advice to NICE[23] is that indeterminate results are also possible from ARFI, although the exact values for an indeterminate ARFI result depend on the device manufacturer.

Clinical advice to NICE[23] is that indeterminate results are possible from the ELF test. ELF test scores between 7.8 and 10.5,[22] or 7.7 and 9.7 are considered to be indeterminate results.[31] In the EASL guidelines,[30] it is recommended that an ELF score <9.8 rules out advanced fibrosis for patients with NAFLD.

In current NHS practice, a biopsy may be considered for patients with indeterminate results from fibrosis testing. MRI-based testing could therefore be used as an additional, non-invasive, diagnostic test to help clinicians assess the need for a liver biopsy. However, the EAG notes that the range of values used to define an indeterminate result can vary across guidelines for the same test and the terms 'indeterminate' and 'intermediate' are used interchangeably. It is

therefore unclear which range of values from non-invasive tests should indicate an indeterminate result and signal that patients should be referred for MRI-based testing.

3.5.2 Patients for whom TE or ARFI is unsuitable

TE and ARFI may not be suitable tests for people with a very high body mass index (BMI), or those with significant ascites because excessive amounts of fat and fluid overlying the liver can prevent the propagation of shear waves necessary to assess liver stiffness.[23] The tests may fail, or the clinicians may decide not to refer patients for these tests because they are likely to fail.

Liver biopsy may be considered for this subgroup of patients to determine the stage of fibrosis. MRI-based testing could be used as an additional, non-invasive, diagnostic test to help assess the need for a liver biopsy.

3.5.3 Patients who have discordant results from fibrosis testing

Patients with NAFLD may undergo multiple tests to confirm the presence of advanced fibrosis. If the results from these tests are discordant, then liver biopsy should be considered. For example, in the EASL guidelines,[30] it is recommended that patients with discordant results, i.e., patients for whom one non-invasive test indicates low-risk of advanced fibrosis (e.g., FIB-4 <1.30, TE <8kPa or ELF <9.8) but another indicates intermediate to high risk of advanced fibrosis (e.g., FIB-4 ≥1.30, TE ≥8kPa or ELF ≥9.8) patients, should be considered for liver biopsy.

Clinical advice to the EAG is that patients who have indeterminate results, patients for whom TE or ARFI is unsuitable, and patients who have discordant results should be considered for a liver biopsy. MRI-based testing could be used as an additional, non-invasive, diagnostic test to help assess the need for a liver biopsy.

3.6 Interventions / index tests

3.6.1 LiverMultiScan

LiverMultiScan (Perspectum Ltd) is a non-invasive multiparametric MRI-based imaging software application that provides quantitative analysis of liver fat content, liver iron concentration and fibro-inflammation from non-contrast MRI images. The topic selection oversight panel identified LiverMultiScan software as potentially suitable for evaluation by the Diagnostics Assessment Programme (DAP) based on a MedTech Innovation Briefing[32] published by NICE and further information provided by the manufacturer.[23]

LiverMultiScan software enables assessment of liver fat content from PDFF, liver iron concentration from T2* mappings and fibro-inflammation from T1 mappings. The T1 analyses for fibro-inflammation are adjusted for iron level to remove artefacts and increase accuracy.[33] This output is referred to as the iron-correct T1 score (cT1). PDFF is an estimate of the percentage of fat within the liver tissue and is calculated from the ratio of fat versus fat and water in MRI images. PDFF can be computed using the IDEAL (Iterative Decomposition of water and fat with Echo Asymmetry and Least squares estimation) or three-point Dixon method.

LiverMultiScan protocols can be integrated into existing abdominal MRI protocols on Siemens, Philips or GE Healthcare scanners and does not require any contrast agent or additional hardware in addition to the MRI scanner.[23] A 15 minute scan acquisition time is typically required to obtain the MR images for analysis by LiverMultiScan software.[23] Training on how to use the LiverMultiScan protocol takes approximately 3 hours.[23] Technical support from imaging application specialists at Perspectum Ltd is provided by the manufacturer as part of the licence.[34] The imaging data from the MRI scan are sent to Perspectum Ltd via an Amazon hosted cloud service and are analysed by Perspectum Ltd trained operators.[35] The quantitative analysis is returned to clinicians electronically in report format as a PDF document.[35]

Perspectum Ltd suggested to NICE[23] that the normal reference range for MRI PDFF is less than 5.6% liver fat content and that the diagnosis indicated by the cT1 output and the clinical recommendations are as follows:

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

Perspectum Ltd does not propose that LiverMultiScan is suitable for staging fibrosis but considers that LiverMultiScan can stage NAFLD and distinguish between patients with NASH

and high risk NASH.[23] However, in the EASL guidelines,[30] liver biopsy is recommended as the reference standard for the diagnosis of NASH for patients with NAFLD.

3.6.2 Magnetic resonance elastography

MRE is a non-invasive MRI-based technique that uses a mechanical driver to generate shear waves across the liver during an MRI scan.[36] An MRI sequence with motion-encoding gradients measures the propagation of the shear waves across the liver to produce an image (elastogram) showing the distribution of liver stiffness.[36] MRE requires additional hardware to an MRI scanner, including an active acoustic driver, a passive pneumatic driver and a connector.[37] MRE can be used alongside standardised MRI PDFF and iron-assessment packages offered by scanner manufacturers, such as Siemens, Philips or GE Healthcare scanners, to assess fat and iron.[38]

The MRE acquisition is performed during breath-holding and takes 12 to 15 seconds, which is typically repeated four times.[23] The total acquisition time can last approximately one minute.[23] Inadequate breath-holding can produce image artefacts which can affect diagnostic accuracy.[37]

NICE guidelines (NG49[9] and NG50[39]) do not consider the routine use of MRE for diagnosing NAFLD or liver fibrosis or cirrhosis. However, MRE is used in some NHS centres where it is available, when other diagnostic tests have returned indeterminate results.

The commercially-available Resoundant, Inc. MRE platform measures the magnitude of the complex shear modulus of propagating waves to provide liver stiffness outputs (kPa).[40] The complex shear modulus is composed of two components, the storage modulus which describes tissue elasticity and the loss modulus which describes tissue viscosity and the ability to absorb energy.[41] The company, Resoundant Inc., has suggested to NICE[23] that MRE liver stiffness outputs (kPa) can be used to stage liver fibrosis as follows:

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

3.7 Place of the intervention in the diagnostic pathway

The proposed positioning of the two MRI-based technologies is as additional, non-invasive diagnostic tests in the NHS diagnostic pathway for patients with NAFLD who have indeterminate results from fibrosis testing, for whom TE or ARFI is unsuitable, or who have discordant results from fibrosis testing before clinicians consider referral for liver biopsy

(Figure 1). Results from an MRI-based assessment could help clinicians make decisions about whether a liver biopsy is needed and about the extent of future monitoring. For patients who require a liver biopsy, results from an MRI assessment could improve targeting for biopsies by identifying the liver region with the most severe disease. Results from an MRI assessment could also help clinicians target lifestyle intervention advice to patients which may improve uptake and compliance with lifestyle interventions and lead to a reduction in the likelihood of progression to more advanced fibrosis and cirrhosis.

3.8 Comparator

In NHS clinical practice, the populations specified in the final scope[23] issued by NICE would not undergo any further investigation prior to deciding whether a biopsy was required. Clinical experts to NICE[23] commented that, in these populations, the probability of having a biopsy is based on clinical suspicion of advanced fibrosis or cirrhosis (e.g., patient age, weight and comorbidities).

3.9 Reference standard

To assess diagnostic test accuracy, index tests results (i.e., LiverMultiScan and MRE) will be compared to the results of a reference standard (i.e., liver biopsy). The reference standard is used to verify the presence or absence of fibrosis, inflammation and steatosis for patients with NAFLD. The reference standard for this assessment is liver biopsy, as performed and interpreted by a trained healthcare professional.

3.9.1 Liver biopsy

Liver biopsy, an invasive procedure, is considered the gold standard for staging liver fibrosis, inflammation and steatosis, and for diagnosing NASH.[9] During liver biopsy, a small sample of tissue is percutaneously or transvenously removed from the liver using a needle.[42] However, liver biopsies are associated with inter- and intra-observer variability and sampling error.[43,44] Liver biopsies are expensive because patients require outpatient care, specialists (a gastroenterologist, hepatologist or radiologist) are needed to carry out the biopsy, pathologists are needed to examine and report the biopsy results and clinicians are required to interpret biopsy results and recommend clinical management for patients.[9] Liver biopsies can be painful and are associated with a high risk of complications, including bleeding from the biopsy site (0.3% to 10.9%) and major intraperitoneal bleeding (0.1% to 4.6%).[42]

In NG50,[39] it is recommended that clinicians should consider a liver biopsy to diagnose cirrhosis in patients for whom TE is not suitable. In NG49,[9] it is stated that a liver biopsy should not be used to diagnose NAFLD or for monitoring disease progression, and that biopsies

should be avoided in children and young people unless there is an unclear diagnosis or concern about rapid disease progression.

Clinical advice to NICE[23] is that in some NHS centres, liver biopsy is carried out in a large proportion of patients with suspected significant or advanced fibrosis to either confirm the suspected diagnosis or to obtain a diagnosis to allow entry into clinical trials. Clinical advice to the ERG is that liver biopsy results provide information that can be used to inform treatment decisions and clinical management.

Clinical advice to the EAG is that, even after an MRI assessment, patients would be referred for biopsy if the following diagnoses were suspected:

- advanced fibrosis (≥F3)
- steatosis with Brunt grade≥2
- advanced NASH (NAS≥4 and ≥F3)
- high risk of progressive disease (NASH or >F1)

Clinicians do not always refer patients for liver biopsy if they suspect the patient has cirrhosis. Reasons for not referring a patient for a liver biopsy include old age, significant co-morbidities, and being contraindicated for biopsy (e.g., patients with extrahepatic biliary obstruction or bacterial cholangitis).[42] Clinical advice to the EAG is that some patients (5% to 10%) do not wish to proceed with liver biopsy, or are treated at centres without access to liver biopsy.

4 METHODS FOR ASSESSING DIAGNOSTIC TEST ACCURACY AND CLINICAL IMPACT

The EAG conducted a systematic literature review that comprised two parts: (1) DTA review of MRI-based technologies for the assessment of fibrosis, inflammation and steatosis for a population of patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed, using liver biopsy as the reference standard, and (2) clinical impact review of MRI-based technologies compared to no further testing. This population consists of:

- patients who have indeterminate results from fibrosis testing (Section 3.5.1)
- patients for whom TE or ARFI is unsuitable (Section 3.5.2)
- patients who have discordant results from fibrosis testing (Section 3.5.3).

The methods for the systematic review followed the general principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for conducting reviews in health care,[45] NICE's Diagnostics Assessment Programme manual[46] and the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.[47] The systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for DTA studies.[48] The PRISMA-DTA[48] checklist and the PRISMA-DTA[48] for abstracts checklist are presented in Appendices 1 and 2, respectively.

4.1 Search strategy

A single search strategy was used to identify relevant studies. The search strategy was designed to focus on the index tests (i.e., LiverMultiScan and MRE) and the target population (i.e., patients with NAFLD). No study design filters were applied, and all electronic databases were searched from inception to 4th October 2021. Details of individual database searches are provided in Appendix 3; the following databases were searched:

- MEDLINE (via Ovid) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations
- Embase (via Ovid)
- Cochrane Database of Systematic Reviews (CDSR)
- Cochrane Central Database of Controlled Trials (CENTRAL)
- Database of Abstracts of Reviews of Effects (DARE) (via Centre for Reviews and Dissemination)
- Health Technology Assessment (HTA) Database (via International HTA Database).

The results of the searches were uploaded to EndNote X9 and duplicates were systematically identified and removed (MM).

4.1.1 Additional searches (clinical impact review)

Where clinical impact outcome data relating specifically to MRI-based technologies were not identified by the initial search strategy, broader searches were carried out to consider studies of NAFLD populations irrespective of whether MRI-based technologies had been used. MEDLINE and Epub Ahead of Print, In-Process & Other Non-Indexed Citations (via Ovid) was searched, and details of the additional searches are provided in Appendix 4.

4.2 Eligibility criteria

The review inclusion criteria are presented in Table 2.

Studies that did not report any outcomes that the EAG considered were relevant to the DTA or the clinical impact of MRI-based technologies were excluded from the review. Studies that did not include original data (i.e., reviews, editorials and opinion papers), case reports and non-English language studies were excluded from the review. Abstracts and manufacturer data were only included if they provided numerical data and sufficient methodological detail to enable assessment of study quality/risk of bias. Further, only outcome data that had not been reported in peer-reviewed full-text papers were extracted from abstracts and manufacturer reports.

Table 2 Review inclusion criteria

Parameter	Final scope[23] issued by NICE		
Population	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed: • who have indeterminate results from fibrosis testing • for whom TE or ARFI is unsuitable • who have discordant results from fibrosis testing		
Setting	Secondary and tertiary care		
Interventions	MRI-based technologies, i.e., LiverMultiScan and MRE		
	Diagnostic test accuracy	**Clinical impact**	
Comparator	LiverMultiScan versus MRE or versus no comparator MRE versus no comparator	No further testing	
Reference standard	Liver biopsy performed and interpreted by a trained healthcare professional	Not applicable	
Outcomes	Test accuracy for: • fibrosis • inflammation • steatosis	Intermediate outcomes: • impact of test result on clinical decision making (such as whether a biopsy is done, frequency of subsequent monitoring, lifestyle advice or intervention offered) • prognostic ability (for example, to predict progression of fibrosis or clinical outcomes) • number of liver biopsies • uptake and maintenance of lifestyle modifications • time to receive test results • time to diagnosis • test failure rate • reduction or remission of liver fibrosis or fibro-inflammation • reduction or remission of liver fat	
		Clinical outcomes: • mortality • morbidity (can be liver-related and non-liver related, and including from complications related to liver biopsy	Patient-reported outcomes: • acceptability of different testing modalities • health-related quality of life
Study design	Diagnostic cross-sectional and case-control studies	RCTs, cross-sectional, case-control/cohort studies & uncontrolled single arm studies	

ARFI=acoustic radiation force impulse; MRE=magnetic resonance elastography; MRI=magnetic resonance imaging; NAFLD=non-alcoholic fatty liver disease; RCT=randomised controlled trial; TE=transient elastography
Source: Final scope[23] issued by NICE

4.3 Study selection

Titles and abstracts identified by the electronic searches were uploaded to Covidence and screened by two reviewers (RB and KE). Full-text articles of any titles and abstracts that were considered potentially eligible for inclusion were obtained via online resources or through the University of Liverpool libraries and uploaded to Covidence. These full-text articles were assessed for inclusion by two reviewers (RB and KE) using the eligibility criteria outlined in Table 2. Discrepancies at each stage of screening were resolved via discussion. Full-text articles that did not meet the inclusion criteria were excluded with reasons for exclusions noted. The reference lists of relevant systematic reviews and eligible studies were hand-searched to identify further potentially relevant studies.

4.4 Data extraction

A data extraction form was designed, piloted and finalised to facilitate standardised data extraction. Data on study and patient characteristics and results were extracted by one reviewer (RB) and independently checked for accuracy by a second reviewer (KE). Any disagreements were resolved through discussion and, if necessary, in consultation with a third reviewer (SN). The manufacturers of the index tests and the corresponding authors of eligible studies were contacted and asked to provide missing data or clarify published data, and to submit individual participant data that would allow the EAG to carry out analyses for the three subgroups identified in the final scope[23] issued by NICE.

4.5 Quality assessment

The methodological quality of DTA studies was assessed using the QUality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[49] The QUADAS-2 tool considers four domains: patient selection, index test(s), reference standard and flow of patients through the study and timing of the tests. Randomised controlled trials (RCTs) evaluating the clinical impact of MRI-based technologies were assessed using the Cochrane Risk of Bias 2.0 tool.[50] National Institute of Health (NIH) study quality assessment tools[51] for cohort studies, case-control studies and before-after (pre-post) studies with no control group were used to assess risk of bias of included non-randomised studies. Qualitative studies were assessed using the Critical Appraisal Skills Programme (CASP) qualitative studies checklist.[52] Quality assessment of the included studies was undertaken by one reviewer (RB) and independently checked by a second reviewer (KE). Any disagreements were resolved by discussion and, if necessary, in consultation with a third reviewer (RD).

4.6 Methods of analysis/synthesis of DTA studies

It was not necessary or possible to use all methods of analysis described in the EAG protocol for this assessment; for details of the methods not used, see Appendix 5.

4.6.1 Statistical analysis and data synthesis

Individual study results

The EAG summarised the sensitivity and specificity of each index test presented in the included DTA studies using forest plots.

Meta-analysis

Where at least three studies provided both sensitivity and specificity data for a specific combination of index test, diagnosis of interest, and cut-off value, the EAG considered performing a bivariate random-effects meta-analysis to provide pooled estimates of sensitivity and specificity. The EAG did not perform bivariate meta-analyses where statistical heterogeneity between the studies (assessed by visually examining forest plots) was so great that pooled estimates of sensitivity and specificity would have been meaningless. The bivariate model was fitted using the meqrlogit command in Stata version 14.

Where at least three studies provided both sensitivity and specificity data for a specific combination of index test and diagnosis of interest, but used different cut-off values for the index test, the EAG used a hierarchical model to estimate a summary receiver operating characteristic (ROC) curve. The hierarchical model was fitted using the nlmixed procedure in SAS version 9.

4.6.2 Subgroup analyses and sensitivity analyses

No subgroup analyses or sensitivity analyses were performed by the EAG (see Appendix 5 for further details).

4.7 Methods of analysis/synthesis of clinical impact studies

It was not necessary or possible to use all methods of analysis described in the EAG protocol for this assessment; for details of the methods not used, see Appendix 5.

Where it was possible and clinically meaningful to perform meta-analysis, the EAG decided whether to use fixed-effects or random-effects models based on the extent of heterogeneity present between the included studies. Clinical and methodological heterogeneity between the included studies was assessed by considering differences in (a) study population, (b) interventions, (c) outcome measures, (d) study quality, and (e) study design. An assessment

of statistical heterogeneity was performed by visually examining forest plots and by considering the I^2 statistic.

Binary data were presented as frequencies and proportions, and were pooled in meta-analyses using the metaprop command in Stata version 14. Pooled proportions with 95% CIs were presented.

Where it was not possible or clinically meaningful to perform meta-analysis, the EAG reported clinical impact/intermediate outcome data narratively.

5 RESULTS OF THE ASSESSMENT OF DIAGNOSTIC TEST ACCURACY AND CLINICAL IMPACT

5.1 EAG study selection process

The EAG's searches of the electronic databases, and reference lists of relevant studies and systematic reviews identified 4489 records. After the removal of duplicate records, 3331 potential records remained. Following initial screening of titles and abstracts, 48 records were considered to be potentially relevant and were retrieved to allow assessment of the full-text publications. Studies excluded at the full-text paper screening stage and the reasons for exclusion are presented in Appendix 6.

The EAG PRISMA[48] flow diagram detailing the review screening process is shown in Figure 2.

Studies identified by the manufacturers

The test manufacturers' evidence submissions included details of studies that were potentially relevant, and should be considered, for inclusion in the EAG review. All the studies suggested by the manufacturers had already been identified by the EAG searches. The studies identified by the manufacturers that were not included in the EAG review are listed in Appendix 7 with reasons for exclusion.

5.2 Studies included in the EAG review

Thirteen studies[29,53-64] reported in 15 publications[29,31,53-65] were included in the DTA review. Two studies[29,59] reported in four publications[29,31,59,65] were evaluations of LiverMultiScan and 10 studies[53-55,57,58,60-64] were evaluations of MRE. One study[56] was an evaluation of LiverMultiScan and MRE.

Eleven studies[29,53,54,57,59,62,64,66-69] reported in 14 publications[29,31,33,53,54,57,59,62,64-69] were included in the clinical impact review of MRI-based technologies. Five studies[29,59,66,68,69] reported in eight publications[29,31,33,59,65,66,68,69] evaluated the clinical impact outcomes associated with LiverMultiScan and six studies[53,54,57,62,64,67] were evaluations of the clinical impact of MRE.

All of the studies included in the DTA review[29,53-64] and ten of the 11 studies included in the clinical impact review[29,53,54,57,59,62,64,66-68] considered patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, only one study[29] provided DTA and clinical impact results for patients with NAFLD who had indeterminate or discordant results from fibrosis testing. One study included in the clinical impact review[69] included patients with NAFLD, however, diagnoses were self-reported by the patients and it is unknown whether patients had previously been diagnosed with advanced fibrosis or cirrhosis.

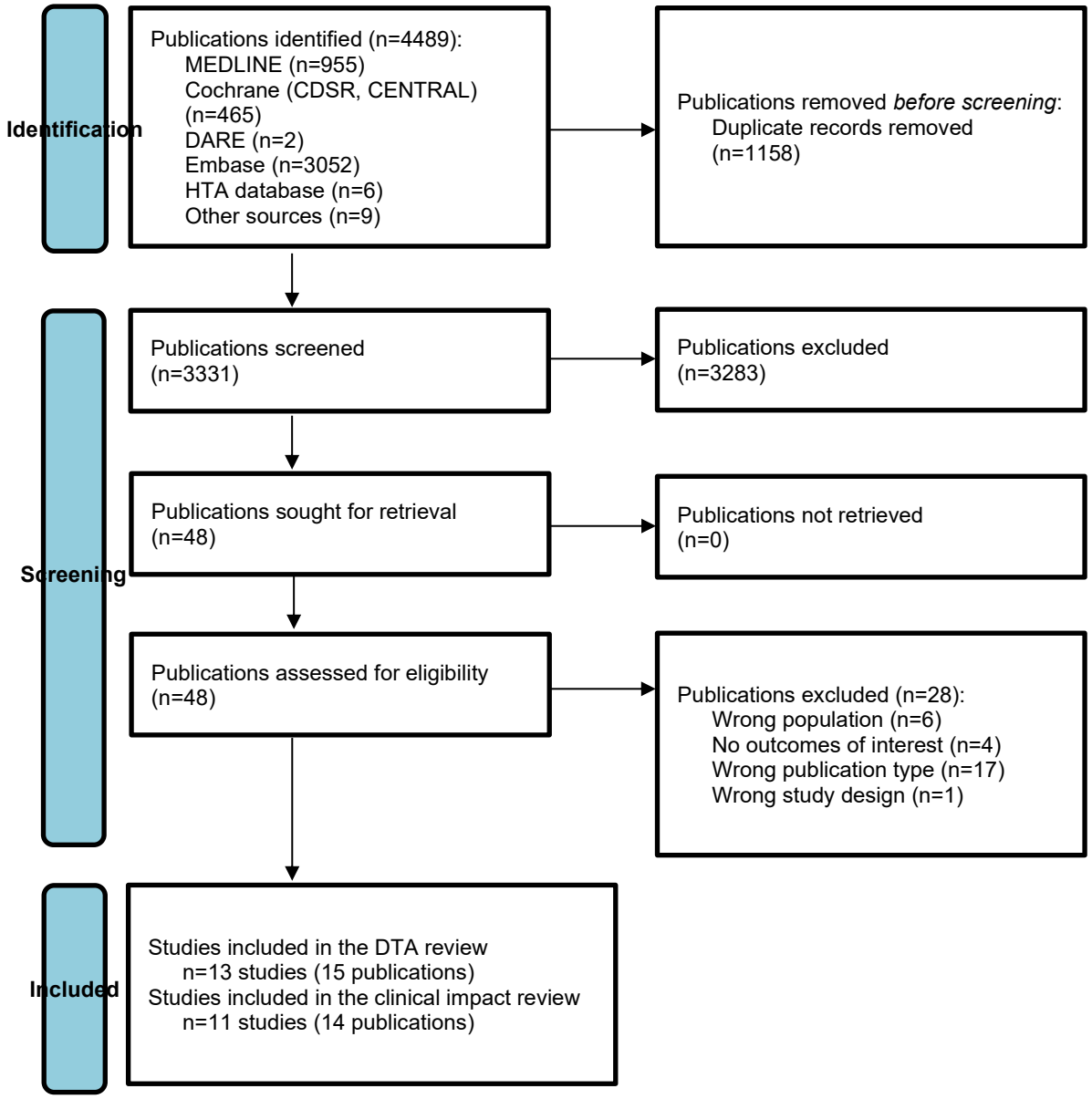

Figure 2 PRISMA flow diagram

DTA=diagnostic test accuracy
Total number of studies included in the review n=17 studies (20 publications)

5.3 Assessment of diagnostic test accuracy

5.3.1 Quality assessment

The included studies that provided DTA[29,53-64] data were assessed for risk of bias using the QUADAS-2 tool.[49] A summary of the results of the assessment using the QUADAS-2 tool is presented in Table 3. The EAG's full assessment is presented in Appendix 8.

Risk of bias

Only one study[53] was judged to have low risk of bias across all domains. One study[64] was judged as having unclear risk of bias for the patient selection domain because there was a lack of information regarding patient recruitment methods and eligibility criteria applied. One study[54] was judged to have a high risk of bias in the index test domain; this study[54] used cut-offs that were not pre-specified and it was unclear whether the index test results were interpreted without knowledge of the results of the reference standard (i.e., liver biopsy). The studies[29,55,57-64] judged as having unclear risk of bias in the index test domain did not use pre-specified thresholds but the index test results were interpreted without knowledge of the results of the reference standard. Four studies[54-56,60] were considered to have unclear risk of bias in the reference standard domain due to not providing details on whether the interpretation of the reference standard results occurred without knowledge of the index test results. Clinical advice to the EAG is that the reference standard would be likely to correctly classify the level of fibrosis; however, with all studies there is a risk of sampling error which means the reference standard may potentially incorrectly classify the condition. Two studies[55,57] were judged to have unclear risk of bias in the flow and timing domain; in one study,[57] the reference standard was performed up to 1 year after the index test and in the other study[55] not all the patients received a liver biopsy.

Applicability concerns

Only one study[29] raised no concerns regarding the applicability of the study population or the index test to the review. The Eddowes 2018[29] study recruited patients who were scheduled for non-targeted liver biopsy to (i) stage fibrosis after inconclusive non-invasive assessment of fibrosis or (ii) make a diagnosis after a range of non-invasive tests had not confirmed a diagnosis. Therefore, the EAG considers that the Eddowes 2018[29] study population is the most relevant to this assessment.

There were concerns regarding the applicability of the study population in six studies.[53,56-59,62] Although these studies[53,56-59,62] included patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed, these were not patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from

fibrosis testing. There were high risk of concerns regarding the applicability of the study population in the remaining six studies[54,55,60,61,63,64] due to the inclusion of patients with other liver disease aetiologies; the authors of these studies did not report or, when requested, provide data specifically for the subpopulation of patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed. Furthermore, it is unclear whether these studies[54,55,60,61,63,64] included patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.

There was a high risk of concern regarding the applicability of the index test in three studies[54,61,62] evaluating MRE. In the Resoundant, Inc. response to the EAG request for information,[70] Resoundant, Inc. highlighted that the Forsgren 2020[54] and the Troelstra 2021[62] studies used an investigational MRE design and not the Resoundant, Inc. MRE platform that is commercially available. The EAG notes that the Troelstra 2021[62] study used two moduli to calculate liver stiffness measurements, the MRE G' shear modulus and the MRE G" loss modulus, and presented data for the two outputs separately throughout the publication. Resoundant, Inc. considers that the data generated by the Toguchi 2017[61] study may not be representative of MRE in clinical practice as it assessed two techniques for drawing regions of interest to calculate liver stiffness (single small round regions of interest per slice [srROIs]) and whole right lobe of the liver (free hand region of interest [fhROI]), which may not be consistent with the method used to analyse MRE in clinical practice. There were no applicability concerns related to the reference standard in any of the studies.

Table 3 QUADAS-2 assessment of DTA studies

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Caussy 2018[53]	☺	☺	☺	☺	?	☺	☺
Eddowes 2018[29]	☺	?	☺	☺	☺	☺	☺
Forsgren 2020[54]	☺	☹	?	☺	☹	☹	☺
Hoffman 2020[55]	☺	?	?	?	☹	☺	☺
Imajo 2021[56]	☺	☺	?	☺	?	☺	☺
Kim 2013[57]	☺	?	☺	?	?	☺	☺
Kim 2020[58]	☺	?	☺	☺	?	☺	☺
Pavlides 2017[59]	☺	?	☺	☺	?	☺	☺
Sofue 2020[60]	☺	?	?	☺	☹	☺	☺
Toguchi 2017[61]	☺	?	☺	☺	☹	☹	☺
Troelstra 2021[62]	☺	?	☺	☺	?	☹	☺
Trout 2018[63]	☺	?	☺	☺	☹	☺	☺
Xanthakos 2014[64]	?	?	☺	☺	☹	☺	☺

☺Low Risk ☹High Risk ? Unclear Risk
DTA=diagnostic test accuracy; QUADAS-2=QUality Assessment of Diagnostic Accuracy Studies-2

5.3.2 Characteristics of the included studies

The characteristics of the 13 studies[29,53-64] included in the DTA review are presented in Table 4.

In line with the final scope[23] issued by NICE, all the studies[29,53-64] included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, only the Eddowes study[29] recruited patients who were scheduled for non-targeted liver biopsy to (i) stage fibrosis after inconclusive non-invasive assessment of fibrosis or (ii) make a diagnosis after a range of non-invasive tests had not confirmed a diagnosis. The EAG considers that the Eddowes study[29] population provides evidence for the population of patients who have indeterminate or discordant results from fibrosis testing. However, it is unclear whether the term 'inconclusive' means indeterminate and/or discordant. The EAG notes that the patients in the study[29] were scheduled for a biopsy and therefore may not represent all patients with indeterminate and/or discordant results from previous fibrosis testing; clinical advice to the EAG is that not all patients with indeterminate and/or discordant results will have a biopsy.

Two studies[29,59] assessed the DTA of LiverMultiScan, ten studies[53-55,57,58,60-64] assessed the DTA of MRE and one study[56] assessed the DTA of LiverMultiScan and MRE. The two studies[29,59] that assessed the DTA of LiverMultiScan were based in the UK, whereas the ten studies[53-55,57,58,60-64] that assessed the DTA of MRE were based in Holland,[62] Japan,[60,61] South

Korea,[58] Sweden[54] and USA.[53,55,57,63,64] The study[56] that assessed the DTA of LiverMultiScan and MRE was based in Japan. Four of the studies[53,57-59] reported that they were conducted in tertiary care. The EAG notes that all of the included studies were conducted in hospitals and therefore considers it likely that all studies were conducted in either secondary or tertiary care settings.

According to the corresponding author, the Pavlides 2017[59] study population included the Banerjee 2014[65] study population and therefore the EAG does not regard the studies as two independent data sets [Michael Pavlides, University of Oxford, 26 November 2021, personal communication].

Six of the included studies[54,55,60,61,63,64] considered patients with liver disease aetiologies other than NAFLD and did not report or provide data upon request specifically for the subpopulation of patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed. Three of the included studies[29,53,57] exclusively considered patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed. However, one of the studies[53] did not report any outcomes of interest and did not provide additional data upon request. For the remaining studies,[56,58,59,62] the EAG obtained data for patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed (see also Table 5) As a result, the EAG quantitative synthesis includes data from only six of the identified studies.[29,56-59,62]

Table 4 Characteristics of studies included in the diagnostic test accuracy review

Study	Study design; country; setting; timeframe	Population; number in analysis and recruitment details	Age (years); Male (n, %); BMI (kg/m^2); T2D (n, %)	Interpreter of index test	Interpreter of liver biopsy
LiverMultiScan					
Eddowes 2018[29]	Prospective cross-sectional; UK; NR; Feb 2014 to Sept 2015	Patients with NAFLD who had indeterminate or discordant results from fibrosis testing (N=46); recruited patients with NAFLD scheduled to undergo clinically indicated liver biopsy	Median age (range): 54 (18 to 73) Male: 28 (56) Mean BMI±SD: 33.6±5.1 T2D: 26 (52)	Analysed by a blinded operator	Assessed by blinded experienced academic liver histopathologists according to the NASH-CRN scoring system
Pavlides 2017[59]	Prospective cross-sectional; UK; tertiary care; May 2011 to Mar 2015	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed; n=48; recruited patients with suspected or known NAFLD within 1 month of liver biopsy (N=71)	Mean age±SD: 54.4±12.2 Male: 35 (72.9) *Median BMI (IQR): 32.7 (28.1 to 38.1) *T2D: 25/71 (35)	Analysed by a blinded operator	Assessed by two blinded experienced liver pathologists using the FLIP algorithm and discussed in a clinic-pathological meeting before a final Consensus report was issued
MRE					
Caussy 2018[53]	Prospective cross-sectional; USA (UCSD and Mayo Clinic); tertiary care; **USCD:** Oct 2011 to Jan 2017; **Mayo clinic:** Mar 2010 to May 2013	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed; **USCD:** N=119; **Mayo clinic:** N=75; recruited from patients with suspected NAFLD who underwent contemporaneous MRE, TE and liver biopsy	**USCD:** Mean age ± SD: 49.8±14.5 Male: 54 (45.4) Mean BMI ± SD: 30.6±5.1 T2D: 44 (37.0) **Mayo clinic:** Mean age ± SD: 47.7±11.5 Male: 25 (33.3) Mean BMI ± SD: 41.7±7.1 T2D: NR	**USCD:** Interpreted by trained image analyst (>6 months of experience with MRE analysis) **Mayo clinic:** Analysed by two experienced readers (11 years; 7 years)	**USCD:** Assessed by a blinded experienced liver pathologist according to the NASH-CRN scoring system **Mayo clinic:** First assessed by staff hepatopathologists in clinical practice according to the Brunt classification and later by an independent blinded hepatopathologist
Forsgren 2020[54]	Prospective cross-sectional; Sweden; NR; 2007 to 2014	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=34/90); recruited from patients scheduled to undergo clinically indicated liver biopsy due to elevated liver enzyme levels	*Median age (range): 52.5 (20 to 81) Male: 49 (54.4) Median BMI (range): 26.4 (19.6 to 35.9) T2D: 18 (20)	ROIs were drawn by an experienced radiologist and were interpreted by two experienced radiologists. The authors did not state whether the radiologists were blinded	Assessed by an experienced histopathologist according to the Batts and Ludwig system. The authors did not state whether the histopathologist was blinded

Hoffman 2020[55]	Retrospective cross-sectional; USA; NR; June 2018 to Sept 2018	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=61/226); recruited from patients with known or suspected hepatic fibrosis who underwent MRE	*Median age (range): 39 (20 to 80) Male: 114 (50.4) BMI: NR T2D: NR	Interpreted by two blinded readers (9 years of experience post fellowship in abdominal imaging; body MRI fellow)	Assessed by a pathologist according to the METAVIR scoring system. The authors did not state whether the pathologist was blinded
Kim 2013[57]	Retrospective cross-sectional; USA; tertiary care; Jan 2007 to Sep 2010	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=142); patients were identified by searching a MRE database for patients who had undergone MRE	Mean age ± SD: 52.8±12.8 Male: 38 (26.8) Mean BMI ± SD: 36.3±7.4 T2D: 39 (27.5)	Interpreted by staff abdominal radiologists	Assessed by blinded hepatopathologists according to the NASH-CRN scoring system
Kim 2020[58]	Prospective cross-sectional; South Korea; tertiary care; Oct 2016 to June 2017	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (N=47); recruited from patients with suspected NASH who were scheduled to undergo or underwent liver biopsy within 2 months (unclear if from recruitment or from MRE)	Mean age ± SD: 51.0±12.7 Male: 16 (34.0) Mean BMI ± SD: 28.3±6.2 T2D: NR	ROIs were drawn and interpreted by two blinded board-certified radiologists (25 years; 6 years of abdominal radiology experience)	Assessed by a blinded pathologist with >15 years of experience according to the NASH-CRN scoring system
Sofue 2020[60]	Retrospective cross-sectional; Japan; NR; 6 month study period but dates NR	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=8/30); recruited from patients with chronic liver disease who underwent MRE at 60Hz and 80Hz vibration frequencies and liver biopsy within 2 months	*Mean age ± SD (range): 61.5±11.5 (39 to 82) Male: 14 (46.7) Mean BMI ± SD (range): 23.9 ± 3.3 (16.2 to 34.5) T2D: NR	Interpreted by a blinded board-certified abdominal radiologist (22 years of experience in abdominal imaging)	Assessed by two pathologists by consensus (12 and 30 years of experience, respectively). The authors did not state whether the pathologists were blinded
Toguchi 2017[61]	Retrospective cross-sectional; Japan; NR; Oct 2013 to Jan 2015	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=23/51); recruited from patients with chronic liver disease who had undergone MRE and TE	*Mean age: 59.9 Male: 21 (41.2) BMI: NR T2D: NR	Interpreted by a blinded radiologist with 8 years of clinical experience	Assessed by three blinded hepatopathologists according to the METAVIR scoring system
Troelstra 2021[62]	Prospective cross-sectional; Holland; NR; Sept 2018 to Oct 2020	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed; N=37; recruited from patients with an	Mean age ± SD: 49.0±13.2 Male: 23 (62.2) Mean BMI ± SD: 33.2±3.8 T2D: 16 (43.2)	NR	Assessed by a blinded hepatopathologist with 15 years of experience according to the SAF score and NASH-CRN scoring system

		incidental finding of hepatic steatosis on abdominal ultrasound			
Trout 2018[63]	Prospective cross-sectional; USA; NR; Jan 2012 to Sept 2016	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=48/86); patients were identified by searching radiology department records for patients who had undergone MRE and liver biopsy	*Median age (range): 14.2 (0.3 to 20.6) Male: 49 (57.0) BMI: NR T2D: NR	Re-interpreted by a blinded MR physicist with 8 years of MRE experience	Re-assessed by a blinded board-certified pathologist with 10 years of experience according to the NASH-CRN scoring system
Xanthakos 2014[64]	Prospective cross-sectional; USA; NR; Aug 2011 to Dec 2012	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=27/35); recruited from patients with chronic liver disease who underwent MRE and liver biopsy	*Median age (IQR): 13 (12 to 16) Male: 28 (51.4) Median BMI (IQR): 33.9 (28.9 to 38.2) T2D: NR	NR	NR
LiverMultiScan and MRE					
Imajo 2021[56]	Prospective cross-sectional; Japan; NR; Jan 2019 to Feb 2020	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed; N=143; recruited patients with suspected NASH scheduled to undergo clinically indicated liver biopsy	Mean age ± SD: 60.2±13.1 Male: 88 (60.7) Mean BMI ± SD 28.8±4.7 **Diabetic: 97 (66.9)	mpMRI data were analysed using LiverMultiScan software by blinded off-site image analysts. MRE images were analysed by abdominal radiologists. The authors did not state whether the abdominal radiologists were blinded	Assessed by three independent histopathologists, one at the time of collection and later by two pathologists using digitalised biopsy slides according to the NASH-CRN scoring system. The paper did not state whether the pathologists were blinded

*The statistics reported are based on the entire study population and not for the subpopulation of patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed
**Does not specify type of diabetes
BMI=body mass index; CRN=Clinical Research Network; FLIP=Fatty Liver Inhibition of Progression; IQR=interquartile range; METAVIR=meta-analysis of histological data in viral hepatitis; mpMRI=multiparametric magnetic resonance imaging; MR=magnetic resonance; MRE=magnetic resonance elastography; NAFLD=non-alcoholic fatty liver disease; NASH=non-alcoholic steatohepatitis; NR=not reported; ROI=region of interest; SAF=steatosis, activity, fibrosis score; SD=standard deviation; T2D=type 2 diabetes; TE=transient elastography; UCSD=University of California at San Diego

5.3.3 Diagnostic test accuracy results

The absolute numbers of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) LiverMultiScan or MRE test results compared to the reference standard of liver biopsy (i.e., 2x2 data) were not presented in any of the included studies. We contacted the authors of all included studies and the test manufacturers to request these data.

Perspectum provided 2x2 data in response to the EAG request for information for the three LiverMultiScan studies[29,56,59] included in the DTA review. The authors of the Troelstra 2021[62] study of MRE provided 2x2 data in response to the EAG request. Data from the Kim 2020[58] study were obtained from a systematic review, and 2x2 data from the Kim 2013[57] study were calculated using the number of patients with and without the diagnosis of interest, and the estimates of sensitivity and specificity reported in the published paper. The full set of data sources is provided in Table 5.

Table 5 Data sources for 2x2 diagnostic test accuracy data

Study	Data source for 2x2 data	*Data provided for population in scope[23]
Eddowes 2018[29]	Perspectum Ltd submission[71]** included 2x2 data	Yes
Imajo 2021[56]	2x2 data were provided in the Perspectum Ltd submission.[71] However, inconsistencies in the data had to be resolved through personal communication with the study authors [Marika French, Perspectum, 3 February 2022]; data provided by the study authors were used in the EAG quantitative analysis. The EAG notes that the LiverMultiScan PDFF output, the LiverMultiScan cT1 output and the MRE test 2x2 data for diagnosis of steatosis and fibrosis provided by the Imajo 2021[56] study authors do not correspond to the numbers of patients with and without these diagnoses reported in Table 2 of the published paper;[56] the EAG was unable to clarify reasons for these discrepancies with the authors of the published paper.[56] The EAG also notes that data for advanced fibrosis (≥F3) were only available for LiverMultiScan tests and not for the MRE test	No
Kim 2013[57]	The EAG calculated 2x2 data using the number of patients with and without fibrosis (≥F3) and the estimates of sensitivity and specificity reported in the published paper	No
Kim 2020[58]	2x2 data were provided in Figure S7, S10 and S14 from the Selvaraj systematic review[72]	No
Pavlides 2017[59]	2x2 data (n=28) were provided in the Perspectum submission[71] and the EAG received IPD (n=48) from the study author [Michael Pavlides, University of Oxford, 9 December 2021]. The EAG used the summary 2x2 data for the quantitative analysis because the IPD used the Ishak staging system[73] to score fibrosis whereas the other included studies use the NASH CRN scoring system[17]	No
Troelstra 2021[62]	2x2 data were made available after personal communication with study authors [Marian Troelstra, Amsterdam University Medical Centers, 24 November 2022]	No

*In line with the final scope[23] issued by NICE, the population of interest consists of the three groups of patients with NAFLD for whom advanced fibrosis or cirrhosis has not yet been diagnosed, namely (i) patients with indeterminate results from fibrosis testing, (ii) patients who are unsuitable for testing with TE or ARFI and (iii) patients with discordant results from fibrosis testing.
** In this EAG report, references to the Perspectum submission[71] are to the evidence submission received by the EAG from Perspectum in response to the EAG request for information.
EAG=External Assessment Group; IPD=individual patient data

The EAG's quantitative synthesis therefore included data from six[29,56-59,62] (out of 13) identified studies for which 2x2 data were available.

Where studies reported 2x2 data (i.e., the number of TP, FP, TN and FN test results), data from individual studies were summarised in forest plots (Figure 3 to Figure 6) alongside estimates of sensitivity and specificity. The individual study results were grouped by diagnosis of interest, and the cut-off value used to indicate a positive result from the index test was also provided.

Where studies reported area under the receiver operating characteristic (AUROC) curve results, these results are summarised in Appendix 9.

Individual study results: LiverMultiScan

For the LiverMultiScan PDFF and LiverMultiScan cT1 outputs (Section 3.6), 2x2 data were available from three studies[29,56,59] as shown in Figure 3 to Figure 5. Diagnosis definitions and cut-off values used to indicate a positive result from the index test were consistent between these studies, and it was therefore possible to draw comparisons between the individual study results. As previously discussed in Section 5.3.2 of this EAG report, the EAG considers that the Eddowes 2018[29] study is the most relevant study to this assessment.

For diagnosis of fibrosis, sensitivity and specificity values for the tests used in the Eddowes 2018[29] study (as reported in the Perspectum Ltd submission[71]) were consistently higher for LiverMultiScan cT1 than for LiverMultiScan PDFF. For LiverMultiScan PDFF, as fibrosis stage increased, sensitivity decreased (≥F1, 80%; ≥F2, 57%; ≥F3, 50%) and specificity decreased or remained the same (≥F1, 50%; ≥F2, 50%; ≥F3, 42%). For LiverMultiScan cT1, as fibrosis stage increased, sensitivity decreased or remained similar (≥F1, 88%; ≥F2, 63%; ≥F3, 64%) and there was no clear pattern to the change in specificity values, with the highest specificity value being reported for fibrosis ≥F2 (≥F1, 67%; ≥F2, 75%; ≥F3, 63%).

For diagnosis of steatosis, sensitivity and specificity values for the outputs used in the Eddowes 2018[29] study were similar between LiverMultiScan cT1 and LiverMultiScan PDFF. The EAG notes that specificity was reported to be 0% for steatosis (Brunt grade ≥1) in the Eddowes 2018[29] study for both LiverMultiScan PDFF and LiverMultiScan cT1, i.e., neither of the outputs was able to correctly identify any patients as not having steatosis (number of true negatives=0). However, this result is highly uncertain (95% CI: 0% to 97%), as it was calculated using data from one patient for whom the reference standard reported a negative result. For the LiverMultiScan PDFF output, the opposite finding was reported by the other two studies,[56,59] i.e., all non-steatosis patients were correctly identified as not having steatosis

(specificity=100%); these results were also based on a small number of true non-steatosis patients (Imajo 2021[56] study: n=7; Pavlides 2017[59] study: n=2). This was the most extreme case of heterogeneity observed between results from the three studies[29,56,59] that assessed the DTA of LiverMultiScan.

For the diagnosis of NASH and advanced NASH, sensitivity was estimated to be 64% in the Eddowes 2018[29] study for both LiverMultiScan PDFF and LiverMultiScan cT1. There was some variation in the specificity estimates from this study for NASH (LiverMultiScan PDFF, 57%; LiverMultiScan cT1, 67%) and advanced NASH (LiverMultiScan PDFF, 54%; LiverMultiScan cT1, 63%).

Fibrosis (≥F1)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	32	3	8	3	46	5%	0.80 [0.64, 0.91]	0.50 [0.12, 0.88]
Imajo 2021	121	5	17	0	143	5%	0.88 [0.81, 0.93]	0.00 [0.00, 0.52]
Pavlides 2017	18	0	10	0	28	5%	0.64 [0.44, 0.81]	Not estimable

Fibrosis (≥F2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	17	8	13	8	46	10%	0.57 [0.37, 0.75]	0.50 [0.25, 0.75]
Imajo 2021	56	20	54	13	143	10%	0.51 [0.41, 0.61]	0.39 [0.23, 0.58]
Pavlides 2017	5	3	14	6	28	10%	0.26 [0.09, 0.51]	0.67 [0.30, 0.93]

Fibrosis (≥F3)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	11	14	11	10	46	10%	0.50 [0.28, 0.72]	0.42 [0.22, 0.63]
Imajo 2021	35	41	43	24	143	10%	0.45 [0.34, 0.57]	0.37 [0.25, 0.50]
Pavlides 2017	2	6	11	9	28	10%	0.15 [0.02, 0.45]	0.60 [0.32, 0.84]

Steatosis (Brunt grade ≥ 1)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	34	1	11	0	46	5%	0.76 [0.60, 0.87]	0.00 [0.00, 0.97]
Imajo 2021	126	0	10	7	143	5%	0.93 [0.87, 0.96]	1.00 [0.59, 1.00]
Pavlides 2017	18	0	8	2	28	5%	0.69 [0.48, 0.86]	1.00 [0.16, 1.00]

Steatosis (Brunt grade ≥ 2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	18	7	5	16	46	10%	0.78 [0.56, 0.93]	0.70 [0.47, 0.87]
Imajo 2021	59	17	10	57	143	10%	0.86 [0.75, 0.93]	0.77 [0.66, 0.86]
Pavlides 2017	7	1	11	9	28	10%	0.39 [0.17, 0.64]	0.90 [0.55, 1.00]

NASH (NAS ≥ 4 with ≥ 1 hepatocyte ballooning and ≥ 1 lobular inflammation)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	16	9	9	12	46	10%	0.64 [0.43, 0.82]	0.57 [0.34, 0.78]
Imajo 2021	59	17	22	45	143	10%	0.73 [0.62, 0.82]	0.73 [0.60, 0.83]
Pavlides 2017	3	5	9	11	28	10%	0.25 [0.05, 0.57]	0.69 [0.41, 0.89]

Advanced NASH (NAS ≥ 4 with fibrosis ≥ 2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	14	11	8	13	46	10%	0.64 [0.41, 0.83]	0.54 [0.33, 0.74]
Imajo 2021	48	28	20	47	143	10%	0.71 [0.58, 0.81]	0.63 [0.51, 0.74]
Pavlides 2017	1	7	9	11	28	10%	0.10 [0.00, 0.45]	0.61 [0.36, 0.83]

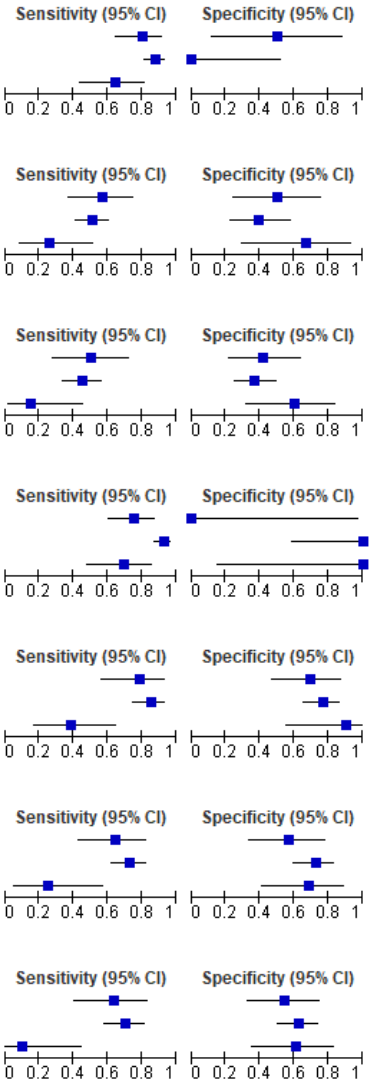

Figure 3 Forest plot displaying 2x2 data, sensitivity and specificity for LiverMultiScan PDFF from the included studies

CI=confidence interval; FN=false negative; FP=false positive; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction; TN=true negative; TP=true positive
Source: EAG report, Table 5

Fibrosis (≥ F1)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	35	2	5	4	46	800ms	0.88 [0.73, 0.96]	0.67 [0.22, 0.96]
Imajo 2021	105	2	33	3	143	800ms	0.76 [0.68, 0.83]	0.60 [0.15, 0.95]
Pavlides 2017	22	0	6	0	28	800ms	0.79 [0.59, 0.92]	Not estimable

Fibrosis (≥ F2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	19	4	11	12	46	875ms	0.63 [0.44, 0.80]	0.75 [0.48, 0.93]
Imajo 2021	56	11	54	22	143	875ms	0.51 [0.41, 0.61]	0.67 [0.48, 0.82]
Pavlides 2017	11	3	8	6	28	875ms	0.58 [0.33, 0.80]	0.67 [0.30, 0.93]

Fibrosis (≥ F3)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	14	9	8	15	46	875ms	0.64 [0.41, 0.83]	0.63 [0.41, 0.81]
Imajo 2021	45	22	33	43	143	875ms	0.58 [0.46, 0.69]	0.66 [0.53, 0.77]
Pavlides 2017	9	5	4	10	28	875ms	0.69 [0.39, 0.91]	0.67 [0.38, 0.88]

Steatosis (Brunt grade ≥ 1)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	36	1	9	0	46	800ms	0.80 [0.65, 0.90]	0.00 [0.00, 0.97]
Imajo 2021	103	4	33	3	143	800ms	0.76 [0.68, 0.83]	0.43 [0.10, 0.82]
Pavlides 2017	21	1	5	1	28	800ms	0.81 [0.61, 0.93]	0.50 [0.01, 0.99]

Steatosis (Brunt grade ≥ 2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	16	7	7	16	46	875ms	0.70 [0.47, 0.87]	0.70 [0.47, 0.87]
Imajo 2021	44	23	25	51	143	875ms	0.64 [0.51, 0.75]	0.69 [0.57, 0.79]
Pavlides 2017	14	0	4	10	28	875ms	0.78 [0.52, 0.94]	1.00 [0.69, 1.00]

NASH (NAS ≥ 4 with ≥ 1 hepatocyte ballooning and ≥ 1 lobular inflammation)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Imajo 2021*	73	34	8	28	143	800ms	0.90 [0.81, 0.96]	0.45 [0.32, 0.58]
Eddowes 2018	16	7	9	14	46	875ms	0.64 [0.43, 0.82]	0.67 [0.43, 0.85]
Imajo 2021	52	15	29	47	143	875ms	0.64 [0.53, 0.75]	0.76 [0.63, 0.86]
Pavlides 2017	10	4	2	12	28	875ms	0.83 [0.52, 0.98]	0.75 [0.48, 0.93]

Advanced NASH (NAS ≥ 4 with fibrosis ≥ 2)

Study	TP	FP	FN	TN	Total N	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)
Eddowes 2018	14	9	8	15	46	875ms	0.64 [0.41, 0.83]	0.63 [0.41, 0.81]
Imajo 2021	44	23	24	52	143	875ms	0.65 [0.52, 0.76]	0.69 [0.58, 0.79]
Pavlides 2017	8	6	2	12	28	875ms	0.80 [0.44, 0.97]	0.67 [0.41, 0.87]

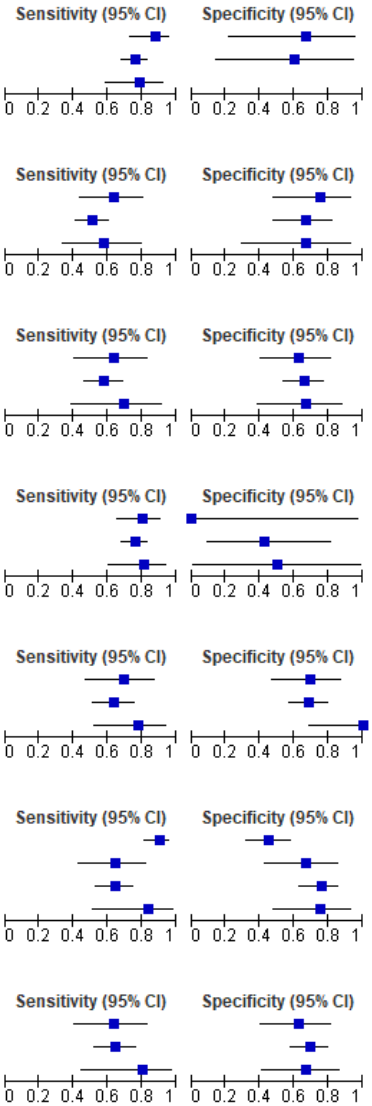

Figure 4 Forest plot displaying 2x2 data, sensitivity and specificity for LiverMultiScan cT1 from the included studies

* Data for NASH was available from the Imajo 2021[56] study for two cut-off values, 800ms and 875ms. All other studies reported data for the 875ms cut-off value only
CI=confidence interval; cT1=iron corrected longitudinal relaxation time; FN=false negative; FP=false positive; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; TN=true negative; TP=true positive
Source: EAG report, Table 5

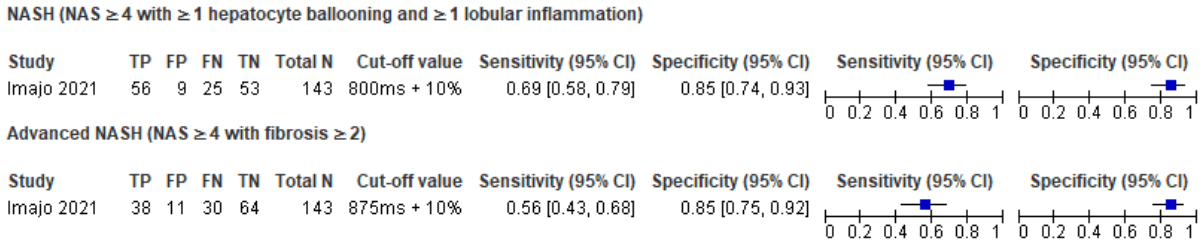

Figure 5 Forest plot displaying 2x2 data, sensitivity and specificity for LiverMultiScan PDFF and cT1 combined from the included studies

CI=confidence interval; cT1= iron corrected longitudinal relaxation time; FN=false negative; FP=false positive; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction; TN=true negative; TP=true positive
Source: EAG report, Table 5

Individual study results: MRE

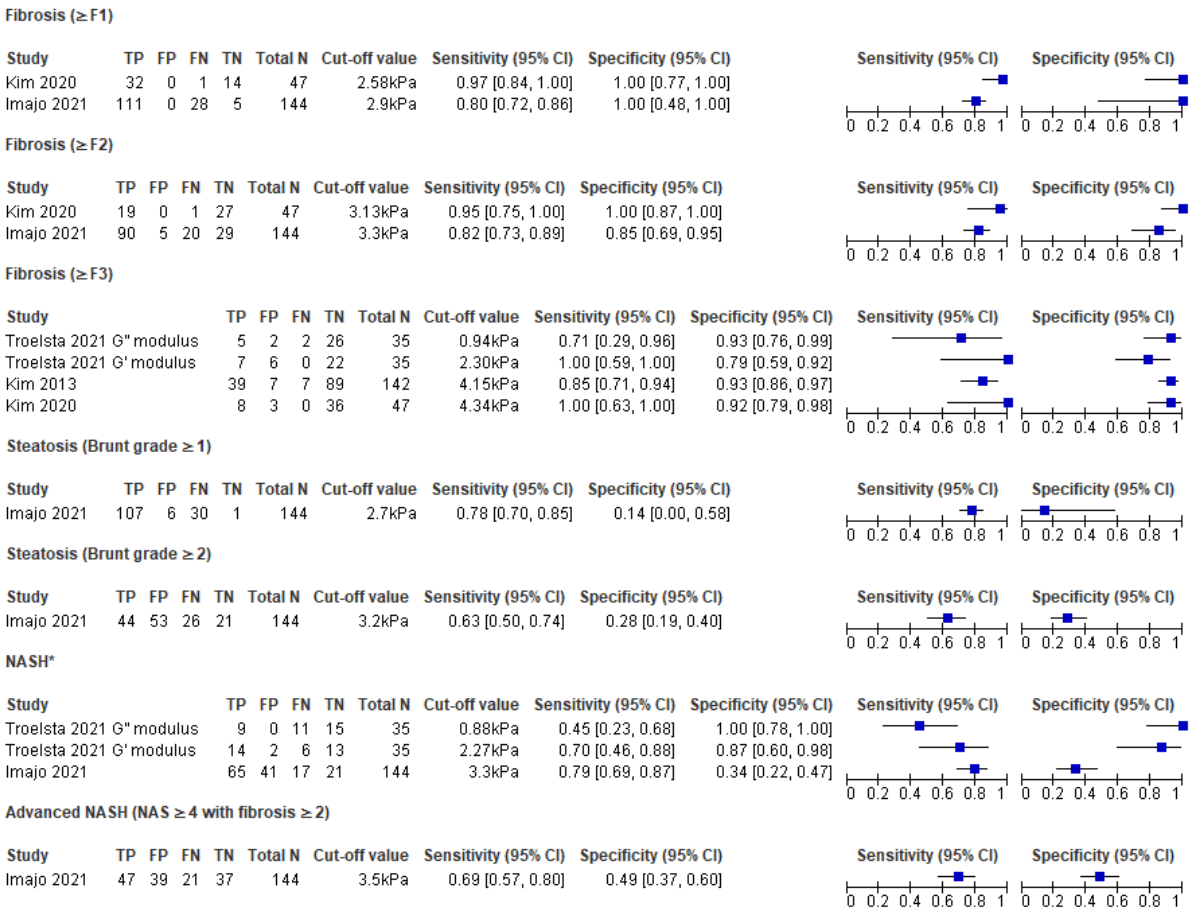

Figure 6 Forest plot displaying 2x2 data, sensitivity and specificity for MRE from the included studies

* NASH was defined in the Imajo 2021[56] study as NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation, and in the Troelstra 2021[62] study as ≥1 steatosis, ≥1 hepatocyte ballooning and ≥1 lobular inflammation
CI=confidence interval; FN=false negative; FP=false positive; MRE=magnetic resonance elastography; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; TN=true negative; TP=true positive
Source: EAG report, Table 5

For MRE, 2x2 data were available from four studies[56-58,62] as shown in Figure 6. Diagnosis definitions were consistent between studies; however, the cut-off values used to indicate a positive result from the index test varied. There were no instances of the same cut-off value being used to indicate the same diagnosis in two of the four[56-58,62] studies. It is therefore difficult to draw comparisons between the results of these four studies.[56-58,62]

Estimates of sensitivity and specificity from the Kim 2020[58] study (as reported in supplementary materials to the Selvaraj 2021[72] systematic review) were high for diagnosis of fibrosis (≥F1: sensitivity=97%, specificity=100%; ≥F2: sensitivity=95%, specificity=100%; ≥F3: sensitivity=100%, specificity=92%).

Compared with estimates from the Kim 2020[58] study, DTA estimates from the Imajo 2021[56] study (provided in communications between the study authors and the EAG) were consistent (≥F1: specificity=100%) or slightly lower (≥F1: sensitivity=80%; ≥F2: sensitivity=82%, specificity=85%); differences between the results from the two studies[56,58] could be explained by the different cut-off values used. The EAG notes that the Imajo 2021[56] study used the cut-off values that Resoundant Inc. suggested to NICE[23] should be used to stage fibrosis (Section 3.6.2). The Kim 2020[58] study calculated optimal cut-off values for fibrosis staging from ROC curve analysis which were lower than those suggested by Resoundant Inc..[23]

For advanced fibrosis (≥F3), data were provided by the authors of the Troelstra 2021[62] study for both the MRE G' shear modulus and the MRE G" loss modulus. The output reported in the other two studies[57,58] providing data for this diagnosis was the MRE complex shear modulus. Clinical advice to the EAG was that the MRE G' shear modulus results were directly comparable with the MRE complex shear modulus results.

Estimates of sensitivity and specificity for advanced fibrosis (≥F3) from the three MRE G' shear modulus (complex shear modulus) studies[57,58,62] varied. The EAG notes that the three studies[57,58,62] calculated optimal cut-off values to stage advanced fibrosis (≥F3) from ROC curve analysis. The cut-off value used by the Troelstra 2021[62] study (2.30kPa) was lower than the value that Resoundant Inc. suggested to NICE[23] should be used to stage advanced fibrosis (>3.9kPa) whereas the cut-off values used by the Kim 2013[57] study (4.15kPa) and the Kim 2020[58] study (4.34kPa) were greater. Sensitivity values were 100% for both the study which used the lowest cut-off value (Troelstra 2021,[62] cut-off value=2.30kPa) and the study that used the highest cut-off value (Kim 2020,[58] cut-off value=4.34kPa). Lower sensitivity (85%) was observed in the remaining study (Kim 2013,[57] cut-off value=4.15 kPa). Specificity was high for the two studies with the highest cut-off values (Kim 2013[57]: specificity=93%, cut-off value=4.15kPa; Kim 2020:[58] specificity=92%, cut-off value=4.34kPa), but a lower specificity

value (79%) was observed for the Troelstra 2021[62] study, which applied a lower cut-off value (2.30kPa).

As cut-off values increase, it would be expected for either sensitivity to increase while specificity decreases, or vice versa. However, this was not the case for ≥F3 data. It is important to note that sensitivity values from the Troelstra 2021[62] study and the Kim 2020[58] study were based on small numbers of patients (n=7 and n=8, respectively). It may be that a clearer pattern would emerge between cut-off values and estimates of DTA if data were available from more patients. There may also be clinical and/or methodological heterogeneity between the included studies[57,58,62] that lead to DTA estimates that do not follow the expected trend.

For the MRE G" loss modulus, estimates of test accuracy for advanced fibrosis (≥F3) from the Troelstra 2021[62] study suggested that this modulus was more specific (specificity=93%) than sensitive (sensitivity=71%).

Data for diagnosis of steatosis were only available from the Imajo 2021[56] study; DTA estimates were lower than those provided for diagnosis of fibrosis from the same study, with specificity values being particularly low (Brunt grade ≥1: sensitivity=78%, specificity=14%; Brunt grade ≥2: sensitivity=63%, specificity=28%). However, the very low specificity value (14%) observed for identifying patients without steatosis (Brunt grade≥1) was based on a very small number of patients (n=7), resulting in a wide confidence interval (0% to 58%).

Data for diagnosis of NASH were available from the Troelstra 2021[62] study (for both the MRE G' shear modulus and the MRE G" loss modulus) and the Imajo 2021[56] study. The two studies used slightly different definitions of NASH (Imajo 2021:[56] NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation; Troelstra 2021:[62] ≥1 steatosis, ≥1 hepatocyte ballooning and ≥1 lobular inflammation). For the shear modulus data, sensitivity was similar between the two studies (Imajo 2021:[56] sensitivity=79%; Troelstra 2021:[62] sensitivity=70%), whereas sensitivity was higher for the Troelstra 2021[62] study than the Imajo 2021[56] study (87% versus 34%, respectively). Differences between the results from the two studies[56,62] could be explained by the different cut-off values used. For the loss modulus, estimates of test accuracy for NASH from the Troelstra 2021[62] study suggested that this modulus was highly specific (specificity=100%), but had poor sensitivity (sensitivity=45%).

Data for diagnosis of advanced NASH were only available from the Imajo 2021[56] study. Comparing estimates of test accuracy from this study for NASH and advanced NASH, MRE was more sensitive for NASH than advanced NASH (79% versus 69%), but less specific (34% versus 49%).

Results from EAG meta-analyses: LiverMultiScan

A summary of meta-analysis results, where available, and justification for not combining results in meta-analysis, where applicable, are provided in Table 6.

It was not possible to perform meta-analysis for fibrosis (≥F1) using LiverMultiScan PDFF or LiverMultiScan cT1 data. For fibrosis (≥F2 and ≥F3), the pooled sensitivity and specificity values were higher for LiverMultiScan cT1 (≥F2: sensitivity=54.1%, specificity=69.0%; ≥F3: sensitivity=60.2%, specificity=65.4%) than for LiverMultiScan PDFF (≥F2: sensitivity=46.8%, specificity=48.6%; ≥F3: sensitivity=38.6%, specificity=43.6%).

For steatosis (Brunt grade ≥1), the EAG did not perform a meta-analysis using the LiverMultiScan PDFF data as heterogeneity between the specificity results of the included studies[29,56,59] was very large (specificity was reported to be 0% one study[29] and 100% for two studies[56,59]). The EAG considered that pooled results from a meta-analysis of these studies would be meaningless. For LiverMultiScan cT1, the meta-analysis results suggested greater sensitivity than specificity, which was particularly poor (sensitivity=77.3% 95% CI: 71.1% to 82.5%; specificity=40.0%, 95% CI: 15.8% to 70.3%).

As the level of steatosis increases (Brunt grade ≥2), results from the EAG meta-analyses suggest that LiverMultiScan cT1 output becomes more specific (specificity=72.0; 95% CI: 62.7% to 79.6%), and slightly less sensitive (sensitivity=67.3%; 95% CI: 58.0% to 75.4%). The steatosis (Brunt grade ≥2) results for LiverMultiScan PDFF (sensitivity=71.9%; 95% CI: 45.3% to 88.3%; specificity=79.0%; 95% CI: 65.4% to 88.3%) are fairly consistent with those for LiverMultiScan cT1.

For NASH and advanced NASH, estimates of DTA were broadly similar between the LiverMultiScan cT1 and LiverMultiScan PDFF outputs, with the exception of sensitivity for detecting advanced NASH (LiverMultiScan cT1: 66.0%; LiverMultiScan PDFF: 49.4%).

Table 6 Results from meta-analyses for the LiverMultiScan index tests

Diagnosis	Definition	Cut-off value	No. of studies	No. of participants	Sensitivity (%, 95% CI)*	Specificity (%, 95% CI)*
LiverMultiScan PDFF						
Fibrosis	≥F1	5%	3	217	The Pavlides 2017[59] study was excluded as it does not contribute specificity data - only two studies remaining so insufficient number of studies to perform meta-analysis	
Fibrosis	≥F2	10%	3	217	46.8 (34.1 to 59.8)	48.6 (32.5 to 65.0)
Fibrosis	≥F3	10%	3	217	38.6 (23.8 to 56.0)	43.6 (30.7 to 57.5)
Steatosis	Brunt Grade ≥1	5%	3	217	Heterogeneity is so great that it is meaningless to meta-analyse (two studies report specificity as 100% and 1 study reports specificity as 0%)	
Steatosis	Brunt Grade ≥2	10%	3	217	71.9 (45.3 to 88.3)	79.0 (65.4 to 88.3)
NASH	NAS ≥4 with at least 1 in ballooning and inflammation	10%	3	217	58.0 (35.3 to 77.8)	67.8 (56.3 to 77.4)
Advanced NASH	NAS ≥4 + fibrosis ≥2	10%	3	217	49.4 (19.1 to 80.1)	60.5 (50.1 to 70.0)
LiverMultiScan cT1						
Fibrosis	≥F1	800ms	3	217	The Pavlides 2017[59] study was excluded as it does not contribute specificity data - only two studies remaining so insufficient number of studies to perform meta-analysis	
Fibrosis	≥F2	875ms	3	217	54.1 (46.3 to 61.7)	69.0 (56.0 to 79.5)
Fibrosis	≥F3	875ms	3	217	60.2 (50.9 to 68.8)	65.4 (55.8 to 73.9)
Steatosis	Brunt Grade ≥1	800ms	3	217	77.3 (71.1 to 82.5)	40.0 (15.8 to 70.3)
Steatosis	Brunt Grade ≥2	875ms	3	217	67.3 (58.0 to 75.4)	72.0 (62.7 to 79.6)
NASH	NAS ≥4 with at least 1 in ballooning and inflammation	800ms	1	143	Insufficient number of studies to perform meta-analysis	
NASH		875ms	3	217	66.1 (57.1 to 74.1)	73.7 (64.2 to 81.5)
Advanced NASH	NAS ≥4 + fibrosis ≥2	875ms	3	217	66.0 (56.2 to 74.6)	67.5 (58.5 to 75.4)
LiverMultiScan PDFF + cT1 combined						
NASH	NAS ≥4 with at least 1 in ballooning and inflammation	800ms + 10%	1	143	Insufficient number of studies to perform meta-analysis	
Advanced NASH	NAS ≥4 + fibrosis ≥2	875ms + 10%	1	143	Insufficient number of studies to perform meta-analysis	

* Where no meta-analysis was performed, justification is provided instead of estimates of sensitivity and specificity
CI=confidence interval; cT1=iron corrected longitudinal relaxation time; NASH=non-alcoholic steatohepatitis; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; PDFF=proton density fat fraction Source: Bivariate random-effects meta-analysis performed using 2x2 data reported in Figure 3 and Figure 4

Results from EAG meta-analyses: MRE

For MRE, there was only one diagnosis (fibrosis ≥F3) where at least three studies[56-58] (224 participants) provided DTA data. For this diagnosis, data were available from the Troelstra 2021[62] study (MRE G' shear modulus and MRE G" loss modulus), the Kim 2013[57] study (complex shear modulus) and the Kim 2020[58] study (complex shear modulus). The EAG considered it appropriate to include data from the Troelstra 2021[62] study for the MRE G' shear modulus rather than for the MRE G" loss modulus in the meta-analysis; clinical advice to the EAG was that the MRE G' shear modulus results was directly comparable with the MRE complex shear modulus results. It would not have been possible to include data for both moduli from the Troelstra 2021[62] study in a meta-analysis as both data sets represented the same group of patients.

As cut-off values varied between the three studies[56-58] that reported data for this diagnosis, a summary ROC curve was estimated (Figure 7).

The summary ROC curve demonstrates how sensitivity and specificity values change as cut-off values vary between the three included studies.[57,58,62] The closer the summary ROC curve is to the top left hand corner in ROC space (where sensitivity and specificity both equal 100%), the greater the discriminatory power of the test. The summary ROC curve for an uninformative test would be the upward diagonal of the summary ROC plot (the dashed line). The summary ROC curve in Figure 7 therefore indicates high DTA. It is also important to note that the observed study results do not all lie close to the summary ROC curve; this may be due to the fact that small studies are likely to estimate values for test accuracy that are further away from the true test accuracy values than larger studies (i.e., statistical error). Two of the included studies had small sample sizes (n=35 in the Troelstra 2021[62] study and n=47 in the Kim 2020[58] study). Clinical and/or methodological heterogeneity between the included studies[57,58,62] may also explain the fact that observed study results do not all lie close to the summary ROC curve. For example, the EAG notes that the Troelstra 2021[62] study used an investigational MRE design and not the Resoundant, Inc. MRE platform that is commercially available and was used in the Kim 2013[57] and Kim 2020[58] studies. Furthermore, the studies were conducted in different countries (Kim 2013,[57] USA; Kim 2020,[58] South Korea; Troelstra 2021,[62] Holland). These differences may have introduced heterogeneity to the analysis.

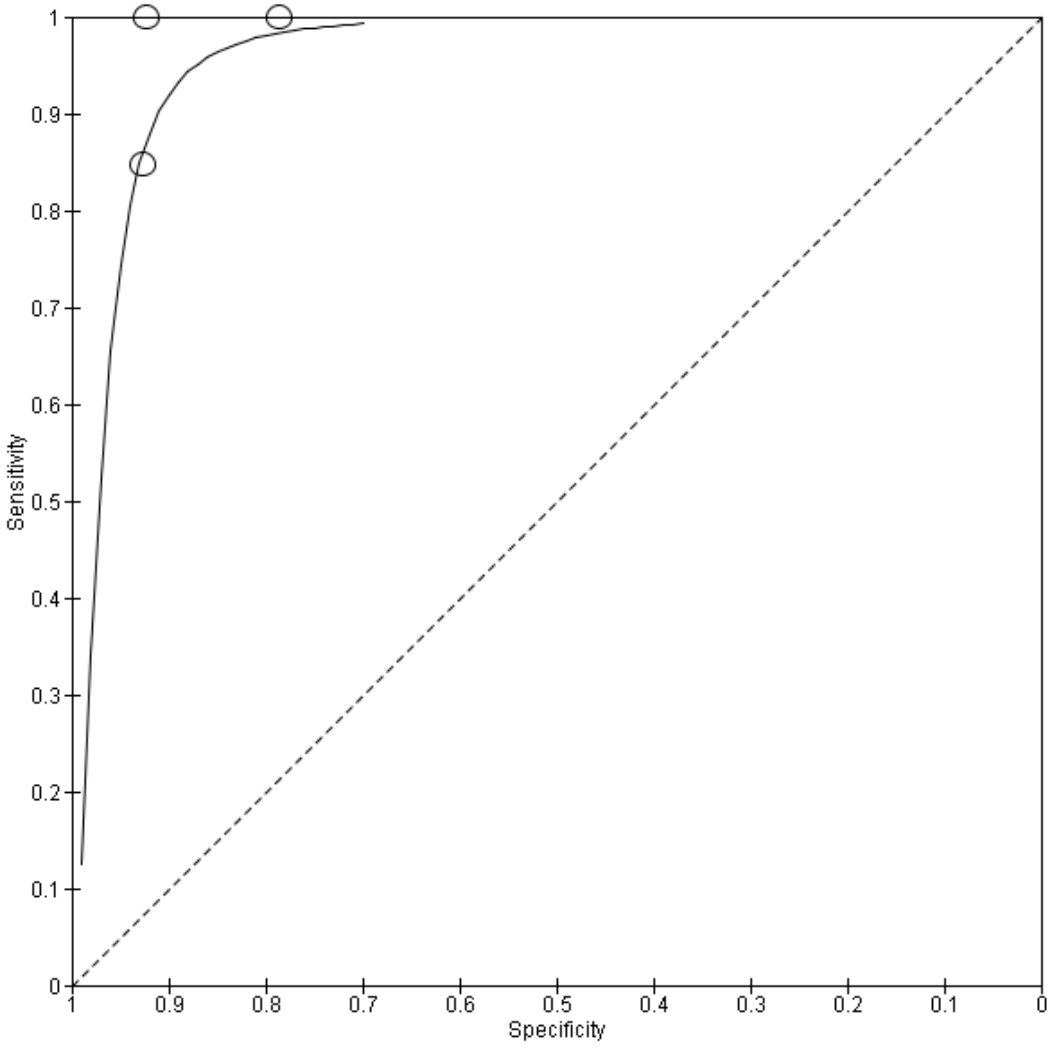

Figure 7 Summary ROC plot for fibrosis (≥F3) data from the MRE test

The solid line is the summary ROC curve. The dashed line indicates sensitivity=1-specificity (i.e. an uninformative test). The circles represent individual study results.
MRE=magnetic resonance elastography; ROC=receiver operator characteristic

5.4 Assessment of clinical impact

Eleven studies[29,53,54,57,59,62,64,66-69] reported in 14 publications[29,31,33,53,54,57,59,62,64-69] were included in the clinical impact review of MRI-based technologies. Five studies[29,59,66,68,69] reported in eight publications[29,31,33,59,65,66,68,69] evaluated the clinical impact outcomes associated with LiverMultiScan and six studies[53,54,57,62,64,67] were evaluations of the clinical impact of MRE.

5.4.1 Quality assessment

Seven[29,53,54,57,59,62,64] of the thirteen[29,53-64] DTA studies were also included in the clinical impact review. The EAG reassessed the methodological quality of the seven DTA studies[29,53,54,57,59,62,64] using the NIH study quality assessment tool.[51] Of the remaining four studies included in the clinical impact review, two were cohort studies,[66,67] one RCT was

described in two publications[68,74] and one was a qualitative study.[69] Full assessments using the NIH study quality assessment tool[51] for the seven DTA studies[29,53,54,57,59,62,64] and the two cohort studies[66,67] are presented in Appendix 10 (Table 23 and Table 24, respectively). The full assessment and summary of the risk of bias assessment for the included RCT[68,74] are presented in Appendix 11. The full assessment using the CASP qualitative studies checklist[52] for the included qualitative study[69] is presented in Appendix 12.

Cross-sectional studies included in the DTA review (n=7)

Five studies[29,53,59,62,64] reported the number of included patients but did not state how many patients were eligible for inclusion, therefore item 3 was rated as cannot determine (CD). Only one study[54] justified study sample size (item 5). The seven DTA studies[29,53,54,57,59,62,64] were cross-sectional studies and therefore did not assess exposure prior to measuring outcomes (item 6), include sufficient timeframes to determine an association between the exposure of interest (item 7) or assess exposure more than once over time (item 10). One study[54] did not report whether assessors were blinded (item 12). None of the seven studies[29,53,54,57,59,62,64] adjusted for the confounding variables in analyses for the outcome test failure rate (item 14).

New included cohort studies (n=2)

The authors of the Jayaswal study[66] only reported the number of included patients and did not state how many patients were eligible for inclusion, therefore item 3 was rated as CD. Neither study[66,67] justified study sample size (item 5). Assessment of liver disease only took place at baseline in both studies.[66,67] There was no mention of the outcome assessors being blinded to the status of the patients in the Gidener study;[67] the EAG assumed that assessors were not blinded given the retrospective study design. Confounding variables were measured in the Jayaswal study[66] but not adjusted for in the analysis.

New randomised controlled trial (n=1)

Information about the RCT was derived from a published protocol [74] (version dated 30th December 2020) and a Clinical Study Report (CSR)[68] provided by Perspectum, rather than from a publication or a manuscript submitted/ accepted for publication. The RCT[68,74] was judged to have low risk of bias for the selection of the reported result domain. However, the RCT[68,74] was judged to have a high risk of bias for the randomisation process because the trial was open-label and the authors did not present any patient characteristics data specifically for patients with NAFLD who underwent LiverMultiScan and liver biopsy. Number of unnecessary liver biopsies avoided data were only available for [redacted] of the [redacted] patients randomised. Therefore, the study was judged to have high risk of bias due to the high level of missing data. The deviations from the intended interventions domain was judged as presenting some concerns due to the open-label trial design and limited data analysis information about the number of

unnecessary liver biopsies avoided described in the protocol[74] and in the CSR.[68] Similarly, the RCT[68,74] was judged as presenting some concerns for outcome measurement due to the open-label design and possibility that the assessors may have known the results of tests that had been carried out prior to liver biopsy. The overall bias for the included RCT[68,74] was judged as high.

New qualitative study (n=1)

The McKay study[69] recruited patients from liver support groups, liver support charities and from Perspectum Ltd social media and online platform. The EAG considered that this was appropriate for the aims of study. However, the EAG notes that patients self-reported their diagnosis and considers this to be a potential source of bias. In the McKay study[69], the study author who conducted and coded the interviews had previously underwent the LiverMultiScan test and had later been diagnosed with liver disease. The McKay study[69] reports that this was a factor in initiating the study and therefore the EAG considers this to be a potential source of bias.

5.4.2 Characteristics of the included studies

Only one study[29] provided clinical impact results for a population of patients with NAFLD who had indeterminate or discordant results from fibrosis testing. Seven studies[29,53,54,57,59,62,64] that were included in the DTA review also provided evidence describing the clinical impact of MRI-based technologies for the assessment of patients with NAFLD. The characteristics of the original seven studies[29,53,54,57,59,62,64] are presented in Table 4. In addition to these seven studies,[29,53,54,57,59,62,64] the EAG identified four new studies.[66-69] Three studies described LiverMultiScan[66,67,69] and one study described MRE.[68] These comprised one prospective cohort study[66] based in the UK, one retrospective cohort study[67] based in the US, one RCT[68,74] based in Germany, Netherlands, Portugal and the UK and one qualitative study[69] based in the UK. The RCT[68,74] (RADIcAL trial) was a phase IV, multicentre, international study that evaluated the impact of using LiverMultiScan in the diagnostic pathway compared to standard of care (SoC) for patients with suspected NAFLD and was sponsored by Perspectum Ltd. Information about the RADIcAL trial[68,74] is presented in Table 7. The characteristics of the four new studies[66-68,69] are presented in Table 8.

Table 7 Key characteristics of the RADIcAL trial

Trial parameter	The RADIcAL trial[68,74]
Design	• Phase IV, multicentre, international study, open-label, RCT • 13 sites across 4 countries (Germany, Netherlands, Portugal and UK) • 5 year study (1 year study setup; 3 year recruitment phase; 12 months follow-up)
Patient population	• Patients (18 to 75 years old) with suspected NAFLD • Dosage of eculizumab stable for ≥3 months prior to screening • Within SoC: ○ 1.5xULN≤ ALT and AST ≤5xULN and GGT≥1.5xULN up to 1 year prior to patient recruitment or; ○ imaging suggestive of fatty liver disease up to 3 years prior to patient recruitment • Or presence of ≥3 of the following criteria: ○ insulin resistance of T2D ○ obesity (BMI>30.0 or waist-to-hip ratio >1.00 for men or >0.85 for women) ○ hypertension (≥130/85mmHg) ○ elevated triglycerides (≥1.7mmol/L) ○ low HDL-cholesterol (<1.05mmol/L for men or <1.25mmol/L for women)
Intervention	• Patients (n=[redacted]) were treated according to LiverMultiScan results. Further diagnostic evaluation was recommended when LiverMultiScan cT1≥800ms or PDFF≥10%. This was not a mandatory study requirement and was left at the discretion of the clinician and patient
Comparator	• SoC (n=[redacted])
Primary outcome	• Proportion of patients with suspected NAFLD incurring of liver-related hospital consultations and/or liver biopsies from the date of randomisation to end of study follow-up
Secondary outcomes	• Patient satisfaction at baseline and follow-up visits • Certainty of diagnosis (binary: yes/no) and frequency at baseline and follow-up visits • Time from randomisation to diagnosis by physician as recorded at final follow-up visit • Rates of liver-related outpatient investigations/consultations/hospital admissions per 400 patients during the study • Cost of LiverMultiScan compared to SoC • Personnel required to perform procedure and tasks from randomisation to end of study follow-up
Sample size calculation	• Sample size calculation based on a 14% reduction for the number of liver biopsies with LiverMultiScan compared to SoC • To maintain statistical significance with more than 80% power (α=0.05) and to show a difference in proportion of patients having consultations with LiverMultiScan compared to SoC, a sample size of 402 patients per arm was required • Upon inclusion of a 25% dropout rate, Perspectum calculated that they would require a cohort of 1072 patients with suspected fatty liver disease to be recruited into the trial

ALT=alanine aminotransferase; AST=aspartate aminotransferase; BMI=body mass index; cT1=iron corrected longitudinal relaxation time; GGT=gamma-glutamyl transferase; NAFLD=non-alcoholic fatty liver disease; PDFF=proton density fat fraction; RCT=randomised controlled trial; SoC=standard of care; T2D=type 2 diabetes; ULN=upper limit of normal
Source: RADIcAL trial[68,74]

Table 8 Characteristics of the new studies included in the clinical impact review

Publications	Study design; country; setting; timeframe	Population; number in analysis and recruitment details	Age (years); male (n, %); BMI (kg/m^2); T2D (n, %)	Interpreter of index test	Interpreter of liver biopsy
LiverMultiScan					
Jayaswal 2020[66]	Prospective cohort study; UK; NR; May 2011 to July 2017	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (n=85/197); recruited patients with compensated liver disease aetiologies scheduled to undergo clinically indicated liver biopsy or with a known diagnosis of liver cirrhosis	*Median age (IQR): 53 (44 to 59) Male: 123 (62) Median BMI (IQR): 28.4 (24.8 to 34.0) T2D: 42 (21)	Analysed using LiverMultiScan software by trained blinded analysts	Assessed for Ishak stage[73] by a blinded specialist liver histopathologist
McKay 2021[69]	Qualitative study; UK; NR	Patients with NAFLD (N=15/101); recruited patients with liver disease (N=90) and patient caregivers (N=11)	*Mean age (range): 51 (20 to 79) Male: 39 (38.6) BMI: NR T2D: NR	Analysed using LiverMultiScan software	NA
Perspectum Ltd. 2021[68,74]	RCT; Germany, Netherlands, Portugal and UK; secondary and tertiary care; Sept 2017 to Dec 2020	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed **(n=[redacted]); recruited patients with suspected or known fatty liver disease. Patients recruited from 7 UK sites (n=253)	***Median age: [redacted] Male: [redacted] Median BMI: [redacted] T2D: [redacted]	NR	NR
MRE					
Gidener 2022[67]	Retrospective cohort study; US; NR; retrospective 10 year follow-up of patients who underwent MRE Jan 2007 to Dec 2009	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed; (n=375/1269); recruited patients with chronic liver disease who underwent MRE for evaluation of liver fibrosis	*Median age (IQR): 55 (47 to 64) Male: 619 (48.8) Median BMI (IQR): 28.8 (25.1 to 33.6) T2D: NR	Drawn ROIs were verified by two expert MRE readers	****NR

*The statistics reported are based on the entire study population and not for the subpopulation of patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed
**Only [redacted] patients had complete histology scores from liver biopsy to confirm stage of fibrosis
*** The statistics reported are based on the entire study population and not for the subpopulation of patients with complete histology scores from liver biopsy
****The publication reported that liver biopsy was used to confirm cirrhosis. However, the proportion of patients who underwent liver biopsy was not reported and no methodological details were provided.
BMI=body mass index; IQR=interquartile range; MRE=magnetic resonance elastography; NA=not applicable; NAFLD=non-alcoholic fatty liver disease; NR=not reported; RCT=randomised controlled trial; ROI=region of interest; T2D=type 2 diabetes

5.4.3 Intermediate outcomes

Prognostic ability

Two studies[66,67] provided information about the prognostic ability of MRI-based technologies. The Jayaswal study[66] assessed the prognostic ability of the LiverMultiScan cT1 output to predict clinical outcomes for a population that included patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed (n=85/197). A subgroup analysis was conducted for the combined subpopulation of patients with the three main liver disease aetiologies (patients with NAFLD [n=85; 43%], alcohol-related liver disease [n=22; 11%] and viral hepatitis [n=50; 25%]). However, data were not provided for the subpopulation of patients with NAFLD only.

In the Jayaswal study,[66] results from LiverMultiScan liver 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 1ms increase in cT1. When a predefined cut-off of cT1>825ms[65] was applied, LiverMultiScan predicted event-free survival (p=0.006); all 11 clinical events that were recorded occurred amongst those who had a cT1 value of >825ms.

The Gidener study[67] reviewed long-term data (≥10 years) from 1269 patients to assess the ability of MRE results to predict clinical outcomes for patients with chronic liver disease who underwent a single MRE between January 2007 to December 2009. The Gidener study[67] reviewed patients' electronic health records for evidence of cirrhosis, decompensation of cirrhosis (defined by at least one decompensation event including oesophageal variceal bleeding, ascites, hepatic encephalopathy, or jaundice), transplant, hepatocellular carcinoma, cholangiocarcinoma or death. The study population included 375 patients with NAFLD. The Gidener study[67] reported that MRE liver stiffness at baseline predicted a lower rate of cirrhosis development (HR=0.37 per 1kPa increase in MRE liver stiffness output, 95% CI: 0.19 to 0.71; p=0.003) for patients with non-cirrhotic NAFLD at baseline compared to patients 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 patients with NAFLD only.

Number of liver biopsies

The RADIcAL trial CSR[68] reported the number of unnecessary liver biopsies avoided by using LiverMultiScan cT1 and LiverMultiScan PDFF results. Unnecessary biopsies were defined as biopsies carried out in patients who had a negative NASH diagnosis. The RADIcAL trial[68]

reported that [redacted] patients with [redacted] and [redacted] underwent unnecessary biopsies in the LiverMultiScan arm (n=[redacted], [redacted]%) compared to the SoC arm (n=[redacted] [redacted]%, EAG calculated odds ratio [OR]=[redacted], 95%CI: [redacted] to [redacted]). The RADIcAL trial[68] also reported that [redacted] patients with [redacted] in the LiverMultiScan arm underwent unnecessary biopsies (n=[redacted], [redacted]%) compared to the SoC arm (n=[redacted] [redacted]%, EAG calculated OR=[redacted], 95% CI: [redacted] to [redacted]).

A [redacted] proportion of patients with [redacted] and [redacted] underwent unnecessary biopsies with elastography [redacted] and without elastography [redacted] prior to biopsy. A [redacted] proportion of patients with [redacted] underwent unnecessary biopsies with elastography [redacted] and without elastography [redacted] prior to biopsy.

The RADIcAL trial[68] reported [redacted] (Appendix 13, Figure 11) and [redacted] (Appendix 13, Figure 12).

Test failure rate

Three studies[29,59,66] reported test failure rate for LiverMultiScan and six studies[53,54,57,62,64,67] reported test failure rate for MRE. However, two of the studies[59,66] that assessed LiverMultiScan and three of the studies[54,64,67] that assessed MRE included patients with other liver disease aetiologies in addition to NAFLD and did not provide data specific to patients with NAFLD.

The test failure rate of LiverMultiScan for patients with all liver aetiologies ranged from 5.3%[59] to 7.6%[66] and the test failure rate of LiverMultiScan for patients with NAFLD only was 5.6%.[29] The reasons for LiverMultiScan test failure specific to patients with NAFLD were technical failure (n=1/3), MRI scan cancelled (n=1/3) and patient unable to tolerate MRI scan (n=1/3).[29]

The MRE test failure rate for patients with all liver aetiologies ranged from 0.0%[54] to 7.6%[53] and the MRE test failure rate for patients with NAFLD only ranged from 3.9%[57] to 7.6%.[53] The EAG performed a fixed-effects meta-analysis to obtain a pooled estimate of test failure rate for patients with NAFLD (test failure rate=4.2%, 95% CI: 2.5% to 6.2%); a forest plot displaying this analysis is provided in Appendix 14 (Figure 13). Minimal statistical heterogeneity was observed between the included studies (I^2=18.9%). The reasons for MRE test failure specific to patients with NAFLD were technical failures (n=11/24),[57,62] patients refusing the test (n=9/24),[53,57] claustrophobia (n=3/24)[53] and the patient being unable to fit in the scanner (n=1/24).[53]

Patient acceptability of different testing modalities

The McKay study[69] collected feedback from patients with liver disease (n=90) and from patient caregivers (n=11) after patients had had a LiverMultiScan. In the McKay study,[69] patients had an MRI scan and MRI data were analysed using LiverMultiScan software. A healthcare professional discussed the LiverMultiScan report with patients in a one-on-one setting and, immediately after the discussion, a study investigator conducted a semi-structured interview that consisted of open-ended questions about the patient's experience of the MRI scan, the patient's understanding of the LiverMultiScan report and ways to improve the scan and report experience. The interviews were transcribed, and thematic analysis was completed.

The McKay study[69] reported that patients considered the MRI scan to be a harmless and tolerable procedure and many highlighted that the non-invasive element of the procedure was important. Although some patients were anxious prior to the scan, most considered that the scan was not particularly stress-inducing. Most patients did not have claustrophobia. However, some patients who did have claustrophobia successfully dealt with the stressor by closing their eyes or using a blindfold during the MRI scan. Many patients considered that, during the MRI scan, sound was a greater psychological stressor than claustrophobia. However, most patients considered that the level of sound was acceptable. Most patients successfully completed the required breath-holding. Some patients struggled with breath-holding (particularly patients with lung-related comorbidities) and reported that a practical demonstration prior to the scan would have been helpful. Some patients considered that the 4 hours fasting required prior to the scan was an issue; fasting may be problematic for some patients with strict medication regimes. However, most patients did not consider this to be an issue.

The McKay study[69] also collected patient feedback on the LiverMultiScan diagnostic report. However, clinical advice to the EAG is that the LiverMultiScan diagnostic report would not usually be made available to patients in NHS clinical practice. According to the McKay study,[69] most patients considered that the diagnostic report was clear and understandable; the statistics reported were clear and the use of imagery, colour and the inclusion of a full liver scan picture improved their understanding of their condition. However, some patients reported that they were confused by some of the terminology and acronyms, e.g., liver inflammation and fibrosis (LIF) and cT1. Most patients considered that the diagnostic report was very important for understanding their disease and helped them to feel empowered and involved in their clinical management. The McKay study[69] reported that careful information delivery by a doctor or health professional was considered essential to assure patients of the quality and validity of their LiverMultiScan results.

In the McKay study,[69] some patients reported that they hoped that the LiverMultiScan results would mean that they could avoid liver biopsy. Patients reported that biopsy was very uncomfortable and caused psychological stress. Patients preferred MRI-based technologies and TE because they were non-invasive, short in duration and results could be delivered quickly.

Clinical impact outcomes (additional targeted searches)

Despite conducting additional targeted searches (Section 4.1.1), the EAG did not identify any relevant studies that provided evidence of the clinical impact of MRI-based technologies for patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed, for the remaining clinical impact outcomes listed in the final scope[23] issued by NICE, namely:

- impact of test result on clinical decision making
- uptake and maintenance of lifestyle modifications
- time to receive test results
- time to diagnosis
- reduction or remission of liver fibrosis or fibro-inflammation
- reduction or remission of liver fat
- mortality
- morbidity
- health-related quality of life

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.[74] However, the company did not provide any data for time to diagnosis in the CSR.[68]

Clinical advice to NICE[23] was that the results generated by MRI-based technologies can motivate people with NAFLD to take up and maintain recommended lifestyle modifications. The EAG performed a broader literature search and identified one study[75] that assessed the relationships between patients with NAFLD and their perceptions about disease consequences and treatment, patient self-efficacy and healthy lifestyle maintenance. This study[75] did not assess the impact of MRI-based technologies; however, the study reported that patient self-efficacy and understanding of their illness were factors that were associated with better nutritional habits, whereas emotional representation (the extent that patients were afraid or concerned about having NAFLD) and perceptions of more severe illness were associated with poorer nutritional habits. Neither of the two companies has assessed whether LiverMultiScan or MRE results affect patient understanding of NAFLD or emotional representation, or whether LiverMultiScan or MRE results impact levels of lifestyle modification compliance.

5.5 Summary of EAG DTA and clinical impact review, and EAG quantitative analysis

EAG DTA and clinical impact review

The EAG DTA review identified 13 studies[29,53-64] reported in 15 publications.[29,31,53-65] The EAG clinical impact review identified 11 studies[29,53,54,57,59,62,64,66-69] reported in 14 publications.[29,31,33,53,54,57,59,62,64-69] However, the EAG was only confident that one study (the Eddowes 2018[29] study) was carried out in the population described in the final scope[23] issued by NICE, namely patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed:

- patients who have indeterminate results from fibrosis testing
- patients for whom TE or ARFI is unsuitable
- patients who have discordant results from fibrosis testing.

The clinical impact review only identified one RCT; the RADIcAL trial,[68] which was carried out by Perspectum Ltd. Results from this study[68] showed that, compared with patients in the standard care arm, [redacted] underwent unnecessary biopsies in the LiverMultiScan arm. Feedback from Perspectum Ltd[71] and the McKay study[69] was that patients and carers experiences of using LiverMultiScan were positive.

EAG quantitative analysis

The only relevant study[29] (n=50) identified by the DTA review focused on the potential of LiverMultiScan to deliver cost savings compared to biopsy and included clinical results (for example, cT1 and PDFF scores). Perspectum Ltd[71] provided more detailed DTA results for the patients included in the Eddowes study.[29] The Eddowes study[29] categorised patients according to low- and high-risk of progressive liver disease; however, Perspectum Ltd[71] suggested six other ways of interpreting the DTA data generated by LiverMultiScan (from the same study): any fibrosis (≥F1), significant fibrosis (≥F2), Brunt Grade ≥1, Brunt Grade ≥2, NASH and advanced NASH. In response to a request from the EAG, Perspectum Ltd[71] also provided data for patients with advanced fibrosis (≥F3).

No DTA data were submitted to NICE by the manufacturer of MRE (Resoundant, Inc). Eleven studies[53-58,60-64] evaluated the DTA of MRE, but none of the studies explicitly included patients with indeterminate or discordant results from previous fibrosis testing.

The EAG carried out a quantitative analysis using data from six studies.[29,56-59,62] Where patients were diagnosed consistently across studies (fibrosis, steatosis, and NASH), the EAG carried out meta-analyses using cT1 and PDFF outputs for LiverMultiScan and for MRE.

Results from the EAG meta-analyses suggested that the LiverMultiScan cT1 output is more sensitive and specific than the LiverMultiScan PDFF output, and that for the diagnosis of fibrosis (≥F3), MRE has high DTA. However, the meta-analyses were populated with data from small numbers of studies and only one[29] of the studies included the population that is the focus of this assessment. This should be considered when interpreting the results from the EAG meta-analyses.

6 METHODS FOR ASSESSING THE COST EFFECTIVENESS

The aim of the EAG economic evaluation was to evaluate whether the use of MRI-based technologies for the assessment of NAFLD represented a cost effective use of NHS resources. The population of interest was patients with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed and:

- who had indeterminate results from fibrosis testing
- for whom TE or ARFI was unsuitable
- who have discordant results from fibrosis testing.

The economic evaluation included a systematic review of existing economic evaluations of MRI-based technologies and the creation of a de novo economic model.

6.1 Systematic review of cost effectiveness evidence

The EAG undertook a systematic review to identify full economic evaluations that were designed to explore the cost effectiveness of the use of MRI-based technologies as diagnostic tools for the three subpopulations of interest with NAFLD for whom advanced fibrosis or cirrhosis had not been diagnosed.

6.1.1 Search strategy

The search strategies used to identify diagnostic and clinical impact evidence for inclusion in the clinical effectiveness review can be found in Appendix 3. To identify published economic evaluations, the EAG appended an economic evaluation-specific search filter to the clinical search strategies (Appendix 15). In addition, two databases of economic publications (EconLit [EBSCO] and the Cost-Effectiveness Analysis [CEA] registry) were searched, using the search strategies presented in Appendix 15, from inception until 4th October 2021. The results of the searches were entered into an Endnote X9 library and de-duplicated (MM) before being exported into Covidence.

6.1.2 Study selection and inclusion criteria

The review inclusion and exclusion criteria (Table 9) reflected the decision problem outlined in the final scope[23] issued by NICE.

The identified publications were assessed for inclusion in the review using a two-stage process. First, two reviewers (DB and RH) independently screened all the titles and abstracts identified by the electronic searches to find potentially relevant records. Second, full-text copies of these records were obtained and assessed independently by two reviewers (DB and

RH) using the inclusion criteria presented in Table 9. Disagreements were resolved through discussion at each stage and in all cases, a consensus was reached.

Table 9 Economic review inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Population	The population of interest is patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed and: • who had indeterminate results from fibrosis testing • for whom TE or ARFI was unsuitable • who had discordant results from fibrosis testing	Publications that do not include analyses of patients with NAFLD
Intervention	MRI-based technologies, i.e., LiverMultiScan (multiparametric MRI), and MRE	Non MRI-based technology
Comparator	• LiverMultiScan • MRE • no comparator	
Outcomes	Cost of test accuracy, cost per intermediate outcomes, incremental cost per LY gained and/or incremental cost per QALY gained	
Study design	Full economic evaluations that consider both costs and consequences (i.e., cost effectiveness analysis, cost utility analysis, cost minimisation analysis and cost benefit analysis)	Partial economic evaluations that only consider either costs or consequences or do not compare two or more treatments with each other Studies that do not present original data (i.e., reviews, editorials and opinion papers)
Language	English only	Non-English language studies

MRE=magnetic resonance elastography; MRI=magnetic resonance imaging; NAFLD=non-alcoholic fatty liver disease; LY=life year; QALY=quality-adjusted life year

6.1.3 Data extraction

A data extraction form was designed in Microsoft Excel. Extracted data included bibliographic information (author[s] and year of publication), type of economic evaluation, country, perspective, population, intervention and comparators, model structure, model outcomes, and sensitivity analyses undertaken. Data extraction was carried out independently by two reviewers (DB and RH) and the two reviewers agreed the final version of the completed data extraction form.

6.1.4 Quality of cost effectiveness evidence

The EAG assessed the quality of the included economic evaluations using the Drummond checklist[76] for assessing economic evaluations and the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.[77] Quality assessment was performed by one reviewer (DB) and checked for accuracy by a second reviewer (RH). All disagreements

were resolved through discussion. There were no unresolved issues and, therefore, it was not necessary to consult with a third reviewer.

6.1.5 Results of the systematic review of existing cost effectiveness evidence

The searches resulted in the identification of 253 publications. Once duplicates (n=49) had been removed, 204 publications remained. Following first-stage screening (titles and abstracts), 31 publications were retrieved for full-text review. After assessing applying inclusion criteria, one publication[29] was identified as being relevant. The PRISMA flow diagram[48] provides an illustration of the screening and selection process (Figure 2). A list of the studies excluded at the full-text stage, along with reasons for exclusion, is provided in Appendix 16.

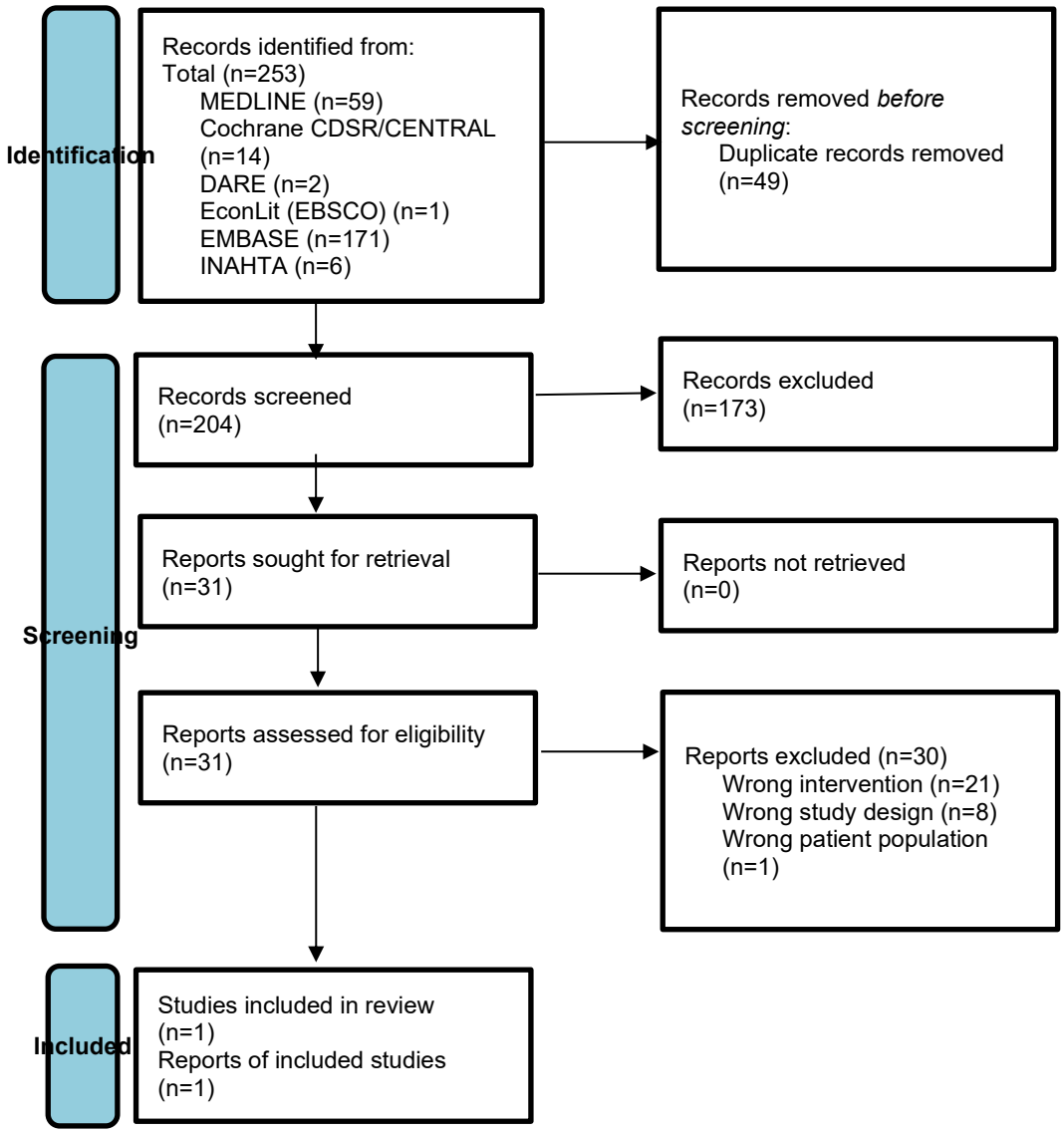

Figure 8 PRISMA flow diagram for the cost effectiveness review

6.1.6 Quality of the included evidence

The quality of the included study[29] was assessed using the Drummond checklist[76] (Table 10) and the CHEERS checklist[77] (Appendix 17).

Table 10 Drummond checklist[76] summary of publication that was included in the EAG's review of economic evidence

Question	Eddowes 2018[29]
Was a well-defined question posed in answerable form?	✗
Was a comprehensive description of the competing alternatives given?	✓
Was the effectiveness of the programme or services established?	Unclear
Where all the important and relevant costs and consequences for each alternative identified?	Unclear
Were costs and consequences measured accurately in appropriate physical units?	Unclear
Were the cost and consequences valued credibly?	Unclear
Were costs and consequences adjusted for differential timing?	✓
Was an incremental analysis of costs and consequences of alternatives performed?	✗
Was allowance made for uncertainty in the estimates of costs and consequences?	Unclear
Did the presentation and discussion of study results include all issues of concern to users?	✗

✓ yes (item properly addressed) ✗ no (item not properly addressed) ✓/✗ partially (item partially addressed)

The population (n=50) described in the published paper is patients with inconclusive results from fibrosis testing. The EAG has assumed that inconclusive is an umbrella term for a group of patients with indeterminate and/or discordant results from previous fibrosis testing. The EAG notes that all patients considered in this analysis were scheduled for a biopsy. This means that the study sample does not represent all patients with indeterminate and/or discordant results from previous fibrosis testing; clinical advice to the EAG is that not all patients with indeterminate and/or discordant results will have a biopsy.

Eddowes 2018[29] repeated the analyses carried out by Blake 2016[78] using DTA results from their study. Blake 2016[78] constructed a simple decision tree to compare the costs for three NAFLD diagnostic pathways that use non-invasive techniques. The patients modelled by Blake 2016[78] did not have inconclusive results from previous fibrosis testing and therefore the Eddowes 2018[29] cost saving results are not relevant to this appraisal.

The information provided in the published paper[29] is limited and, therefore, it is unclear whether all important costs and consequences were included in the analysis, or whether the included costs and consequences were valued credibly. An incremental analysis was not performed and there is no evidence that any sensitivity or scenario analyses were performed. The authors did not describe the limitations of the cost effectiveness analysis, nor the generalisability of results.

6.1.7 Characteristics of the included study

The characteristics of the included study[29] are summarised in Table 4. This study[29] was also included in the EAG DTA and clinical impact review.

The included study, Eddowes 2018,[29] reported results from a cost utility analysis. The population was adult patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and who were scheduled for non-targeted liver biopsy to stage fibrosis after inconclusive non-invasive assessment of fibrosis or to make a diagnosis after a range of non-invasive tests had not confirmed a diagnosis. Three diagnostic tools were considered LiverMultiScan (two cut-offs: 822ms and 875ms), TE (two cut-offs: 5.8kPa and 7.0kPa), ELF (two cut-offs: 7.7 and 9.8); LiverMultiScan plus TE (four combinations of cut-offs) was also considered. The perspective of the analysis was the UK NHS, and the time horizon was 2 weeks (i.e., LiverMultiScan and TE were performed within 2 weeks of biopsy).

Results were generated by a decision tree model. The model was populated with clinical effectiveness evidence from a cross-sectional study undertaken at the Queen Elizabeth Hospital Birmingham and the Royal Infirmary of Edinburgh (ISRCTN39463479). Costs were sourced from the NHS tariff and the cost year was 2016. The short time horizon of the model meant that it was not necessary to discount costs and benefits.

6.1.8 Study results and conclusions

Study results

The model generated results in terms of biopsies avoided, total costs, cost saving versus biopsy and total cost per correct diagnosis. cost per correct diagnosis and the number of biopsies avoided for a hypothetical cohort of 1000 patients. Results (Table 11) show that, of the interventions considered, LiverMultiScan (875ms) plus TE (7.0kPa) generated the highest number of biopsies avoided (848.7 per 1000 patients) at the lowest cost (£237,488 per 1000 patients). This approach also delivered the highest cost saving versus biopsy (£402,122) and the lowest cost per correct diagnosis (£307.92).

Table 11 Cost effectiveness results

Intervention	Biopsies avoided	Total costs	Cost savings versus biopsy	Total costs per correct diagnosis
	Per 1000 patients			
LMS cT1 822ms[a]	381.9	£538,345	£101,265	£649.57
LMS cT1 875ms[b]	458.4	£489,392	£150,218	£554.26
TE 5.8 kPa[a]	297.2	£517,530	£122,080	£814.16
TE 7.0 kPa[b]	491.6	£393,146	£246,464	£590.14
ELF 7.7[a]	151.1	£654,010	-£14,400	£1,138.43
ELF 9.8[b]	858.9	£201,322	£438,288	£363.97
LMS cT1 822ms+TE 5.8kPa[a]	734.6	£338,260	£301,359	£415.37
LMS cT1 875ms+TE 5.8kPa	722.7	£345,851	£293,759	£414.60
LMS cT1 822ms+TE 7.0kPa	841.1	£242,309	£397,301	£315.60
LMS cT1 875ms+TE 7.0kPa[b]	848.7	£237,488	£402,122	£307.92

[a] Patients with simple steatosis and ≤F1 fibrosis
[b] Patients with either NASH or >F1 fibrosis
cT1=iron corrected longitudinal relaxation time; ELF=enhanced liver fibrosis; LMS=LiverMultiScan; NASH=non-alcoholic steatohepatitis; TE=transient elastography
Source: Eddowes 2018[29] study

Study conclusions

The authors concluded that LiverMultiScan combined with TE, delivered the lowest cost per correct diagnosis.

6.1.9 EAG cost effectiveness review conclusions

The EAG searches for published economic evaluations that assessed the cost effectiveness of LiverMultiScan and MRE only identified one study.[29] The included study[29] assessed the comparative cost savings versus biopsy of LiverMultiScan, TE, ELF and LiverMultiScan plus TE. The authors provided limited data describing the study methods and results and, therefore, study quality and the generalisability of results are unclear.

In the Eddowes 2018[29] study, clinical effectiveness evidence was collected from a population with inconclusive results from previous fibrosis testing. To generate cost effectiveness results, Eddowes 2018[29] study clinical effectiveness data were used to populate the Blake 2016[78] model. However, the focus of the Blake 2016[78] model was not to explore cost effectiveness for patients with inconclusive results from previous fibrosis testing. Therefore, Eddowes 2018[29] study cost effectiveness results are not relevant to this appraisal.

6.2 Development of a de novo model

6.2.1 Introduction

The EAG cost effectiveness review did not identify any published economic evaluations that were relevant to this appraisal; the EAG has therefore developed a de novo economic model.

Perspectum Ltd suggest[23] that LiverMultiScan results can be used by clinicians to help diagnose patients with fatty liver, NASH and high-risk NASH. Perspectum Ltd[71] also provided LiverMultiScan DTA results for a range of other diagnoses, including advanced fibrosis (≥F3). Whilst LiverMultiScan results are unlikely to inform patient treatment plans, they can potentially be used to help identify patients for whom a biopsy may not be appropriate. In contrast, biopsy results provide an accurate diagnosis and data that can be used to inform patient treatment plans, for example, identification of co-factors for liver injury (such as alcohol, iron, or auto-immune hepatitis). However, biopsy is an expensive invasive procedure that is not without risks. If LiverMultiScan results could be used to help identify patients who do not require a biopsy, this would benefit patients by reducing the number of unnecessary biopsies and would save NHS resources. The primary clinical outcome from the EAG model is therefore the number of biopsies avoided if LiverMultiScan were introduced into the diagnostic pathway.

The EAG cost effectiveness results will be driven by the proportion of patients who, if they had a biopsy, would test positive, i.e., population prevalence. This estimate is independent of LiverMultiScan test accuracy (or the accuracy of any other test introduced into the diagnostic pathway). Population prevalence estimates vary depending on two factors, the diagnosis and the population investigated. Published evidence[56,58] shows that population prevalence varies by population investigated; it is essential that the prevalence data used to populate the EAG model relate to the population described in the final scope[23] issued by NICE and are generalisable to patients treated in NHS clinical practice.

The EAG only identified one study (Eddowes 2018[29]) that provided LiverMultiScan DTA and population prevalence data that were focused on patients, who were scheduled for, and received, a biopsy, and who had inconclusive results from previous fibrosis testing.

As DTA data are only available for patients with inconclusive results who received a biopsy, the Eddowes 2018[29] study population represents a subset of the population described in the final scope[23] issued by NICE. Clinical advice to the EAG is that not all patients with inconclusive results from previous fibrosis testing would be referred for a biopsy; reasons for not referring a patient for a biopsy include presence of co-morbidities, personal choice, old age and medical contraindications. The utility of positive LiverMultiScan results for patients

who would not be referred for biopsy is unclear and is not considered in the EAG model. No DTA or population prevalence data are available for the full population described in the final scope[23] issued by NICE.

Further, LiverMultiScan data are not available for patients for whom TE or ARFI were unsuitable. In addition, no DTA or population prevalence data were available for any of the population described in the final scope[23] issued by NICE for patients who had had an MRE.

The EAG cautions that the data presented in the Eddowes 2018[29] study relate to 50 patients and the data presented by Perspectum Ltd[71] relate to 46 patients; however, both sets of data appear to be from the same group of patients, i.e., as described in the Eddowes 2018[29] publication, and are referred to as Eddowes 2018/Perspectum Ltd.[29,71]

The EAG model has been developed based on the assumption that the LiverMultiScan DTA results are robust and will be used to stop clinicians from sending patients with a negative result for a biopsy. However, if this assumption does not hold then results from the EAG model should not be used to inform decision making.

6.2.2 Model structure

The EAG built a decision tree in Microsoft Excel® to estimate the costs and quality adjusted life years (QALYs) associated with two diagnostic pathways, LiverMultiScan plus biopsy and liver biopsy only. Eight different diagnostic test strategies described in the literature or by Perspectum Ltd[71] were investigated. Eddowes 2018[29] chose to categorise patients according to low- and high-risk of progressive liver disease; however, Perspectum Ltd[71] has provided data for seven other ways of interpreting the DTA data generated by LiverMultiScan (from the same study). The eight different diagnostic test strategies considered by the EAG were:

- T1: Any fibrosis (≥F1)
- T2: Significant fibrosis (≥F2)
- **T3: Advanced fibrosis (≥F3)**
- T4: Brunt Grade ≥1
- **T5: Brunt Grade ≥2**
- T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)
- **T7: Advanced NASH (NAS≥4 plus ≥F2)**
- **T8: High risk of progressive disease (NASH or >F1)**

In the EAG model, for each of the eight diagnostic test strategies (T1 to T8), if a patient's LiverMultiScan result exceeds the specific cT1 or PDFF thresholds associated with the test strategy, then the patient is defined as having a positive result and will have a biopsy. The

EAG asked the Specialist Committee members to consider the eight diagnostic test strategies and identify any strategies for which a positive LiverMultiScan result would not change their decision to send a patient for a biopsy. The advice from the Specialist Committee was that, for patients with LiverMultiScan test results suggesting a diagnosis of T3, T5, T7 and T8, the decision whether to send the patient for a biopsy would not change, i.e., patients who had a positive LiverMultiScan test result would proceed to biopsy. The EAG has presented results for all strategies but considers that the findings from the strategies in bold are the most important.

In the model, LiverMultiScan cT1 or PDFF results lead to the following consequences:

- true positive (TP); LiverMultiScan result and biopsy result are both positive – correctly identified by LiverMultiScan results and patient is appropriately sent for a biopsy
- false positive (FP); LiverMultiScan result positive and biopsy result negative – incorrectly identified by LiverMultiScan results and patient is inappropriately sent for a biopsy
- true negative (TN); LiverMultiScan result negative and biopsy, if performed, would have been negative – correctly identified by LiverMultiScan results and the patient was appropriately not sent for a biopsy, LiverMultiScan repeated at 6 months, result is negative and no biopsy required
- false negative (FN); LiverMultiScan result negative but biopsy, if performed, would have been positive – incorrectly identified by LiverMultiScan results and patient inappropriately not sent for a biopsy, LiverMultiScan repeated at 6 months, biopsy following repeat LiverMultiScan (assumed always to be positive)
- test failure – patients go to straight to biopsy

The accuracy of liver biopsy does not influence EAG cost effectiveness results; the model is driven by the congruence of the LiverMultiScan and biopsy results and not by the diagnoses reached following a biopsy.

The assumption that all patients with a negative LiverMultiScan test result will go on to have a repeat LiverMultiScan at 6 months and will then be correctly diagnosed is optimistic and favours the LiverMultiScan plus biopsy pathway for two reasons. First, it seems implausible that the accuracy of a second LiverMultiScan test will be 100%; some patients are likely to have a second FN result and some patients with an initial TN result will have a FP result and will go straight to (an unnecessary) biopsy. Second, the EAG has assumed that patients whose second LiverMultiScan test results are negative will have no further tests as this result is assumed to be a TN.

The population prevalence can be estimated by adding together the number of patients with TP and FN results.

Perspectum Ltd[71] has suggested that patients will receive a second LiverMultiScan if their cT1 score is between 800 and 875ms; however, the EAG has assumed that patients with cT1 scores less than 800ms will also receive a second LiverMultiScan. The EAG considers that this assumption is appropriate as all tests for this cohort have low specificity (i.e., high rates of FNs).

As all patients are assumed to be correctly diagnosed by 6 months, the LiverMultiScan plus biopsy pathway benefits arise from identifying people who are TNs and removing the costs and lost QALYs arising from these patients having unnecessary biopsies. These benefits are balanced against the LiverMultiScan plus biopsy pathway costs and the QALY loss associated with FNs.

Currently, NHS patients with inconclusive results from previous fibrosis testing may be sent for a biopsy or receive no further diagnostic tests (Figure 9); the proposed LiverMultiScan plus biopsy pathway is shown in Figure 10.

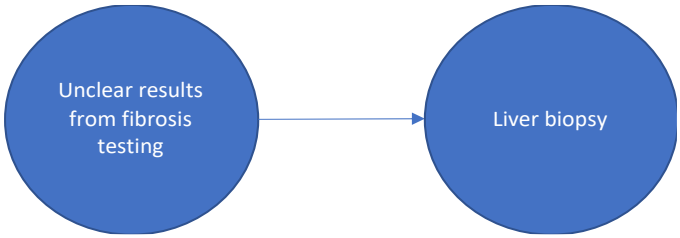

Figure 9 Current NHS diagnostic test pathway for patients with inconclusive results from previous fibrosis testing

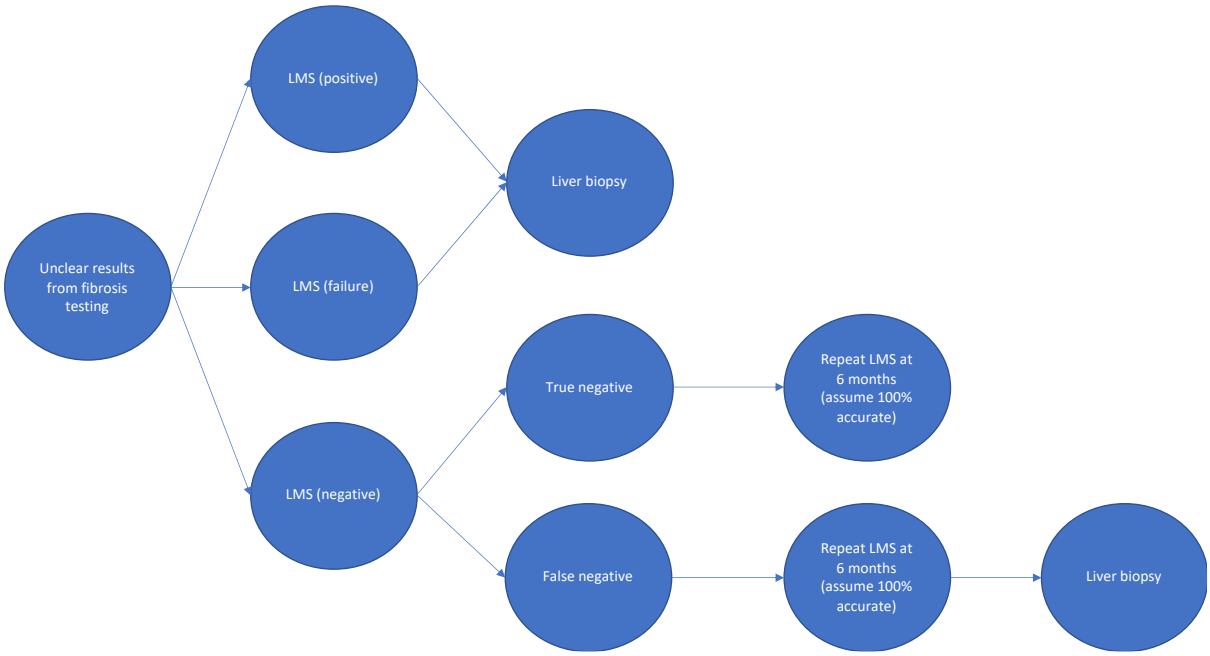

Figure 10 Proposed LiverMultiScan plus biopsy diagnostic test pathway for patients with inconclusive tests from fibrosis testing

6.2.3 Population

The modelled population is patients with inconclusive results from fibrosis testing who, without access to LiverMultiScan, would be scheduled for and would receive a biopsy. Patient characteristics are based on the population described in the Eddowes 2018[29] study. All patients (n=50) had a histologically confirmed diagnosis of NAFLD without secondary causes and without history of alcohol excess; 32 patients had an inconclusive non-invasive assessment of fibrosis and 18 patients had undergone a range of non-invasive tests without a firm diagnosis being made. Over half of the patients were male (56%), their average age was 54 years, 86% were Caucasian, 58% were non-smokers and 10% of patients in the study were post-transplant.

6.2.4 Intervention

For patients with inconclusive results from fibrosis testing who were scheduled for, and received, a biopsy, DTA data and population prevalence data were only available from a population of patients who had received a LiverMultiScan.[29] No MRE DTA data and population prevalence data were available for the population described in the final scope[23] issued by NICE.

Cut-off values have been proposed by Perspectum Ltd[71] for the staging of fibro-inflammation, associated diagnoses, and clinical management options. The normal reference range for PDFF is ≤5.6% liver fat content. The proposed cT1 cut-off values are:

- Less than 800ms: “Fatty liver”
    - Reassure as no inflammation present
    - Reassess with MRI in 3 years
- 800 to 875ms: “Non-alcoholic steatohepatitis (NASH)”
    - Lifestyle modification
    - Management of type 2 diabetes and cardiovascular disease
    - Monitor disease status with MRI after 6 months
- More than 875ms: “High-risk NASH”
    - Reassess with MRI every 6 months
    - Consider liver biopsy if cirrhosis is suspected
    - Cancer surveillance
    - Consider inclusion in NASH therapeutic trials

When compared with the PDFF values from the same cohort of patients and using the same diagnostic test strategies, the cT1 scores always generated the same or higher sensitivity and specificity values. The EAG cost effectiveness analysis is, therefore, populated with

LiverMultiScan cT1 scores. For completeness, the cost effectiveness results generated using PDFF values are presented in Appendix 18.

6.2.5 Comparator

The comparator is liver biopsy only which represents current standard of care.

6.2.6 Time horizon, discounting and perspective

The model has a maximum time horizon of 6 months and ends when a patient has a biopsy or has been accurately diagnosed following a repeat LiverMultiScan test. The short model time horizon means that discounting of costs and benefits is not relevant. The cost perspective of the model is the NHS. For patients in the LiverMultiScan plus biopsy pathway, only the costs and outcomes associated with the LiverMultiScan test and biopsy are considered. For patients in the biopsy only pathway, only the costs and outcomes associated with biopsy are considered.

6.2.7 EAG model parameters

Diagnostic test accuracy

LiverMultiScan rates of TP, FP, TN and FN are a function of the sensitivity and specificity of the LiverMultiScan test and the population prevalence. These rates vary depending on the diagnostic test strategy considered and have been estimated from evidence provided by Eddowes 2018/Perspectum Ltd.[29,71] The DTA estimates have been used to populate the different decision tree nodes for different diagnostic test strategies (Table 12). The LiverMultiScan test failure rate reported by Eddowes 2018[29] was 5.5%. In the EAG model, any patient who had a test failure result was referred for a biopsy.

Table 12 LiverMultiScan diagnostic test accuracy strategies and values (per 1,000 successful tests)

Diagnostic test strategy		cT1 cut-off value	Population prevalence	True positive	True negative	False positive	False negative	Sensitivity	Specificity
T1	Any fibrosis (≥F1)	800ms	87.0%	761	87	43	109	0.88	0.67
T2	Significant fibrosis (≥F2)	875ms	65.2%	413	261	87	239	0.63	0.75
T3	Advanced fibrosis (≥F3)	875ms	47.8%	304	196	326	174	0.64	0.63
T4	Brunt Grade ≥1	800ms	97.8%	782	0	22	196	0.8	0
T5	Brunt Grade ≥2	875ms	50.0%	348	348	152	152	0.7	0.7
T6	NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	54.4%	348	304	152	196	0.64	0.67
T7	Advanced NASH (NAS≥4 plus ≥F2)	875ms	47.8%	304	326	196	174	0.64	0.62
T8*	High risk (NASH or >F1)	875ms	79.4%	772	107	99	22	0.975	0.5

* Only sensitivity and specificity values were available from the Eddowes 2018[29] study the other values were calculated by the EAG
cT1= iron corrected longitudinal relaxation time; DTA=diagnostic test accuracy; F=fibrosis stage; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis
Source: Eddowes 2018 study/Perspectum Ltd[29,71]

6.2.8 Intervention and comparator costs

Unless otherwise stated, the intervention costs are presented in 2019/20 GBP. The costs prior to receiving a LiverMultiScan or biopsy, which ever test comes first in the pathway, are not included in the EAG analysis. Intervention costs are displayed in Table 13.

Table 13 Intervention costs

Intervention	Cost	Description	Source
Biopsy	£1,513	YG10Z Percutaneous transvascular* biopsy of lesion of liver	NHS Reference Costs 2019/20[79]
	£770	YG11A Percutaneous punch† biopsy of lesion of liver, 19 years and over	
	£805	Weighted average of YG10Z and YG11A	
MRI	£148.24	RD01A Scan of one area, without contrast, 19 years and over	
LiverMultiScan	£199	Cost per scan for data analysis and reporting	Perspectum Ltd[29,71]

* transjugular
† standard biopsy procedure
MRI=magnetic resonance imaging

6.2.9 Biopsy complications

The Stevenson study[80] estimated the average costs (per biopsy) of treating complications associated with a percutaneous biopsy and a transjugular biopsy to be £7 and £13 respectively. An EAG targeted literature search failed to identify more robust estimates. The EAG weighted the Stevenson study[80] costs by the proportions of patients (NHS Reference Costs 2019/20[79]) who had percutaneous and transjugular biopsies (£7.30) and inflated the weighted cost to 2019/20 prices (£8.54) using the NHS Cost Inflation Index (pay and prices index).

6.2.10 Utility values

The only utility values required in the EAG model are the disutilities associated with having a biopsy. The EAG carried out a targeted search of the literature; however, the EAG did not identify any primary studies that reported disutility values specifically associated with liver biopsy for patients with inconclusive results from fibrosis testing. There is no information in NG49,[9] the NICE guideline for the assessment and management of NAFLD, about the disutility associated with having a biopsy. However, the Stevenson study[80] identified that a loss of utility due to biopsy can be caused by direct pain and anxiety, serious adverse events and death. The EAG also considers that loss of utility can arise from failure to treat patients with advanced liver disease (i.e., LiverMultiScan test FN results).

1a Disutilities associated with having a liver biopsy: direct pain and anxiety

The EAG considers that it is not unreasonable that there would be a loss in utility due to the pain and anxiety associated with a liver biopsy. Clinical advice to the EAG is that it would be appropriate to use a level 3 decrement for pain, lasting for 1 day (utility loss=0.386, QALY loss=0.00105) and a level 3 decrement for anxiety lasting for a week prior to the biopsy (utility loss=0.236, QALY loss=0.00453) in the EAG base case analysis. The uncertainty around the total QALY loss value (0.00558) has been explored in an EAG threshold analysis.

1b Disutilities associated with having a liver biopsy: serious adverse events

The Stevenson study[80] included a systematic review and an economic evaluation of non-invasive diagnostic tools for the detection of liver fibrosis in patients with alcohol-related liver disease. In the Stevenson study[80] base case analysis it was assumed that serious adverse events were associated with QALY losses of 0.000142 and 0.000254 per patient for percutaneous and transjugular biopsies respectively. The EAG weighted these values by the proportions of NHS patients receiving percutaneous and transjugular biopsies (NHS Reference Costs 2019/20);[79] this led to a QALY loss associated with serious adverse events of 0.000147 per biopsy.

1c Disutilities associated with having a liver biopsy: death

It has been reported that death directly related to percutaneous liver biopsy occurs in a maximum of 1 in 10,000 people biopsied; this value has been used in the EAG model. In line with the population modelled in the Eddowes 2018[29] study, the EAG has assumed that the average age of patients who have a percutaneous liver biopsy is 54 years. Based on average life expectancy in the UK, patients aged 54 years are expected to live a further 32.5 years. However, patients with NAFLD have a lower than average life expectancy, living, on average, 6 years less than the general population.

The age dependent utility value for someone aged 60 in the UK is 0.80. This means that the undiscounted total QALY loss for every biopsy related death is 21.2 (discounted at an annual rate of 3.5% leads to a loss of 14.14 QALYs). Applying a probability of death of 1 in 10,000 people biopsied generates a QALY loss of 0.00141 per biopsy.

2. Failure to treat advanced liver disease

The disutility associated with failure to treat liver disease will depend on the severity of the undiagnosed disease. In NG49,[9] the NICE guideline for the assessment and management of NAFLD, it was assumed that the QALY loss associated with untreated NASH was 0.03. The EAG has applied this QALY loss to the 6-month period before patients with FN LiverMultiScan test results undergo a second LiverMultiScan test.

6.2.11 Summary of base case assumptions

Parameter assumptions and sources used in the base case model are summarised in Table 14.

Table 14 EAG base case model assumptions

Parameter	Assumption	Source/justification
Percentage of patients with a positive LiverMultiScan who go to biopsy	100%	Clinical advice
Percentage of patients with FN results who are retested and correctly diagnosed at 6 months	100%	Conservative assumption that would favour LiverMultiScan (i.e., produce optimistic ICERs per QALY gained for the use of LiverMultiScan)
Time horizon	6 months	Sufficient to capture key differences in costs and benefits between LiverMultiScan plus biopsy and a biopsy only pathways
Discount rate	NA	As model time horizon was under 12 months, no discounting was included in the model
Population prevalence		
Any fibrosis (≥F1)	87.0%	Eddowes 2018/Perspectum Ltd[29,71]
Significant fibrosis (≥F2)	65.2%	
Advanced fibrosis (≥F3)	47.8%	
Brunt Grade ≥1	97.8%	
Brunt Grade ≥2	50.0%	
NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	54.4%	
Advanced NASH (NAS≥4 plus ≥F2)	47.8%	
High risk (NASH or >F1)	79.4%	
LiverMultiScan Test Accuracy		
Sensitivity		
Any fibrosis (≥F1)	0.88	Eddowes 2018/Perspectum Ltd[29,71]
Significant fibrosis (≥F2)	0.63	
Advanced fibrosis (≥F3)	0.64	
Brunt Grade ≥1	0.8	
Brunt Grade ≥2	0.7	
NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	0.64	
Advanced NASH (NAS≥4 plus ≥F2)	0.64	
High risk (NASH or >F1)	0.98	
Specificity		
Any fibrosis (≥F1)	0.67	Eddowes 2018/Perspectum Ltd[29,71]
Significant fibrosis (≥F2)	0.75	
Advanced fibrosis (≥F3)	0.63	
Brunt Grade ≥1	0	
Brunt Grade ≥2	0.7	

NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	0.67	
Advanced NASH (NAS≥4 plus ≥F2)	0.62	
High risk (NASH or >F1)	0.5	
Costs		
Biopsy	£805	Weighted average of YG10Z Percutaneous transvascular biopsy of lesion of liver and YG11A Percutaneous punch biopsy of lesion of liver, 19 years and over from NHS Reference Costs[79]
MRI	£148.24	RD01A Scan of one area, without contrast, 19 years and over from NHS Reference Costs[79]
LiverMultiScan	£199	Cost per scan for data analysis and reporting provided by Perspectum Ltd[71]
Utilities		
QALY losses associated with having a liver biopsy		
Direct pain and anxiety	0.00453	Assumption based upon clinical advice
Serious adverse events	0.000147	Sourced from literature
Death	0.00141	Assumption based upon risk of death from biopsy
Other QALY losses		
QALY loss from failure to treat advanced liver disease	0.03 pa	QALY loss from untreated NASH from NG49[9]

F=stage of fibrosis; FN=false negative; ICER=incremental cost effectiveness ratio; MRI=magnetic resonance imaging; NA=not applicable; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; NG=NICE guideline; QALY=quality adjusted life year

6.2.12 Uncertainty

Uncertainty around parameter values and the impact this could have on cost effectiveness results has been explored by the EAG by running threshold and scenario analyses.

The EAG undertook three threshold analyses:

- LiverMultiScan test results were assumed to be 100% accurate. For each of the diagnostic test strategies, the proportion of patients who would test positive using the reference standard (biopsy) was varied until the LiverMultiScan plus biopsy pathway versus biopsy pathway only was cost effective at a threshold of £20,000 (£30,000) per QALY gained.
- For each of the eight diagnostic test strategies, the QALY loss associated with liver biopsy threshold analysis was varied until the LiverMultiScan plus biopsy pathway versus biopsy pathway only was cost effective at a threshold of £20,000 (£30,000) per QALY gained.
- For each of the eight diagnostic test strategies, the cost at which LiverMultiScan was cost effective at a threshold of £20,000 (£30,000) per QALY gained was estimated.

The EAG also carried out scenario analyses, for all eight diagnostic test strategies, in which the effects of LiverMultiScan failure rates of 0% and 10% were explored.

6.2.13 EAG base case cost effectiveness analysis results

The EAG has generated base case analysis cost effectiveness results for a hypothetical cohort of 1,000 patients with inconclusive results from fibrosis testing. Eight diagnostic test strategies were investigated in the EAG base case analysis:

- T1: Any fibrosis (≥F1)
- T2: Significant fibrosis (≥F2)
- T3: Advanced fibrosis (≥F3)
- T4: Brunt Grade ≥1
- T5: Brunt Grade ≥2
- T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)
- T7: Advanced NASH (NAS≥4 plus ≥F2)
- T8 High risk (NASH or ≥F1).

The EAG base case cost effectiveness analysis results show that there is wide variation between the eight diagnostic test strategies in terms of the number of biopsies that could be avoided if the LiverMultiScan test were introduced into the current diagnostic pathway (minimum: Brunt Grade ≥1 [n=0]; maximum: Brunt Grade ≥2 (n=328.9).

For all eight diagnostic test strategies, the inclusion of the LiverMultiScan test into the pathway increases costs per patient; range: £244 (Brunt Grade ≥2) to £412 (Brunt Grade ≥1).

For six of the diagnostic test strategies (Any fibrosis [≥F1], significant fibrosis [≥F2], advanced fibrosis [≥F3], Brunt Grade ≥1, NASH [NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning] and advanced NASH [NAS≥4 plus ≥F2]), QALY losses were greater for the LiverMultiScan plus biopsy pathway than for the biopsy only pathway. For the remaining two diagnostic test strategies, the QALY losses were greater for the biopsy only pathway.

For six of the diagnostic test strategies (any fibrosis [≥F1], significant fibrosis [≥F2], advanced fibrosis [≥F3], Brunt Grade ≥1, NASH [NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning] and advanced NASH [NAS≥4 plus ≥F2]), the base case ICERs per QALY gained show that the LiverMultiScan plus biopsy pathway is dominated by the biopsy only pathway i.e., the biopsy only pathway is less expensive and leads to fewer QALY losses than the LiverMultiScan plus biopsy pathway.

The two most cost effective diagnostic test strategies are high risk (NASH or >F1) and Brunt Grade ≥2. The ICERs, for these strategies, for the comparison of the LiverMultiScan plus biopsy pathway versus the biopsy only pathway are £749,886 and £1,266,511 per QALY gained respectively. Clinicians suggested that, when considering both these strategies, a

positive result from a LiverMultiScan test would indicate that a patient should be referred for a biopsy. EAG base case cost effectiveness results are provided in Table 15 to Table 19.

Table 15 Initial LiverMultiScan outcomes generated by the EAG model (per 1,000 tests)

Diagnostic test strategy	cT1 cut-off value	True Positive	True Negative	False Positive	False Negative	Failed tests
T1: Any fibrosis (≥F1)	800ms	719.1	82.2	40.6	103.0	55.0
T2: Significant fibrosis (≥F2)	875ms	390.3	246.6	82.2	225.9	55.0
T3: Advanced fibrosis (≥F3)	875ms	287.6	184.9	308.2	164.3	55.0
T4: Brunt Grade ≥1	800ms	739.9	0.0	20.8	185.2	55.0
T5: Brunt Grade ≥2	875ms	328.9	328.9	143.6	143.6	55.0
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	328.9	287.3	143.6	185.2	55.0
T7: Advanced NASH (NAS≥4 plus ≥F2)	875ms	287.3	308.1	185.2	164.4	55.0
T8: High Risk (NASH or >F1)	875ms	729.5	101.1	93.6	20.8	55.0

cT1=iron corrected longitudinal relaxation time; EAG=External Assessment Group; F=stage of fibrosis; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis

Table 16 LiverMultiScan plus biopsy pathway: biopsies performed and averted (per 1,000 patients)

Diagnostic test strategy	cT1 cut-off value	Total number of biopsies, including those following a repeated LiverMultiScan at 6 months	Biopsies averted
T1: Any fibrosis (≥F1)	800ms	917.8	82.2
T2: Significant fibrosis (≥F2)	875ms	753.4	246.6
T3: Advanced fibrosis (≥F3)	875ms	815.1	184.9
T4: Brunt Grade ≥1	800ms	1000.0	0.0
T5: Brunt Grade ≥2	875ms	671.1	328.9
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	712.7	287.3
T7: Advanced NASH (NAS≥4 plus ≥F2)	875ms	691.9	308.1
T8: High Risk (NASH or >F1)	875ms	898.9	101.1

cT1=iron corrected longitudinal relaxation time; F=stage of fibrosis; MRI=magnetic resonance imaging; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis

Table 17 Pathway diagnostic test strategy costs (per 1,000 patients)

Diagnostic test strategy	cT1 cut-off value	LiverMultiScan plus biopsy pathway costs				Biopsy only pathway costs			Additional cost for the LMS pathway
		Biopsy procedures	Biopsy complications	LiverMultiScan test	Total costs	Biopsy procedures	Biopsy complications	Total costs	
T1: Any fibrosis (≥F1)	800ms	£738,817	£7,838	£411,556	£1,158,211	£805,000	£8,540	£813,540	£344,671
T2: Significant fibrosis (≥F2)	875ms	£606,451	£6,434	£511,311	£1,124,195	£805,000	£8,540	£813,540	£310,655
T3: Advanced fibrosis (≥F3)	875ms	£656,163	£6,961	£468,510	£1,131,633	£805,000	£8,540	£813,540	£318,093
T4: Brunt Grade ≥1	800ms	£805,000	£8,540	£411,556	£1,225,096	£805,000	£8,540	£813,540	£411,556
T5: Brunt Grade ≥2	875ms	£540,268	£5,732	£511,311	£1,057,310	£805,000	£8,540	£813,540	£243,770
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	£573,740	£6,087	£511,311	£1,091,137	£805,000	£8,540	£813,540	£277,597
T7: Advanced NASH (NAS≥4 plus ≥F2)	875ms	£557,004	£5,909	£511,311	£1,074,224	£805,000	£8,540	£813,540	£260,684
T8: High Risk (NASH or >F1)	875ms	£723,602	£7,676	£389,570	£1,120,849	£805,000	£8,540	£813,540	£307,309

cT1=iron corrected longitudinal relaxation time; F=stage of fibrosis; LMS=LiverMultiScan; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis

Table 18 QALY analyses for the two diagnostic pathways (per 1,000 patients)

Diagnostic test strategy	cT1 cut-off value	LiverMultiScan plus biopsy pathway					Biopsy only pathway				Incremental QALYs (LMS+biopsy pathway)*
		Biopsy procedure	Biopsy complications	Biopsy death	False negatives	Total QALY losses	Biopsy procedure	Biopsy complications	Biopsy death	Total QALY losses	
T1: Any fibrosis (≥F1)	800ms	5.12	0.13	1.29	1.55	8.10	5.58	0.15	1.41	7.14	-0.96
T2: Significant fibrosis (≥F2)	875ms	4.20	0.11	1.06	3.39	8.76	5.58	0.15	1.41	7.14	-1.63
T3: Advanced fibrosis (≥F3)	875ms	4.55	0.12	1.15	2.47	8.28	5.58	0.15	1.41	7.14	-1.15
T4: Brunt Grade ≥1	800ms	5.58	0.15	1.41	2.78	9.92	5.58	0.15	1.41	7.14	-2.78
T5: Brunt Grade ≥2	875ms	3.74	0.10	0.95	2.15	6.94	5.58	0.15	1.41	7.14	0.19
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	3.98	0.10	1.00	2.78	7.86	5.58	0.15	1.41	7.14	-0.73
T7: Advanced NASH (NAS≥4 plus ≥F2)	875ms	3.86	0.10	0.98	2.47	7.40	5.58	0.15	1.41	7.14	-0.27
T8: High risk (NASH or >F1)	875ms	5.02	0.13	1.27	0.31	6.73	5.58	0.15	1.41	7.14	0.41

* A negative value means that the biopsy only pathway generates more QALYs than LMS+biopsy pathway; a positive value means that the LiverMultiScan plus biopsy pathway generates more QALYs than biopsy only pathway

cT1=iron corrected longitudinal relaxation time; F=stage of fibrosis; LMS=LiverMultiScan; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; QALY=quality adjusted life year

Table 19 Incremental analyses for LiverMultiScan plus biopsy versus biopsy (1,000 patients)

Diagnostic test strategy	cT1 cut-off value	Incremental		ICER per QALY gained (versus biopsy)
		Costs	QALYs	
T1: Any fibrosis (≥F1)	800ms	£344,671	-0.96	LMS+biopsy dominated by biopsy
T2: Significant fibrosis (≥F2)	875ms	£310,655	-1.63	LMS+biopsy dominated by biopsy
T3: Advanced fibrosis (≥F3)	875ms	£318,093	-1.15	LMS+biopsy dominated by biopsy
T4: Brunt Grade ≥1	800ms	£411,556	-2.78	LMS+biopsy dominated by biopsy
T5: Brunt Grade ≥2	875ms	£243,770	0.19	£1,266,511
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875ms	£277,597	-0.73	LMS+biopsy dominated by biopsy
T7: Advanced NASH (NAS≥4 plus ≥F2)	875ms	£260,684	-0.27	LMS+biopsy dominated by biopsy
T8: High risk (NASH or >F1)	875ms	£307,309	0.41	£749,886

cT1=iron corrected longitudinal relaxation time; F=stage of fibrosis; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; QALY=quality adjusted life year

6.2.14 Threshold analyses

Population prevalence

The EAG base case cost effectiveness analyses results showed that if LiverMultiScan test results were 100% accurate, the ICERs for all the diagnostic test strategies would **ONLY** fall below £20,000 (£30,000) per QALY gained if the population prevalence was ≤39.7% (≤45.9%). In the dataset[29] used to populate the model, the diagnostic test strategy with the lowest population prevalence was advanced NASH (NAS≥4 plus ≥F2; 47.8%); however, for this diagnostic test strategy, the accuracy of the LiverMultiScan test was not close to 100% (sensitivity=0.64; specificity=0.62). Clinicians suggested that, when considering this strategy, a positive result from a LiverMultiScan test would result in a patient being referred for a biopsy.

The most cost effective diagnostic test strategy was high risk (NASH or >F1). Clinicians suggested that, when considering this strategy, a positive result from a LiverMultiScan test would result in a patient being referred for a biopsy. The population prevalence for the high risk diagnostic test strategy (79.4%) was not close to the threshold values required for this strategy to be considered cost effective at thresholds of £20,000 (39.7%) or £30,000 (45.9%) per QALY gained; the accuracy of the LiverMultiScan test for this strategy was not close to 100% (sensitivity=0.98; specificity=0.5)

QALY losses associated with each biopsy

The value that QALY losses associated with a biopsy would need to be for the two most cost effective diagnostic test strategies, in the EAG base case analysis, to become cost effective at thresholds of £20,000 and £30,000 per QALY gained are shown in Table 20.

Table 20 QALY loss associated with biopsy: results from threshold analyses

Diagnostic test strategy	Threshold: £20,000 per QALY			Threshold: £30,000 per QALY		
	Original QALY loss	Threshold QALY loss	Increase from original	Original QALY loss	Threshold QALY loss	Increase from original
Brunt Grade ≥2	0.007	0.044	514%	0.007	0.031	340%
High risk (NASH or >F1)	0.007	0.156	2,080%	0.007	0.105	1,368%

NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; ICER=incremental cost effectiveness ratio; QALY=quality adjusted life year

Cost analysis

The EAG threshold cost analysis focused on high risk (NASH or >F1), which was the most cost effective diagnostic test strategy (£749,886 per QALY gained) for the comparison of LiverMultiScan plus biopsy pathway versus biopsy only pathway. If the cost of carrying out a LiverMultiScan test (i.e., MRI and LiverMultiScan) fell from £347.24 to £80.63 (£84.28) per

patient, then the ICER per QALY gained for this comparison would fall to £20,000 (£30,000). This LiverMultiScan cost is substantially below the NHS Reference Cost 2019/20[79] for a single area MRI scan (£148.24).

6.2.15 EAG scenario analyses

A zero failure rate

Compared to the base case analyses (failure rate 5.5%), assuming a LiverMultiScan test failure rate of 0% improved the cost effectiveness of the LiverMultiScan plus biopsy pathway versus the biopsy only pathway for all the diagnostic test strategies considered. However, the LiverMultiScan plus biopsy pathway remained dominated by the biopsy only pathway for any fibrosis stage (≥F1), significant fibrosis (≥F2), advanced fibrosis (≥F3), Brunt Grade ≥1, NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning) and advanced NASH (NAS≥4 plus ≥F2). The high risk (NASH or >F1) test strategy remained the most cost effective diagnostic strategy with the ICER falling from £749,886 to £703,283 per QALY gained. Brunt Grade ≥2 remained the second most cost effective diagnostic strategy with the ICER falling from £1,266,511 to £1,167,286 per QALY gained.

A 10% failure rate

Assuming a 10% LiverMultiScan failure rate reduced the cost effectiveness of the LiverMultiScan plus biopsy pathway versus the biopsy only pathway. However, the LiverMultiScan plus biopsy pathway remained dominated by the biopsy only pathway for any fibrosis stage (≥F1), significant fibrosis (≥F2) and Brunt Grade ≥1, NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning) and advanced NASH (NAS NASH [NAS≥4 plus ≥F2]). The high risk (NASH or >F1) test strategy remained the most cost effective diagnostic strategy with the ICER increasing from £749,886 to £792,252 per QALY gained. Brunt Grade ≥2 remained the second most cost effective diagnostic strategy with the ICER increasing from £1,266,511 to £1,356,715 per QALY gained.

Removal of QALY loss associated with a delayed diagnosis

The EAG carried out a scenario in which there were no QALY losses associated with a delayed diagnosis. Cost effectiveness results from this analysis showed that, the most cost effective diagnostic test strategy was Brunt Grade ≥2 and the ICER per QALY gained was £103,861.

6.2.16 EAG analyses of uncertainty considered and rejected

Probabilistic sensitivity analysis

The EAG model is linear (single node decision tree). Therefore, using probabilistic sensitivity analysis (PSA) to explore the impact of model non-linearity on cost effectiveness results is not required. Further, as the distributions around most of the model inputs are unknown, any PSA would largely be populated with arbitrary choices of distributions, and this would lead to cost effectiveness results that were no more informative than deterministic results.

Deterministic one-way sensitivity analyses

The EAG considered undertaking deterministic one-way sensitivity analyses for the following parameters: sensitivity, specificity, population prevalence and utility values.

The EAG population prevalence threshold analysis showed that the sensitivity and specificity of any diagnostic test strategy could be 100% and the ICER per QALY gained would still be above £30,000. If sensitivity and specificity values were lower than those used in the EAG base case, then this would decrease the cost effectiveness of LiverMultiScan plus biopsy versus biopsy for any diagnostic test strategy. Therefore, varying these DTA parameters in one-way sensitivity analyses would not generate useful results.

The EAG used binomial distributions to construct CIs around base case population prevalence estimates. Results showed that for advanced fibrosis (≥F3), Brunt Grade ≥2 and advanced NASH (NAS≥4 plus ≥F2), the CI lower bounds were 33.4%, 35.6% and 33.4% respectively. For these three diagnostic test strategies, the population prevalence estimates may be low enough that the LiverMultiScan plus biopsy pathway could be cost effective versus the biopsy only pathway; however, the LiverMultiScan plus biopsy pathway could only be cost effective if LiverMultiScan sensitivity and specificity values were 100%. There is no evidence that LiverMultiScan sensitivity and specificity values are both close to 100% for any of the diagnostic test strategies. Further, clinical advice to the EAG was that LiverMultiScan patients classed as high risk (NASH or >F1) (population prevalence was 79.4%), would all be sent for a biopsy; the population prevalence CI lower bound for this population was 67.7%

Results from the EAG utility threshold and scenario analyses showed that plausible changes to QALY losses associated with diagnoses (FN) or biopsies do not change the conclusions that can be drawn from the EAG base case cost effectiveness results. Therefore, varying utility values in sensitivity analyses would not generate useful results.

Alternative sources of population prevalence data

Population prevalence data were only available, by diagnosis, from the Eddowes 2018[29] study for patients with inconclusive results from previous fibrosis testing who were scheduled for and received a biopsy (i.e., a subgroup of the population described in the final scope[23] issued by NICE). Population prevalence estimates are independent of the diagnostic test used (LiverMultiScan or MRE) as they are generated from biopsy results only. Population prevalence data were available from other populations, however, the population prevalence for the same diagnoses varied significantly. For example, for the diagnosis of significant fibrosis (≥F2) in populations with suspected NASH who were sent for a biopsy, population prevalence estimate calculated using Imajo 2021[56] study data was approximately 75%, whereas the estimate calculated using Kim[58] 2020 study data was 43.6%. Neither of these estimates are more suitable than the value from the Eddowes 2018[29] study used in the EAG model as they do not specifically relate to the patients described in the final scope[23] issued by NICE. However, the disparity between the estimates calculated using values from these two studies[56,58] highlights that there may be uncertainty around the population prevalence estimates calculated from Eddowes 2018[29] study data; other studies carried out in the same population may lead to substantially different population prevalence estimates.

Alternative sources of DTA data

It would be possible for the EAG to use DTA data from patients who did not have indeterminate results from fibrosis testing but who did have a LiverMultiScan or MRE in the EAG model, for example, data from the Imajo 2021[56] or Kim 2020[58] studies. Results from the Imajo 2021[56] study suggest that in a population not described in the final scope[23] issued by NICE, MRE is generally more sensitive and less specific than LiverMultiScan.

However, populating the EAG with different DTA data would not the change the conclusions that can be drawn from EAG base case cost effectiveness results as threshold analysis showed that even if tests were 100% accurate, it is unlikely that ICERs would fall below £30,000 per QALY gained using the best available population prevalence estimates. Therefore, the EAG did not consider analyses using LiverMultiScan sensitivity and specificity estimates from other sources or analyses using published MRE sensitivity and sensitivity estimates.

The potential impact of MRI-based technology use for patients who will not receive a biopsy

There are no population prevalence or DTA data for patients with indeterminate results from previous fibrosis testing who would not be sent for a biopsy. Clinical advice to the EAG is that patients with indeterminate results from previous fibrosis testing are referred for a biopsy

unless there are clear reasons for not doing so, for example, presence of co-morbidities, personal choice, old age and medical contraindications. If these patients were to receive a LiverMultiScan, cT1 and PDFF results would be available; however, this information is unlikely to influence treatment decisions and the reasons for not referring these patients for biopsy will remain despite access to LiverMultiScan results. Further, there are no specific population prevalence, sensitivity or specificity data (LiverMultiScan or MRE) for these patients. The only parameter values that could be used in this analysis would be the EAG base case parameter values.

Assumption that all patients with a positive LiverMultiScan results are referred for a biopsy

Based on clinical advice, including that from a Specialist Committee member, the EAG has assumed that all patients with a positive result from a LiverMultiScan test would be referred for a biopsy. Without further information about why patients with a positive LiverMultiScan test result are not sent for a biopsy, it is impossible to make informed variations to the EAG model to accommodate a pathway in which patients who are identified as needing a biopsy (TP and FP) are not referred for a biopsy.

Extend model 6 month time horizon

If the EAG model time horizon were extended beyond 6 months, then this would reduce the cost effectiveness of LiverMultiScan due to the increased QALY losses associated with missed diagnoses that would be accrued, and the increased costs associated with further diagnostic tests.

6.3 EAG cost effectiveness discussion

Clinical advice to the EAG is that LiverMultiScan (or MRE) does not provide the level of detailed information that may be required to make treatment decisions, for example, clinical features that suggest additional cofactors for liver injury; this information is only available from a biopsy. Results from the EAG cost effectiveness analyses showed that, for patients with inconclusive results from previous fibrosis testing, LiverMultiScan (or MRE) can, potentially, identify patients for whom a biopsy is not necessary and reduce the proportion of patients who have an unnecessary biopsy.

The Eddowes 2018[29] study evidence suggests that, regardless of the diagnostic test strategy used, the proportion of patients with inconclusive results from fibrosis testing who would require a biopsy means that the LiverMultiScan plus biopsy pathway is unlikely to be cost effective versus biopsy using a willingness to pay threshold of £30,000 per QALY gained. For six of the eight diagnostic test strategies considered, LiverMultiScan plus biopsy pathway was

dominated by the biopsy only pathway. Threshold analysis showed that even when assuming that the LiverMultiScan test was 100% accurate, the population prevalence, for any of the eight diagnostic test strategies, would have to be significantly lower than suggested by evidence from the Eddowes 2018[29] study. Therefore MRE, although potentially more accurate than LiverMultiScan, is unlikely to have an ICER below £30,000 per QALY gained.

The EAG cost effectiveness analyses are limited to eight diagnostic test strategies proposed by Eddowes 2018/Perspectum Ltd.[29,71] It is not known whether all the diagnostic strategies would be acceptable to clinicians working in NHS practice. In response to a question from the EAG, one Specialist Committee member identified four of the eight strategies (T3, T5, T7 and T8) where a positive LiverMultiScan test result would mean that they would still refer the patient for a biopsy.

EAG cost effectiveness results for the LiverMultiScan plus biopsy pathway are optimistic as they have been generated using the assumption that patients will be correctly diagnosed following a maximum of two LiverMultiScan tests. Any deviation from this assumption would decrease the cost effectiveness of the LiverMultiScan plus biopsy pathway versus the biopsy pathway.

The EAG base case cost effectiveness results should be used with caution due to the limited DTA and population prevalence data available to populate the model; the only relevant DTA and population prevalence estimates are from a small study (n=46 patients).[29] This is of concern as, in a different population to that described in the final scope[23] issued by NICE, population prevalence estimates for a specific diagnosis that were calculated using data from two studies[56,58] were different. Despite this limitation, EAG model results are informative and provide an indication of the likely cost effectiveness of LiverMultiScan and MRE (despite the absence of evidence on test accuracy for MRE in the scope[23] population).

7 DISCUSSION

7.1 Statement of principal findings

7.1.1 Diagnostic test accuracy

In line with the final scope[23] issued by NICE, the 13 studies[29,53-64] included in the DTA review considered patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, no studies were identified that provided evidence for the DTA of MRI-based technologies for patients with NAFLD for whom TE or ARFI was unsuitable. Of the 13 studies[29,53-64] that were included in the DTA review, the EAG was confident that only one study[29] provided evidence for the DTA of MRI-based technologies for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and who had indeterminate or discordant results from fibrosis testing; the Eddowes 2018[29] study evaluated LiverMultiScan and reported both PDFF and cT1 outputs. When assessing study quality, for most of the risk of bias and applicability concerns domains, the EAG considered that most studies had low risk of bias. For diagnosis of fibrosis, sensitivity ranged from 50% to 88% and specificity ranged from 42% to 75%. Sensitivity and specificity values for fibrosis testing were consistently higher when using LiverMultiScan cT1 data than when using LiverMultiScan PDFF.

Data from three studies were included in the meta-analyses for LiverMultiScan. For fibrosis (≥F2 and ≥F3), the pooled sensitivity and specificity values were higher for LiverMultiScan cT1 than for LiverMultiScan PDFF. For steatosis (Brunt Grade ≥1), the meta-analysis results suggested that LiverMultiScan cT1 had greater sensitivity than specificity. The steatosis (Brunt Grade ≥2) results for LiverMultiScan PDFF were fairly consistent with those for LiverMultiScan cT1. For NASH and advanced NASH, meta-analysis results were broadly similar between the LiverMultiScan cT1 and LiverMultiScan PDFF outputs, with the exception of sensitivity for detecting advanced NASH (LiverMultiScan cT1: 66.0%; LiverMultiScan PDFF: 49.4%). All other estimates of sensitivity and specificity ranged from 58.0% to 73.7%.

The sensitivity (fibrosis ≥F2) and specificity (fibrosis ≥F1 and ≥F2) reported for MRE in the four individual studies[56-58,62] identified by the EAG were consistently greater when compared to those observed with LiverMultiScan. For fibrosis (≥F2) the sensitivity of MRE ranged from 82% to 95% and specificity ranged from 85% to 100%. For fibrosis (≥F3) the sensitivity of MRE ranged from 71% to 100% and specificity ranged from 79% to 93%. Data from three studies[56-58] were used to estimate a summary ROC curve for MRE for advanced fibrosis (≥F3). The summary ROC indicated high DTA but not all observed study results lay close to the curve. The sensitivity and specificity observed in the two studies[57,58] that used the Resoundant, Inc. MRE platform that is commercially available ranged from 85% to 100% and from 92% to 93%,

respectively. The EAG notes that the DTA results for MRE are for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, the studies did not specify whether these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.

7.1.2 Clinical impact

Eleven studies[29,53,54,57,59,62,64,66-69] evaluated the clinical impact of MRI-based technologies for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. As in the DTA review, no studies were identified that provided evidence for the clinical impact of MRI-based technologies for patients with NAFLD for whom TE or ARFI was unsuitable. Only one study[29] provided evidence for the clinical impact of MRI-based technologies for patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and who had indeterminate or discordant results from fibrosis testing.

The two studies[66,67] that evaluated the prognostic ability of MRI-based technologies included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, the studies[66,67] also included patients with other liver disease aetiologies and did not present results specifically for patients with NAFLD.

One study[68] reported that LiverMultiScan could [redacted] the number of unnecessary biopsies for patients with [redacted] and [redacted] when compared to standard care.

Test failure rate in a population of patients with NAFLD was reported in four studies.[29,53,57,62] The test failure rate of the index tests for patients with NAFLD was 5.6%[29] for LiverMultiScan and ranged from 3.9%[57] to 7.6%[53] for MRE. The test failure rate of MRE for patients with NAFLD was estimated by the EAG meta-analysis to be 4.2% (95% CI: 2.5% to 6.2%).

Acceptability of LiverMultiScan from patient feedback was generally positive.[69] Patients considered the MRI scan was a painless and comfortable procedure and many highlighted that the 'non-invasive' element of the procedure was important.[69]

No studies were identified that evaluated the remaining clinical impact outcomes specified in the final scope[23] issued by NICE (Table 2).

7.1.3 Cost effectiveness

Eddowes 2018[29] study clinical effectiveness data were collected from a population with inconclusive results from previous fibrosis testing and used to populate the Blake 2016[78] model. However, the Blake 2016[78] model was not designed to explore cost effectiveness for

patients with inconclusive results from previous fibrosis testing. Therefore, the Eddowes 2018[29] study cost savings estimates are not relevant to this appraisal.

The EAG developed a de novo economic model that enabled a comprehensive assessment (eight different diagnostic test strategies) of the cost effectiveness of two different diagnostic pathways: LiverMultiScan plus biopsy versus biopsy only. The base case ICER per QALY gained results for six diagnostic pathways showed that LiverMultiScan plus biopsy was dominated by biopsy only and, for the other two diagnostic pathways, the ICERs per QALY gained were £749,886 and £1,266,511. The results from the EAG threshold and scenario analyses demonstrated that these results were robust to plausible variations in the magnitude of key parameters.

7.2 Strengths and limitations of the assessment

7.2.1 Strengths of the assessment

This assessment is the first to evaluate the DTA, clinical impact and cost effectiveness of MRI-based technologies for three groups of patients with NAFLD for whom advanced fibrosis or cirrhosis has not yet been diagnosed, namely (i) patients with indeterminate results from fibrosis testing, (ii) patients who are unsuitable for testing with TE or ARFI and (iii) patients with discordant results from fibrosis testing. The clinical and cost effectiveness systematic review processes included extensive literature searches and followed best practice recommendations.[45-48]

Perspectum Ltd[71] has provided DTA data that were not previously available from published sources. These DTA data could allow LiverMultiScan outputs to be used to inform treatment decisions for patients with NAFLD (eight different diagnostic test strategies). The EAG used these data, as well published data, to carry out quantitative analyses.

A key strength of the EAG economic evaluation is that the de novo model provides a simple, flexible framework that allows the comparison of eight different diagnostic strategies. It is based on the best available DTA and population prevalence evidence (identified through the systematic review and provided by Perspectum Ltd) and captures the trade-off between high upfront costs of diagnostic tests and the reduction in subsequent biopsies that they may offer. The model design captures all of the main factors that are relevant to the decision problem. It is user friendly and calculations are transparent. Furthermore, the model can easily be updated to incorporate new DTA and population prevalence evidence if they become available.

7.2.2 Limitations of the assessment

The DTA and population prevalence data available from Eddowes 2018/Perspectum Ltd[29,71] are from patients with inconclusive results from previous fibrosis testing. The EAG has assumed that inconclusive is an umbrella term that includes the three subgroups of patients described in the final scope[23] issued by NICE; however, the EAG is not confident that the term inconclusive includes patients for whom TE and ARFI are unsuitable.

The EAG quantitative synthesis only included data from six studies.[29,56-59,62] Furthermore, the meta-analyses were populated with data from small numbers of studies and only one[29] of the studies included the population that is the subject of this assessment. This should be considered when interpreting results from the EAG meta-analyses.

Data on the clinical impact of MRI-based technologies was scarce for some outcomes (prognostic ability, number of liver biopsies and test failure rate). No data were available for the remaining clinical outcomes listed in the in the final scope[23] issued by NICE.

Eddowes 2018/Perspectum Ltd[29,71] provided LiverMultiScan DTA data for the relevant population. These data were included in the EAG DTA review and were used to inform the EAG economic model. However, Resoundant, Inc. did not provide any MRE DTA evidence for the relevant population and therefore MRE could not be considered as a comparator in the EAG economic model, although the cost effectiveness of MRE can be inferred from the model results, i.e., MRE is unlikely to be cost effective in the population described in the final scope[23] issued by NICE (using data from Eddowes 2018/Perspectum Ltd[29,71]) even if test accuracy was 100%.

In the EAG model, LiverMultiScan is positioned as a triage test, i.e., LiverMultiScan would be added to the current NHS diagnostic pathway to avoid a more invasive downstream test (biopsy). The LiverMultiScan test is not 100% sensitive or specific for any of the eight diagnostic test strategies considered; the levels of sensitivity and specificity required to provide clinicians with sufficient confidence to use LiverMultiScan test results for patients described in the final scope[23] issued by NICE is not known.

Potentially, different proportions of patients with advanced disease will receive a LiverMultiScan test FN result depending on the diagnostic test strategy used. If this did occur, the average impact of a FN result (costs and, notably, QALY losses) would vary depending on diagnostic test strategy used. The inability to resolve this issue is unlikely to be a major limitation of the EAG analyses as results from an EAG scenario analysis that removed the

QALY loss associated with a LiverMultiScan test FN result showed that the conclusions that can be drawn from the EAG base case cost effectiveness analyses results did not change.

7.3 Uncertainties

There is substantial evidence on the DTA of MRI-based technologies for liver related conditions. However, there is limited DTA, clinical impact and cost effectiveness data for patients who have indeterminate results from fibrosis testing, for whom TE or ARFI are unsuitable or patients who have discordant results from fibrosis testing.

The clinical value of MRI-based technologies to support decision making for the clinical management of NAFLD and to improve the uptake and maintenance of lifestyle modifications remains uncertain. It is plausible that use of MRI-based technologies may inform the target area for a liver biopsy, however no evidence is available to suggest that MRI-based technologies would be used for that purpose. The clinical impact of MRI-based technologies on intermediate, clinical and patient-reported outcomes also remains uncertain. The RADIcAL trial[68] evaluated the clinical impact of LiverMultiScan for patients with suspected NAFLD (completed December 2020) reported the number of liver biopsies avoided by using LiverMultiScan. However, only a small proportion of patients recruited to the trial contributed data to this analysis. It is unclear if the patients included in the RADIcAL trial[68] consisted of those who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing. The clinical value of LiverMultiScan to help avoid unnecessary biopsies therefore remains uncertain.

If the population prevalence estimate calculated using data from the 46 patients in the Eddowes 2018[29] study reflects the population prevalence of patients treated in NHS clinical practice in England and Wales, then the EAG cost effectiveness results are certain. However, if the population prevalence in NHS clinical practice is different, then results from the EAG cost effectiveness results will no longer be valid.

7.4 Conclusions

7.4.1 Clinical effectiveness

MRI-based technologies may be useful to identify patients that may benefit from additional testing in the form of liver biopsy and those for whom this additional testing may not be necessary. However, there is a paucity of DTA and clinical impact data for a population that may benefit from implementation of this technology, namely patients with indeterminate or discordant results from previous fibrosis testing or patients for TE and ARFI are not suitable.

7.4.2 Cost effectiveness

Only one small LiverMultiScan study[29] provided DTA and population prevalence data for patients described in the final scope[23] issued by NICE. It is unclear whether sensitivity and specificity estimates reported by this small study[29] will give clinicians sufficient confidence to use LiverMultiScan test results to triage patients with inconclusive results from previous fibrosis testing to biopsy. Cost effectiveness results from the EAG model are only informative if clinicians have confidence in LiverMultiScan DTA data. Using the available DTA and population prevalence data, EAG cost effectiveness results showed that LiverMultiScan is unlikely to be cost effective at current prices when used to triage patients with inconclusive results from previous fibrosis testing to biopsy.

LiverMultiScan data are not available for patients for whom TE or ARFI were unsuitable. Further, no MRE DTA data were available for the population described in the final scope[23] issued by NICE. The EAG was unable to generate cost effectiveness results for this technology; however, even if MRE was 100% accurate, due to high population prevalence estimates, it is unlikely that MRE would be cost effective at current prices.

7.4.3 Implications for service provision

If LiverMultiScan were to be recommended by NICE, the implications for NHS service provision would be significant due to the increased staffing levels and changes in infrastructure that would be required to accommodate the high demand for MRI scans for patients with NAFLD.

7.4.4 Suggested research priorities

Only Eddowes 2018/Perspectum Ltd[29,71] provided data for a relevant population. Other published studies may also have included these patients; however, this information was not available from the published studies. If, in future, information about results from previous fibrosis testing could be recorded at the time of study enrolment, study DTA results from individual patients or subgroups could be used to inform treatment decisions.

Large, prospective studies are required to assess the DTA and clinical impact of MRI-based technologies for the population of interest to this assessment. DTA studies should ideally use liver biopsy as the reference standard, pre-specify and justify the use of thresholds, and use the MRI-based technologies as intended in routine clinical practice. Prospective database registries could be set up to collect long-term data (prognostic ability, reduction or remission of liver fibrosis or fibro-inflammation, reduction or remission of liver fat, mortality, and morbidity outcomes) from patients with NAFLD who undergo non-invasive diagnostic procedures. Qualitative studies are required to investigate the impact of MRI-based technology test results

on clinical decision making, their potential to influence the uptake and maintenance of lifestyle modifications and the acceptability of the technologies to patients.

8 ACKNOWLEDGEMENTS

The authors are grateful to Gideon Hirschfield (Chair in Autoimmune Liver Disease, Toronto General Hospital, Canada) for clinical advice and comments on a draft version of the EAG report. The authors would like to thank Yemisi Takwoingi (Professor in Test Evaluation and Evidence Synthesis, University of Birmingham) for advice given on statistical analysis methods for assessment of diagnostic test accuracy and Chris Hyde (Professor of Public Health and Clinical Epidemiology, University of Exeter) for comments on a draft version of the EAG report.

8.1 Contributions of authors

All authors contributed to the conception and design of the study or the analysis and interpretation of the data, drafting or revising the report, and final approval of the version to be published.

Rui Duarte (Deputy Director, LR*i*G, Health Technology Assessment Lead) managed the project, contributed to the development of the methods for the systematic review, conducted the review of diagnostic test accuracy and clinical impact and supervised the statistical analysis and economic modelling work.

Rebecca Bresnahan (Research Associate, Clinical Effectiveness) conducted the systematic review of diagnostic test accuracy and clinical impact and acted as the first reviewer in the systematic review.

James Mahon (Director, Coldingham Analytical Services, Health Economics and Modelling) developed the health economic model, identified inputs to the economic model, and conducted the economic evaluation.

Sophie Beale (Director, Hare Research, Health Economics and Modelling) provided input to the health economic model and provided senior advice to the project.

Angela Boland (Director, LR*i*G, Health Economics and Modelling) provided input to the health economic model and provided senior advice to the project.

Marty Chaplin (Research Associate, Statistician) contributed to the statistical analysis methods, performed the statistical analysis for the systematic review of diagnostic test accuracy and clinical impact.

Devarshi Bhattacharyya (Health Economic Modeller, Health Economics and Modelling) conducted the review of cost effectiveness evidence.

Rachel Houten (Health Economic Modeller, Health Economics and Modelling) contributed to the review of cost effectiveness evidence.

Katherine Edwards (Senior Research Fellow, Systematic Reviewer) acted as the second reviewer in the systematic review.

Sarah Nevitt (Research Associate, Statistician) contributed to the statistical analysis for the diagnostic test accuracy review.

Michelle Maden (Research Associate, Information Specialist) devised and performed the literature searches.

9 REFERENCES

1. National Health Service. Non-alcoholic fatty liver disease (NAFLD). Published 13 January 2022; Available from: www.nhs.uk/conditions/non-alcoholic-fatty-liver-disease/. Accessed 20 January 2022.
2. Bedossa P. Diagnosis of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: Why liver biopsy is essential. Liver Int. 2018; 38 Suppl 1:64-6.
3. West J, Card TR. Reduced mortality rates following elective percutaneous liver biopsies. Gastroenterology. 2010; 139:1230-7.
4. Kleiner DE, Brunt EM. Nonalcoholic fatty liver disease: Pathologic patterns and biopsy evaluation in clinical research. Semin Liver Dis. 2012; 32:3-13.
5. Gonzalez HC, Jafri SM, Gordon SC. Role of liver biopsy in the era of direct-acting antivirals. Curr Gastroenterol Rep. 2013; 15:307.
6. British Liver Trust. The alarming impact of liver disease in the UK. Facts and statistics. Published June 2019; Available from: www.britishlivertrust.org.uk/wp-content/uploads/The-alarming-impact-of-liver-disease-FINAL-June-2019.pdf. Accessed 8 December 2021.
7. British Liver Trust. NAFLD, NASH and fatty liver disease. Causes and symptoms. Published 9 June 2021; Available from: www.britishlivertrust.org.uk/information-and-support/living-with-a-liver-condition/liver-conditions/non-alcohol-related-fatty-liver-disease/#causes. Accessed 15 December 2021.
8. Ratziu V, Bellentani S, Cortez-Pinto H, Day C, Marchesini G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. J Hepatol. 2010; 53:372-84.
9. National Institute for Health and Care Excellence. Non-alcoholic fatty liver disease. Assessment and management. Published 6 July 2016; Available from: www.nice.org.uk/guidance/ng49/evidence/full-guideline-pdf-2548213310. Accessed 8 December 2021.
10. Scapaticci S, D'Adamo E, Mohn A, Chiarelli F, Giannini C. Non-alcoholic fatty liver disease in obese youth with insulin resistance and type 2 diabetes. Front Endocrinol. 2021; 12:639548.
11. Dyson JK, Anstee QM, McPherson S. Non-alcoholic fatty liver disease: A practical approach to diagnosis and staging. Frontline Gastroenterol. 2014; 5:211.
12. Morgan A, Hartmanis S, Tsochatzis E, Newsome PN, Ryder SD, Elliott R*, et al.* Disease burden and economic impact of diagnosed non-alcoholic steatohepatitis (NASH) in the United Kingdom (UK) in 2018. Eur J Health Econ. 2021; 22:505-18.
13. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: A systematic review and meta-analysis of paired-biopsy studies. Clin Gastroenterol Hepatol. 2015; 13:643-54.e9.
14. Taylor RS, Taylor RJ, Bayliss S, Hagström H, Nasr P, Schattenberg JM*, et al.* Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: A systematic review and meta-analysis. Gastroenterology. 2020; 158:1611-25.e12.
15. Stål P. Liver fibrosis in non-alcoholic fatty liver disease - diagnostic challenge with prognostic significance. World J Gastroenterol. 2015; 21:11077-87.
16. Tapper EB, Parikh ND. Mortality due to cirrhosis and liver cancer in the United States, 1999-2016: Observational study. BMJ. 2018; 362:k2817.
17. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW*, et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005; 41:1313-21.
18. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z*, et al.* Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017; 65:1557-65.

19. de Alwis NMW, Day CP. Non-alcoholic fatty liver disease: The mist gradually clears. J Hepatol. 2008; 48:S104-S12.
20. Zubair R, Mirza M, Iftikhar J, Saeed N. Frequency of incidental fatty liver on ultrasound and its association with diabetes mellitus and hypertension. Pak J Med Sci. 2018; 34:1137-41.
21. Horn P, Newsome PN. NAFLD – diagnosis, assessment and management. Published 17 November 2021 Available from: www.bsg.org.uk/clinical-articles-list/nafld-diagnosis-assessment-and-management/. Accessed 8 December 2021.
22. Byrne CD, Patel J, Scorletti E, Targher G. Tests for diagnosing and monitoring non-alcoholic fatty liver disease in adults. BMJ. 2018; 362:k2734.
23. National Institute for Health and Care Excellence. MRI-based technologies for the assessment of non-alcoholic fatty liver disease: Final scope. Published September 2021; Available from: www.nice.org.uk/guidance/gid-dg10045/documents/final-scope. Accessed 22 November 2021.
24. Williams R, Alessi C, Alexander G, Allison M, Aspinall R, Batterham RL*, et al.* New dimensions for hospital services and early detection of disease: A review from the Lancet Commission into liver disease in the UK. Lancet. 2021; 397:1770-80.
25. Jarvis H, Worsfold J, Hebditch V, Ryder S. Engagement with community liver disease management across the UK: A cross-sectional survey. BJGP Open. 2021:BJGPO.2021.0085.
26. Smeuninx B, Boslem E, Febbraio MA. Current and future treatments in the fight against non-alcoholic fatty liver disease. Cancers. 2020; 12:1714.
27. National Institute for Health and Care Excellence. Obesity: Identification, assessment and management. Published 6 July 2016; Available from: https://www.nice.org.uk/guidance/cg189. Accessed 14 December 2021.
28. Bazick J, Donithan M, Neuschwander-Tetri BA, Kleiner D, Brunt EM, Wilson L*, et al.* Clinical model for NASH and advanced fibrosis in adult patients with diabetes and NAFLD: Guidelines for referral in NAFLD. Diabetes Care. 2015; 38:1347-55.
29. Eddowes PJ, McDonald N, Davies N, Semple SIK, Kendall TJ, Hodson J*, et al.* Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther. 2018; 47:631-44.
30. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol. 2021; 75:659-89.
31. McDonald N, Eddowes PJ, Hodson J, Semple SIK, Davies NP, Kelly CJ*, et al.* Multiparametric magnetic resonance imaging for quantitation of liver disease: A two-centre cross-sectional observational study. Sci Rep. 2018; 8:9189.
32. National Institute for Health and Care Excellence. LiverMultiScan for liver disease. Medtech innovation briefing. Published 26 April 2019; Available from: https://www.nice.org.uk/advice/mib181/resources/livermultiscan-for-liver-disease-pdf-2285963692990405. Accessed 22 February 2022.
33. Pavlides M, Banerjee R, Sellwood J, Kelly CJ, Robson MD, Booth JC*, et al.* Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. J Hepatol. 2016; 64:308-15.
34. Perspectum Diagnostics Ltd. Data on file. LiverMultiScan® Case for Adoption. July 2021.
35. Perspectum Diagnostics Ltd. LiverMultiScan. A guide to interpreting liver tissue characterisation for clinicians. Published July 2019; Available from: www.perspectum.com/media/1731/a-guide-to-interpreting-liver-tissue-characterisation-for-clinicians_pdm135-1-1.pdf. Accessed 8 December 2021.
36. Venkatesh SK, Yin M, Ehman RL. Magnetic resonance elastography of liver: Clinical applications. J Comput Assist Tomogr. 2013; 37:887-96.
37. Venkatesh SK, Ehman RL. Magnetic resonance elastography of liver. Magn Reson Imaging Clin N Am. 2014; 22:433-46.

38. Dulai PS, Sirlin CB, Loomba R. MRI and MRE for non-invasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: Clinical trials to clinical practice. J Hepatol. 2016; 65:1006-16.
39. National Institute for Health and Care Excellence. Cirrhosis in over 16s: Assessment and management. Published 6 July 2016; Available from: www.nice.org.uk/guidance/ng50/resources/cirrhosis-in-over-16s-assessment-and-management-pdf-1837506577093. Accessed 10 December 2021.
40. Resoundant Inc. Measure what matters. MR elastography for clinical trials. Published 2021; Available from: www.resoundant.com/rct-white-paper-download. Accessed 17 March 2022.
41. Manduca A, Bayly PJ, Ehman RL, Kolipaka A, Royston TJ, Sack I, *et al.* MR elastography: Principles, guidelines, and terminology. Magn Reson Med. 2021; 85:2377-90.
42. Neuberger J, Patel J, Caldwell H, Davies S, Hebditch V, Hollywood C, *et al.* Guidelines on the use of liver biopsy in clinical practice from the British Society of Gastroenterology, the Royal College of Radiologists and the Royal College of Pathology. Gut. 2020; 69:1382-403.
43. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology. 2005; 128:1898-906.
44. Pavlides M, Birks J, Fryer E, Delaney D, Sarania N, Banerjee R, *et al.* Interobserver variability in histologic evaluation of liver fibrosis using categorical and quantitative scores. Am J Clin Pathol. 2017; 147:364-9.
45. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking systematic reviews in health care. Published January 2009; Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm. Accessed 8 December 2021.
46. National Institute for Health and Care Excellence. Diagnostic assessment programme manual. Published December 2011; Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf. Accessed 15 December 2021.
47. Cochrane Diagnostic Test Accuracy Working Group. Cochrane handbook for systematic reviews of diagnostic test accuracy. Published 2009; Available from: http://srdta.cochrane.org/handbook-dta-reviews. Accessed 7 December 2021.
48. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, *et al.* Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. JAMA. 2018; 319:388-96.
49. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155:529-36.
50. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al.* RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ. 2019; 366:l4898.
51. National Heart Lung and Blood Institute (NHLBI). Study quality assessment tools. Published July 2021; Available from: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed 10 December 2021.
52. Critical Appraisal Skills Programme. CASP Qualitative Studies Checklist. Published 2018; Available from: www.casp-uk.b-cdn.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf. Accessed 17 March 2022.
53. Caussy C, Chen J, Alquiraish MH, Cepin S, Nguyen P, Hernandez C, *et al.* Association between obesity and discordance in fibrosis stage determination by magnetic resonance vs transient elastography in patients with nonalcoholic liver disease. Clin Gastroenterol Hepatol. 2018; 16:1974-82.e7.

54. Forsgren MF, Nasr P, Karlsson M, Dahlstrom N, Noren B, Ignatova S*, et al.* Biomarkers of liver fibrosis: Prospective comparison of multimodal magnetic resonance, serum algorithms and transient elastography. Scand J Gastroenterol. 2020; 55:848-59.
55. Hoffman DH, Ayoola A, Nickel D, Han F, Chandarana H, Shanbhogue KP. T1 mapping, T2 mapping and MR elastography of the liver for detection and staging of liver fibrosis. Abdom Radiol. 2020; 45:692-700.
56. Imajo K, Tetlow L, Dennis A, Shumbayawonda E, Mouchti S, Kendall TJ*, et al.* Quantitative multiparametric magnetic resonance imaging can aid non-alcoholic steatohepatitis diagnosis in a Japanese cohort. WJG. 2021; 27:609-23.
57. Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: Noninvasive assessment with MR elastography. Radiology. 2013; 268:411-9.
58. Kim JW, Lee Y-S, Park YS, Kim B-H, Lee SY, Yeon JE*, et al.* Multiparametric MR index for the diagnosis of non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. Sci Rep. 2020; 10:2671.
59. Pavlides M, Banerjee R, Tunnicliffe EM, Kelly C, Collier J, Wang LM*, et al.* Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int. 2017; 37:1065-73.
60. Sofue K, Onoda M, Tsurusaki M, Morimoto D, Yada N, Kudo M*, et al.* Dual-frequency MR elastography to differentiate between inflammation and fibrosis of the liver: Comparison with histopathology. J Magn Reson Imaging. 2020; 51:1053-64.
61. Toguchi M, Tsurusaki M, Hyodo T, Numoto I, Matsuki M, Imaoka I*, et al.* Magnetic resonance elastography in the assessment of hepatic fibrosis: A study comparing transient elastography and histological data in the same patients. Abdom Radiol. 2017; 42:1659-66.
62. Troelstra MA, Witjes JJ, van Dijk A-M, Mak AL, Gurney-Champion O, Runge JH*, et al.* Assessment of imaging modalities against liver biopsy in nonalcoholic fatty liver disease: The Amsterdam NAFLD-NASH cohort. JMRI. 2021.
63. Trout AT, Serai SD, Sheridan RM, Xanthakos SA, Zhang B, Su W*, et al.* Diagnostic performance of MR elastography for liver fibrosis in children and young adults with a spectrum of liver diseases. Radiology. 2018; 287:824-32.
64. Xanthakos SA, Podberesky DJ, Serai SD, Miles L, King EC, Balistreri WF*, et al.* Use of magnetic resonance elastography to assess hepatic fibrosis in children with chronic liver disease. J Pediat. 2014; 164:186-8.
65. Banerjee R, Pavlides M, Tunnicliffe EM, Piechnik SK, Sarania N, Philips R*, et al.* Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol. 2014; 60:69-77.
66. Jayaswal ANA, Levick C, Selvaraj EA, Dennis A, Booth JC, Collier J*, et al.* Prognostic value of multiparametric magnetic resonance imaging, transient elastography and blood-based fibrosis markers in patients with chronic liver disease. Liver Int. 2020; 40:3071-82.
67. Gidener T, Yin M, Dierkhising RA, Allen AM, Ehman RL, Venkatesh SK. Magnetic resonance elastography for prediction of long-term progression and outcome in chronic liver disease: A retrospective study. Hepatology. 2022; 75:379-90.
68. Perspectum Diagnostics Ltd. Data on file. Non-invasive rapid assessment of patients with non-alcoholic fatty liver disease (NAFLD) using magnetic resonance imaging with LiverMultiScan^{TM}: RADIcAL-1 clinical data report. 30 December 2020.
69. McKay A, Pantoja C, Hall R, Matthews S, Spalding P, Banerjee R. Patient understanding and experience of non-invasive imaging diagnostic techniques and the liver patient pathway. JPRO. 2021; 5.
70. Resoundant Inc. Data on file. Resoundant comments. List of studies for inclusion MRE. January 2022.
71. Perspectum Diagnostics Ltd. Data on file. Request for information. December 2021.

72. Selvaraj EA, Mózes FE, Jayaswal ANA, Zafarmand MH, Vali Y, Lee JA*, et al.* Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: A systematic review and meta-analysis. Supplementary material. J Hepatol. 2021; 75:770-85.
73. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F*, et al.* Histological grading and staging of chronic hepatitis. J Hepatol. 1995; 22:696-9.
74. Tonev D, Shumbayawonda E, Tetlow LA, Herdman L, French M, Rymell S*, et al.* The effect of multi-parametric magnetic resonance imaging in standard of care for nonalcoholic fatty liver disease: Protocol for a randomized control trial. JMIR Res Protoc. 2020; 9:e19189.
75. Zelber-Sagi S, Bord S, Dror-Lavi G, Smith ML, Towne SD, Jr., Buch A*, et al.* Role of illness perception and self-efficacy in lifestyle modification among non-alcoholic fatty liver disease patients. World J Gastroenterol. 2017; 23:1881-90.
76. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford: Oxford University Press; 2015.
77. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D*, et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Value Health. 2013; 16:e1-5.
78. Blake L, Duarte RV, Cummins C. Decision analytic model of the diagnostic pathways for patients with suspected non-alcoholic fatty liver disease using non-invasive transient elastography and multiparametric magnetic resonance imaging. BMJ Open. 2016; 6:e010507.
79. National Health Service. National schedule of reference costs 2019/20. Published 15 September 2021; Available from: www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/. Accessed 9 March 2022.
80. Stevenson M, Lloyd-Jones M, Morgan MY, Wong R. Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation. Health Technol Assess. 2012; 16:1-174.
81. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat Methods Med Res. 2017; 26:1896-911.
82. Imajo K, Nagai K, Iwaki M, Kobayashi T, Honda Y, Kessoku T*, et al.* Comparative performance of non-invasive imaging modalities for the diagnosis of NASH in a japanese NAFLD population. AASLD 2020: The Liver Meeting; 2020; Virtual.
83. Eddowes P, McDonald N, Davies N, Semple S, Hübscher S, Kendall T*, et al.* Validation of multiparametric MRI in the assessment and staging of non-alcoholic fatty liver disease. Aliment Pharmacol Ther; 2016; England.

10 APPENDICES

Appendix 1 PRISMA-DTA checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	2
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	114
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	19 to 30
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	21 to 29
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	19
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	18
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	33
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	31
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	115 to 117
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	34
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	34

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	34
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	34
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	35 to 36
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	35 to 36
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	35 to 36
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	122 to 123
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	37 and Figure 2
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	Table 4
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	39 to 41
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figure 3 to 6
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Table 6
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	62

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	97 to 98
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	100
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	101 to 103
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	2

FN=false negative; FP=false positive; PRISMA-DTA=Preferred Reporting Items for Systematic Reviews and Meta-Analyses for diagnostic test accuracy studies; TN=true negative; TP=true positive

Appendix 2 PRISMA-DTA for Abstracts checklist

Section/topic	#	PRISMA-DTA for Abstracts Checklist item	Reported on page #
TITLE and PURPOSE			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	4
Objectives	2	Indicate the research question, including components such as participants, index test, and target conditions.	4
METHODS			
Eligibility criteria	3	Include study characteristics used as criteria for eligibility.	4
Information sources	4	List the key databases searched and the search dates.	4
Risk of bias & applicability	5	Indicate the methods of assessing risk of bias and applicability.	4
Synthesis of results	A1	Indicate the methods for the data synthesis.	4
RESULTS			
Included studies	6	Indicate the number and type of included studies and the participants and relevant characteristics of the studies (including the reference standard).	4
Synthesis of results	7	Include the results for the analysis of diagnostic accuracy, preferably indicating the number of studies and participants. Describe test accuracy including variability; if meta-analysis was done, include summary results and confidence intervals.	5
DISCUSSION			
Strengths and limitations	9	Provide a brief summary of the strengths and limitations of the evidence	5
Interpretation	10	Provide a general interpretation of the results and the important implications.	5
OTHER			
Funding	11	Indicate the primary source of funding for the review.	6
Registration	12	Provide the registration number and the registry name	5

Appendix 3 Search strategies

MEDLINE (R) ALL (via Ovid)

1 exp Non-alcoholic Fatty Liver Disease/
2 non-alcoholic fatty liver disease.tw,kw.
3 NAFLD.tw,kw.
4 non-alcoholic steatohepatitis.tw,kw.
5 NASH.tw,kw.
6 metabolic dysfunction associated fatty liver disease.tw,kw.
7 MAFLD.tw,kw.
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 exp Magnetic Resonance Imaging/
10 MRI.tw,kw.
11 magnetic resonance imag*.tw,kw.
12 LiverMultiScan.tw,kw.
13 Magnetic resonance elastograph*.tw,kw.
14 MRE.tw,kw.
15 9 or 10 or 11 or 12 or 13 or 14
16 8 and 15
17 exp animals/
18 human/
19 17 not 18
20 16 not 19
21 limit 20 to english language

Embase (via Ovid)

1 exp nonalcoholic fatty liver/
2 non-alcoholic fatty liver disease.tw,kw.
3 NAFLD.tw,kw.
4 exp nonalcoholic steatohepatitis/
5 non-alcoholic steatohepatitis.tw,kw.
6 NASH.tw,kw.
7 exp metabolic fatty liver/
8 metabolic dysfunction associated fatty liver disease.tw,kw.
9 MAFLD.tw,kw.
10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
11 exp nuclear magnetic resonance imaging/

12 MRI.tw,kw.
13 magnetic resonance imag*.tw,kw.
14 LiverMultiScan.tw,kw.
15 exp magnetic resonance elastography/
16 Magnetic resonance elastograph*.tw,kw.
17 MRE.tw,kw. 3770
18 11 or 12 or 13 or 14 or 15 or 16 or 17
19 10 and 18
20 Animal experiment/
21 human experiment/ or human/
22 20 not 21
23 19 not 22
24 limit 23 to english language
25 limit 24 to embase
26 limit 24 to conference abstracts
27 25 or 26

Cochrane Central Database of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) (via The Cochrane Library)

1 MeSH descriptor: [Non-alcoholic Fatty Liver Disease] explode all trees
2 ("non-alcoholic fatty liver disease"):ti,ab,kw
3 (NAFLD):ti,ab,kw
4 ("non-alcoholic steatohepatitis"):ti,ab,kw
5 (NASH):ti,ab,kw
6 ("metabolic dysfunction associated fatty liver disease"):ti,ab,kw
7 (MAFLD):ti,ab,kw
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 MeSH descriptor: [Magnetic Resonance Imaging] explode all trees
10 (MRI):ti,ab,kw
11 (magnetic NEXT resonance NEXT imag*):ti,ab,kw
12 (LiverMultiScan):ti,ab,kw
13 (Magnetic resonance elastograph*):ti,ab,kw
14 (MRE):ti,ab,kw
15 9 or 10 or 11 or 12 or 13 or 14
16 #8 AND #15

Database of Abstracts of Reviews of Effects (DARE) (via Centre for Reviews and Dissemination)

1 MeSH DESCRIPTOR Non-alcoholic Fatty Liver Disease EXPLODE ALL TREES
2 ("non-alcoholic fatty liver disease")
3 (NAFLD)
4 ("non-alcoholic steatohepatitis")
5 (NASH)
6 ("metabolic dysfunction associated fatty liver disease")
7 (MAFLD)
8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
9 MeSH DESCRIPTOR Magnetic Resonance Imaging EXPLODE ALL TREES
10 (MRI)
11 ("magnetic resonance imag*")
12 (LiverMultiScan)
13 ("Magnetic resonance elastograph*")
14 (MRE)
15 #9 OR #10 OR #11 OR #12 OR #13 OR #14
16 #8 AND #15

Health Technology Assessment Database (HTA) (via International HTA Database)

(MAFLD) OR ("metabolic dysfunction associated fatty liver disease") OR (NASH) OR ("non-alcoholic steatohepatitis") OR (NAFLD) OR ("non-alcoholic fatty liver disease") OR ("Non-alcoholic Fatty Liver Disease"[mhe])

Appendix 4 Additional searches

MEDLINE (R) ALL (via Ovid)

Intermediate outcomes

1	exp Non-alcoholic Fatty Liver Disease/
2	non-alcoholic fatty liver disease.tw,kw.
3	NAFLD.tw,kw.
4	non-alcoholic steatohepatitis.tw,kw.
5	NASH.tw,kw.
6	metabolic dysfunction associated fatty liver disease.tw,kw.
7	MAFLD.tw,kw.
8	or/1-7
9	exp Magnetic Resonance Imaging/
10	MRI.tw,kw.
11	magnetic resonance imag*.tw,kw.
12	LiverMultiScan.tw,kw.
13	Magnetic resonance elastograph*.tw,kw.
14	MRE.tw,kw.
15	or/9-14
16	8 and 15
17	exp animals/
18	human/
19	17 not 18
20	16 not 19
21	limit 20 to english language
22	Clinical Decision-Making/
23	"clinical decision making".tw,kw.
24	22 or 23
25	20 and 24
26	8 and 24
27	exp "Predictive Value of Tests"/
28	((predict* or prognos*) adj (value or ability)).tw,kw.
29	(predict* adj2 (progression or regression)).tw,kw.
30	27 or 28 or 29
31	20 and 30
33	exp *"Predictive Value of Tests"/
34	((predict* or prognos*) adj (value or ability)).ti,kw.
35	(predict* adj2 (progression or regression)).ti,kw.

36 33 or 34 or 35
37 8 and 36
38 *Biopsy/ and Liver/
39 "number of liver biops*".tw,kw.
40 ("number of biops*" adj3 liver).tw,kw.
41 38 or 39 or 40
42 20 and 41
43 8 and 41
44 (lifestyle adj modif*).tw,kw.
45 20 and 44
46 (lifestyle adj modif*).ti,kw.
47 8 and 46
48 (time adj3 result*).tw,kw.
49 20 and 48
50 8 and 48
51 (time adj5 diagnos*).tw,kw.
52 Delayed Diagnosis/
53 Early Diagnosis/
54 51 or 52 or 53
55 20 and 54
56 "time to diagnosis".tw,kw.
57 8 and 56
58 (fail* adj3 (rate* or detect* or diagnos*)).tw,kw.
59 20 and 58
60 8 and 58
61 ((reduc* or remission) adj5 (fibrosis or inflammation)).tw,kw.
62 20 and 61
63 ((reduc* or remission) adj3 (liver fibrosis or fibro inflammat* or fibro-inflammat*)).tw,kw.
64 8 and 63
72 ((reduc* or remission) adj3 (liver adj fat*)).tw,kw.
73 20 and 72
74 8 and 72

Clinical outcomes and patient-reported outcomes

1 exp Non-alcoholic Fatty Liver Disease/
2 non-alcoholic fatty liver disease.tw,kw.
3 NAFLD.tw,kw.

4 non-alcoholic steatohepatitis.tw,kw.
5 NASH.tw,kw.
6 metabolic dysfunction associated fatty liver disease.tw,kw.
7 MAFLD.tw,kw.
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 exp Magnetic Resonance Imaging/
10 MRI.tw,kw.
11 magnetic resonance imag*.tw,kw.
12 LiverMultiScan.tw,kw.
13 Magnetic resonance elastograph*.tw,kw.
14 MRE.tw,kw.
15 9 or 10 or 11 or 12 or 13 or 14
16 8 and 15
17 exp animals/
18 human/
19 17 not 18
20 16 not 19
21 limit 20 to english language
22 exp Mortality/
23 (mortalit* or death* or died).tw,kw.
24 22 or 23
25 20 and 24
26 (mortalit* or death* or died).ti,kw.
27 22 or 26
28 8 and 27
29 28 not 25
30 exp Morbidity/
31 morbidit*.tw,kw.
32 contraindicat*.tw,kw.
33 complication*.tw,kw.
34 30 or 31 or 32 or 33
35 8 and 34
36 (morbidit* or complication* or contraindicat*).ti,kw.
37 exp *Morbidity/
38 36 or 37
39 8 and 38
40 20 and 34

41 39 not 40
42 exp "Quality of Life"/
43 "quality of life".tw,kw.
44 "Chronic Liver Disease Questionnaire ".tw,kw.
45 CLDQ.tw,kw.
46 42 or 43 or 44 or 45
47 8 and 46
48 20 and 46
49 47 not 48
50 exp "Patient Acceptance of Health Care"/ or exp Patient Satisfaction/
51 acceptab*.tw,kw.
52 (patient* adj3 satisf*).tw,kw.
53 "perceived effectiveness".tw,kw.
54 claustrophobi*.tw,kw.
55 50 or 51 or 52 or 53 or 54
56 8 and 55
57 20 and 55
58 56 not 57

Appendix 5 Methods of analysis/synthesis: Differences between protocol and review

DTA studies

The EAG did not plot the sensitivity and specificity of each index test in ROC space. There was only one combination of index test and diagnosis where studies reported diagnostic test accuracy for a variety of different cut-off values. For other combinations of index test and diagnosis, data were reported for two cut-off values at most, and plotting studies in ROC space would not have been informative. For the combination of index test and diagnosis where studies reported accuracy for a variety of different cut-off values, the results from individual studies were plotted in ROC space, along with the summary ROC curve from the hierarchical model.

The EAG did not encounter issues with sparse data when performing the meta-analyses, and so it was not necessary to reduce the bivariate model to two univariate random-effects logistic regression models by assuming no correlation between sensitivity and specificity across studies.[81]

Study characteristics, populations and results were not sufficiently homogenous to perform additional meta-analyses using fixed-effects models (i.e., simplifying the regression models to fixed-effects models by eliminating the random-effects parameters for sensitivity and specificity). All meta-analyses were conducted using random-effects models. The bivariate model was fitted using the meqrlogit command in Stata 14 (meqrlogit replaces xtmelogit in Stata 14).

If data had been available, the EAG would have examined the impact of the following variables on the diagnostic accuracy of MRI-based technologies by performing subgroup analyses or meta-regression (by inclusion of the variable as a covariate in a bivariate model):

- prior tests for fibrosis (i.e., an indicator variable for whether FIB-4, NFS, ELF, TE and/or ARFI tests have previously been performed)
- age (i.e., adults [≥18 years] compared to children and young people [<18 years] and/or mean / median age of patients in the study included as a continuous covariate in the bivariate model).

If data had been available, the EAG would have conducted sensitivity analyses by excluding studies judged to have a high risk of bias for at least one domain of the QUADAS-2 tool, or studies that the EAG was uncertain about the appropriateness of including them in the primary meta-analyses.

Data were insufficient to perform any subgroup analyses or sensitivity analyses.

Clinical impact studies

No studies provided data for the clinical impact outcomes of interest, and limited data were available for intermediate outcomes. There were only sufficient data to perform a meta-analysis for MRE test failure rate. It was not necessary or useful to plot or tabulate the data reported for other outcomes; these data were therefore reported narratively.

If the EAG had tabulated or plotted other clinical and/or intermediate outcome data, binary and categorical data would have been presented as frequencies and proportions, and continuous data would have been presented as means and standard deviations, or medians and interquartile ranges, according to the distribution of the data. If it had been possible to perform meta-analyses for continuous outcomes, the EAG would have expressed continuous data as means and standard deviations or standard errors (calculated from standard deviations or confidence intervals where appropriate), and pooled these data an inverse-variance meta-analysis using the metan command in Stata version 14.

Very little heterogeneity was observed in the conducted meta-analyses, and therefore it was not necessary to perform subgroup analyses. The EAG also did not perform sensitivity analyses, as there were no studies that the EAG considered to be important to exclude in sensitivity analyses (to investigate the impact of the inclusion of these studies on the overall pooled estimate).

Appendix 6 Excluded studies

1. Alquraish M, Cepin S, Nguyen P, Hernandez C, Bettencourt R, Fortney L, et al. Obesity predicts discordancy between magnetic resonance elastography and transient elastography for the stage of fibrosis in patients with nonalcoholic fatty liver disease. Hepatology 2017; 66:335A. **Wrong publication type**

2. Anna O, Michihiro I, Takashi K, Asako N, Yasushi H, Takaomi K, et al. Influence of liver stiffness heterogeneity on concordance of MR elastography-based liver fibrosis staging and biopsy results in patients with nonalcoholic fatty liver disease. Journal of Gastroenterology and Hepatology 2021; 36:104-5. **Wrong publication type**

3. Aslam F, Mouchti S, Kelly M, Dennis A, Imajo K, Nakajima A. Investigation of a composite imaging biomarker for identification of non-alcoholic steatohepatitis (NASH) patients in a Japanese population. Journal of Hepatology 2020; 73:S411-S2. **Wrong publication type**

4. Beyer C, Hutton C, Andersson A, Imajo K, Nakajima A, Kiker D, et al. Comparison between magnetic resonance and ultrasound-derived indicators of hepatic steatosis in a pooled NAFLD cohort. PloS one 2021; 16:e0249491. **No outcomes of interest**

5. Bravo S, Kelly M, Xu P, Banerjee R, Neubauer S, Hollar K, et al. Evaluation of multiparametric MRI in comparison with MR elastography in patients evaluated for chronic liver disease. Journal of Hepatology 2018; 68:S638-S9. **Wrong publication type**

6. Caussy C, Ajmera VH, Puri P, Hsu CL-S, Bassirian S, Mgdsyan M, et al. Serum metabolites detect the presence of advanced fibrosis in derivation and validation cohorts of patients with non-alcoholic fatty liver disease. Gut 2019; 68:1884-92. **Wrong population**

7. Choi SJ, Kim SM, Kim YS, Kwon OS, Shin SK, Kim KK, et al. Magnetic Resonance-Based Assessments Better Capture Pathophysiologic Profiles and Progression in Nonalcoholic Fatty Liver Disease. Diabetes & metabolism journal 2020. **Wrong population**

8. Dennis A, Kelly M, Fernandes C, Mouchti S, Banerjee R, Fallowfield J, et al. Utility and interpretation of the quantitative MRI metrics PDFF and cT1 as biomarkers for non-alcoholic steatohepatitis. American Journal of Gastroenterology 2020; 115:S589-S90. **Wrong publication type**

9. Dennis A, Kelly MD, Fernandes C, Mouchti S, Fallowfield JA, Hirschfield G, et al. Correlations Between MRI Biomarkers PDFF and cT1 With Histopathological Features of

Non-Alcoholic Steatohepatitis. Frontiers in endocrinology 2020; 11:575843. **No outcomes of interest**

10. Dzyubak B, Li J, Chen J, Mara KC, Therneau TM, Venkatesh SK, et al. Automated analysis of multiparametric magnetic resonance imaging/magnetic resonance elastography exams for prediction of nonalcoholic steatohepatitis. Journal of magnetic resonance imaging 2021; 54:122-31. **No outcomes of interest**

11. Eddowes PJ, Newsome PN, Hirschfield GM, McDonald N, Fallowfield J, Davies NP, et al. Exclusion of clinically significant non-alcoholic fatty liver disease with multi-parametric magnetic resonance imaging: A prospective evaluation. Hepatology 2016; 64:572A-3A. **Wrong publication type**

12. Eddowes P, Newsome P, Hirschfield G, McDonald N, Fallowfield J, Davies N, et al. Validation of multiparametric MRI in the assessment and staging of non-alcoholic fatty liver disease. Gut 2016; 65:A157-A8. **Wrong publication type**

13. Filza A, Sofia M, Andrea D, Matt K, Rajarshi B, Kento I, et al. Non-invasive imaging modalities for assessment of fibrosis, inflammation and steatosis in a Japanese NASH population. Hepatology International 2020; 14:S321. **Wrong publication type**

14. Freitag CE, Andersson I, Chen W, Hinton A, Levin D, Yearsley MM, et al. Comparison of histologic and magnetic resonance methodologies for the estimation of hepatic steatosis. Laboratory Investigation 2018; 98:805. **Wrong publication type**

15. Imajo K, Iwaki M, Kobayashi T, Honda Y, Kessoku T, Ogawa Y, et al. Impact of liver stiffness heterogeneity on discordance between pathological liver fibrosis stage and mr elastography-based liver stiffness measurements in patients with NAFLD. Hepatology 2020; 72:923A. **Wrong publication type**

16. Imajo K, Nagai K, Iwaki M, Kobayashi T, Honda Y, Kessoku T, et al. Comparative performance of non-invasive imaging modalities for the diagnosis of nash in a japanese NAFLD population. Hepatology 2020; 72:905A-6A. **Wrong publication type**

17. Kawada T. Validation Study of Elastographies in Patients With Nonalcoholic Fatty Liver Disease for Detecting Liver Fibrosis. Clinical Gastroenterology and Hepatology 2019; 17:2139-40. **Wrong publication type**

18. Lee Y-S, Lee M-J, Kim JH, Seo YS, Yim HJ, Yeon JE, et al. Multiparametric MRI effectively evaluated disease severity of nonalcoholic fatty liver disease. Hepatology 2019; 70:1085A. **Wrong publication type**

19. Lee Y-S, Yoo YJ, Jung YK, Kim JH, Seo YS, Yim HJ, et al. Multiparametric MR Is a valuable modality for evaluating disease severity of nonalcoholic fatty liver disease. Clinical and translational gastroenterology 2020; 11:e00157. **Wrong population**

20. McDonald N, Fallowfield J, Eddowes PJ, Hirschfield GM, Semple SI, Davies NP, et al. Multiparametric assessment of liver disease using quantitative magnetic resonance imaging: A two-centre prospective validation study. Hepatology 2016; 64:323A-4A. **Wrong publication type**

21. Miles L, King E, Kohli R, Xanthakos S, Podberesky D, Serai S. Assessment of hepatic fibrosis in pediatric chronic liver disease with MR elastography. Pediatric Radiology 2014; 44:S71. **Wrong publication type**

22. Murphy-Lavallee J, Olivie D, Ilinca A, Lefebvre T, Wartelle-Bladou C, Giard J-M, et al. Prospective comparison of transient, point shear wave, and magnetic resonance elastography for staging liver fibrosis. European Radiology 2019; 29:6477-88. **Wrong population**

23. Nogami A, Iwaki M, Kobayashi T, Kessoku T, Honda Y, Saito S, et al. Assessment of hepatic fibrosis by vibration-controlled transient elastography and MR elastography have equivalent diagnostic performance, but in the assessment of hepatic steatosis, MRI PDFF methods are better than controlled attenuation parameter in over. Journal of Gastroenterology and Hepatology 2021; 36:246. **Wrong population**

24. Sharpton SR, Bettencourt R, Jung J, Heilman J, Pepin K, Ehman RL, et al. Automated analysis of magnetic resonance elastography and its reproducibility with manual analysis in adults with nonalcoholic fatty liver disease: A goldmine study. Hepatology 2020; 72:893A-4A. **Wrong publication type**

25. Sohn W, Kwon H-J, Chang Y, Ryu S, Cho YK. Liver fibrosis in asians with metabolic dysfunction-associated fatty liver disease. Clinical gastroenterology and hepatology 2021. **Wrong study design**

26. Tamaki N, Imajo K, Sharpton S, Jung J, Kawamura N, Yoneda M, et al. MRE plus FIB-4 (MEFIB) versus FAST in detection of candidates for pharmacological treatment of NASH-related fibrosis. Hepatology 2021. **Wrong population**

27. Tonev D, Shumbayawonda E, Tetlow LA, Herdman L, French M, Rymell S, et al. The effect of multi-parametric magnetic resonance imaging in standard of care for nonalcoholic fatty liver disease: Protocol for a randomized control trial. JMIR research protocols 2020; 9:e19189. **Wrong publication type**

28. Yin M, Glaser KJ, Talwalkar JA, Chen J, Manduca A, Ehman RL. Hepatic MR elastography: Clinical performance in a series of 1377 consecutive examinations1. Radiology 2016; 278:114-24. **No outcomes of interest**

Appendix 7 Studies suggested by manufacturers and reasons for exclusion

Excluded studies from the reference list suggested by Perspectum Ltd.

1. Alkhouri et al. (unpublished) 'MRI assessment (cT1) with LiverMultiScan following VCTE improves the diagnostic yield for high-risk NASH' (Submitted to EASL 2022). **Wrong publication type**

2. Alkhouri et al. (unpublished) 'Sequential testing for high-risk NASH by cT1 from LiverMultiScan improves diagnostic yield compared to the use of MRE alone' (Submitted to DDW 2022). **Wrong publication type**

3. Andersson A, Kelly M, Imajo K, Nakajima A, Fallowfield JA, Hirschfield G et al. Clinical utility of MRI biomarkers for identifying NASH patients at high risk of progression: A multi-center pooled data and meta-analysis. Clin Gastroenterol Hepatol 2021; In press. **Wrong study design**

4. Beyer C, Hutton C, Andersson A, Imajo K, Nakajima A, Kiker D et al. Comparison between magnetic resonance and ultrasound-derived indicators of hepatic steatosis in a pooled NAFLD cohort. Plos One 2020; 16(4):e0249491. **Wrong study design**

5. Brown E, Waddell T, Mouchti S, Roca-Fernandez A, Thomaides-Brears H, Wilton M et al. Multiparametric magnetic resonance imaging of the liver demonstrates the prevalence of steatohepatitis in patients with type 2 diabetes. Diabetologia 2020; 876. **Wrong publication type**

6. Carolan JE, Dennis A, Hutton C, Kelly M (unpublished) Investigating the cost-effectiveness of quantitative MRI for identifying adults with suspected NAFLD in Europe. (Accepted to ICFL 2022). **Wrong publication type**

7. Cruz M, Ferreira AA, Papanikolaou N, Banerjee R, Alves FC. New boundaries of liver imaging: from morphology to function. Eur J Intern Med; 2020 79:12-22. **Wrong study design**

8. Dennis A, Mouchti S, Kelly M, Fallowfield, JA, Hirschfield G, Pavlides M et al. A composite biomarker using multiparametric magnetic resonance imaging and blood analytes accurately identifies patients with non-alcoholic steatohepatitis and significant fibrosis. Sci Rep 2020; 10(1):15308. **Wrong patient population**

9. Dennis A, Kelly M, Fernandes C, Mouchti S, Banerjee R, Fallowfield J et al. Utility and interpretation of the quantitative MRI metrics PDFF and cT1 as biomarkers for Non-alcoholic Steatohepatitis. The American College of Gastroenterology 2020; 115(Suppl):S589-90. **Wrong publication type**

10. Harrison S, Roberts K, Paredes A, Lisanti C, Schwope R, Whitehead J et al. Prospective liver biopsy-based prevalence of NAFLD and steatohepatitis among a large middle-aged population utilizing FibroScan, LiverMultiScan and MRE to guide liver biopsy. Journal of Hepatology. 2019; 70(1):e770-1 **Wrong publication type**

11. Harrison SA, Dennis A, Fiore MM, Kelly MD, Kelly CJ, Paredes AH et al. Utility and variability of three non-invasive liver fibrosis imaging modalities to evaluate efficacy of GR-MD-02 in subjects with NASH and bridging fibrosis during a phase-2 randomized clinical trial. PLoS ONE 2018; 13(9):e0203054. **Wrong population**

12. Harrison SA, Gawrieh S, Roberts K, Lisanti CJ, Schwope RB, Cebe KM et al. Prospective evaluation of the prevalence of non-alcoholic fatty liver disease and steatohepatitis in a large middle-aged US cohort. Journal of Hepatology 2021; 75(2):284-91. **Wrong study design**

13. Hydes TJ, Summers N, Brown E, Wilding JPH, Cuthbertson DJ, Alam U et al. Mechanisms, screening modalities and treatment options with NAFLD and type 2 diabetes. Diabet Med 2020; 37(11):1793-1806. **Wrong study design**

14. Noureddin M, Beyer C, Loomba R, Harisinghani M, Harrison S, Alkhouri N et al. (unpublished) Decreases in liver cT1 accurately reflect histological improvement induced by therapies in NASH with enhanced sensitivity to fibrosis change: a multi-centre pooled cohort analysis. (Submitted to EASL 2022). **Wrong publication type**

15. Samur SS, Carolan JE, Chhatwal J et al. Comparative cost-effectiveness of multiparametric magnetic resonance imaging for detection of high-risk NASH. Hepatology 2020; 904A-5A. **Wrong publication type**

16. Thomaides-Brears HB, Lepe R, Banerjee R, Duncker C et al. Multiparametric MR mapping in clinical decision-making for diffuse liver disease. Abdom Radiol 2020; 45(11):3507-3522. **Wrong study design**

Excluded studies from the reference list suggested by Resoundant, Inc.

1. Ajmera V, Loomba R. Imaging biomarkers of NAFLD, NASH, and fibrosis. Mol Metab 2021; 50:101167. **Wrong study design**

2. Allen AM, Shah VH, Therneau TM, Venkatesh SK, Mounajjed T, Larson JJ et al. Multiparametric magnetic resonance elastography improves the detection of NASH regression following bariatric surgery. Hepatol Commun 2020; 4(2):185-92. **Wrong population**

3. Chen J, Allen A, Therneau T, Chen J, Li J, Hoodeshenas S et al. Liver stiffness measurement by magnetic resonance elastography is not affected by hepatic steatosis. Eur Radiol 2021; 32:950-8. **Wrong population**

4. Costa-Silva L, Ferolla SM, Lima AS, Vidigal P, Ferrari T. MR elastography is effective for the non-invasive evaluation of fibrosis and necroinflammatory activity in patients with nonalcoholic fatty liver disease. Eur J Radiol 2018; 98: 82-9. **Wrong population**

5. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodríguez-Perálvarez 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). **Wrong intervention**

6. Cui J, Ang B, Haufe W, Hernandez C, Verna EC, Sirlin CB et al. Comparative diagnostic accuracy of magnetic resonance elastography vs. eight clinical prediction rules for non-invasive diagnosis of advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease: A prospective study. Aliment Pharmacol Ther 2015; 41:1271-80. **Wrong population**

7. Dzyubak B, Li J, Chen J, Mara K, Therneau T, Venkatesh S et al. Automated analysis of multiparametric magnetic resonance imaging/magnetic resonance elastography exams for prediction of nonalcoholic steatohepatitis. JMRI 2021; 54(1):122-131. **No outcomes of interest**

8. Gidener T, Yin M, Dierkhising R, Allen A, Ehman R, Venkatesh S. MRE for prediction of long-term progression and outcome in chronic liver disease: A retrospective study. Hepatology 2021; 75: 10. **Wrong publication type**

9. Han M, Vipani A, Noureddin N, Ramirez K, Gornbein J, Saouaf R et al. MR elastography-based liver fibrosis correlates with liver events in nonalcoholic fatty liver patients: A multicenter study. Liver Int 2020; 40(9):2242–2251. **No outcomes of interest**

10. Higuchi M, Tamaki N, Kurosaki M, Inada K, Kirino S, Yamashita K et al. Changes of liver stiffness measured by magnetic resonance elastography during direct-acting antivirals treatment in patients with chronic hepatitis C. J Med Virol 2020; 93:3744-51. **Wrong population**

11. Hsu C, Caussy C, Imajo K, Chen J, Singh S, Kaulback K et al. Magnetic resonance vs. transient elastography analysis of patients with nonalcoholic fatty liver disease: A systematic review and pooled analysis of individual participants. Clin Gastroenterol Hepatol 2019; 17:630–637. **Wrong study design**

12. Jung J, Loomba RR, Imajo K, Madamba E, Gandhi S, Bettencourt R et al. MRE combined with FIB-4 (MEFIB) index in detection of candidates for pharmacological treatment of NASH-related fibrosis. Gut 2021; 70(10):1946–53. **Wrong population**

13. Lee YS, Yoo YJ, Jung YK, Kim JH, Seo YS, Yim HJ et al. Multiparametric MR is a valuable modality for evaluating disease severity of nonalcoholic fatty liver disease. CTG 2020; 11(4):e00157. **Wrong population**

14. Li J, Allen A, Shah V, Manduca A, Ehman R, Yin M. (2021) Longitudinal Changes in MR Elastography-based Biomarkers in Obese Patients Treated with Bariatric Surgery. Clin Gastroenterol Hepatol 2021; In press. **Wrong population**

15. Liang Y, Li D. Magnetic resonance elastography in staging liver fibrosis in non-alcoholic fatty liver disease: A pooled analysis of the diagnostic accuracy. BMC Gastroenterol 2020; 20(1):89. **Wrong study design**

16. Noureddin M, Truong E, Gornbein JA, Saouaf R, Guindi M, Todo T et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. J Hepatol 2021; In press. **Wrong intervention**

17. Patel NS, Hooker J, Gonzalez M, Bhatt A, Nguyen P, Ramirez K et al. Weight loss decreases magnetic resonance elastography estimated liver stiffness in nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2017; 15:463-4. **No outcomes of interest**

18. Selvaraj EA, Mozes FE, Jayaswal ANA, Zafarmand MH, Vali Y, Lee JA et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: A systematic review and meta-analysis. J Hepatol 2021; 75 (4):770–785. **Wrong study design**

19. Singh S, Fujii LL, Murad MH, Wang Z, Asrani SK, Ehman RL et al. Liver stiffness is associated with risk of decompensation, liver cancer, and death in patients with chronic liver diseases: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2013; 11(12):1573–e89. **Wrong study design**

20. Vilar-Gomez E, Lou Z, Kong N, Vuppalanchi R, Imperiale TF, Chalasani N. Cost effectiveness of different strategies for detecting cirrhosis in patients with nonalcoholic fatty liver disease based on United States health care system. Clin Gastroenterol Hepatol 2020; 18(10):2305–2314.e12. **Wrong intervention**

21. Zhang E, Wartelle-Bladou C, Lepanto L, Lachaine J, Cloutier G, Tang A. Cost-utility analysis of nonalcoholic steatohepatitis screening. Eur Radiol 2015; 25(11):3282-94. **Wrong population**

Appendix 8 QUADAS-2 quality assessment of DTA studies

Caussy 2018[53]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Patients with suspected NAFLD indicated for a liver biopsy were recruited consecutively into a cross-sectional study.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.		
Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST

A. Risk of Bias

The tests were performed by a radiologist blinded to the patient's clinical data. The thresholds of MRE were pre-defined.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		Yes
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Liver biopsy performed by an experienced liver pathologist who was blinded to the patient's clinical and radiologic data.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Liver biopsy conducted 48 hours to one month after MRE.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Eddowes 2018[29]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Consecutive patients across two sites recruited to a cross-sectional study.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		No
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and scheduled for non-targeted liver biopsy to stage fibrosis after inconclusive non-invasive assessment of fibrosis or to make a diagnosis after a range of non-invasive tests had not confirmed a diagnosis.		
Is there concern that the included patients do not match the review question?	**Concerns**	**LOW**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Index tests interpreted by a single operator blinded to the clinical findings and biopsy results. Unclear if the thresholds used were pre-specified.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		Unclear
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Liver biopsy assessed by experienced academic liver histopathologists blinded to the MRI findings.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Two weeks interval between index test and reference standard.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Forsgren 2020[54]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Study recruited all patients who required a liver biopsy between 2007 and 2014.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.		
Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

No information provided on whether the results were interpreted without knowledge of the results of the reference standard. Thresholds were not pre-specified. Applicability concerns were judged to be high because the study used an investigational MRE design and not the Resoundant MRE platform that is commercially-available.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Unclear
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**HIGH**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**HIGH**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

No information provided on whether the results of the liver biopsy were interpreted without knowledge of the results of the index test.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Index test and reference standard performed on the same day.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Hoffman 2020[55]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

The study recruited all patients with known or suspected hepatic fibrosis who underwent MRE between June and September 2018.

Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Unclear
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.

Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Index test interpreted by two readers blinded to the histopathology or other clinical or laboratory findings.

Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

No information provided on whether the results of the liver biopsy were interpreted without knowledge of the results of the index test.

Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Three months between index test and reference standard. Not all patients received a reference standard.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		No
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**UNCLEAR**

Imajo 2021[56]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Patients who were being screened clinically on suspicion of NASH between January 2019 and February 2020.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.		
Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST (LiverMultiScan)

A. Risk of Bias

Interpreted by image analysts who were blinded to the clinical data and risk grouping.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		Yes
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 2: INDEX TEST (MRE)

A. Risk of Bias

No information provided on whether the MRE results were interpreted without knowledge of the results of the reference standard.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Unclear
If a threshold was used, was it pre-specified?		Yes
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

No information provided on whether the results of the liver biopsy were interpreted without knowledge of the results of the index test.

Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

All tests conducted at clinical visit 1.

Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Kim 2013[57]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Consecutive patients with NAFLD underwent MRE and/or liver biopsy.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.		
Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST

A. Risk of Bias

MRE performed prior to the reference standard. Thresholds were not pre-specified.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Liver biopsy results were examined by dedicated hepatopathologists who were unaware of the MRE results.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Liver biopsy performed within one year of the index test.		
Was there an appropriate interval between index test and reference standard?		No
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**UNCLEAR**

Kim 2020[58]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Patients with clinically suspected NASH who were scheduled to undergo or underwent liver biopsy within 2 months were identified from October 2016 to June 2017.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.		
Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST

A. Risk of Bias

MRE performed prior to the reference standard. Thresholds were not pre-specified.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Reference standard interpreted by an experienced pathologist who was blinded to the patients' clinical and radiologic data.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Two months interval between index test and reference standard.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Pavlides 2017[59]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Patients with suspected or known NAFLD were invited to participate between May 2011 and March 2015.

Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.

Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Analysis of index tests were performed by a blinded investigator. Unclear if the thresholds used were pre-specified.

Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		Unclear
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Liver biopsies were evaluated by liver pathologists blinded to the MR data.

Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

One month interval between index test and reference standard.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Sofue 2020[60]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Consecutive patients recruited during a six-months period.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Unclear
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.		
Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Interpreted by a radiologist blinded to the patient clinical demographics and histopathologic findings. Thresholds were not pre-specified.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

No information provided on whether the results of the liver biopsy were interpreted without knowledge of the results of the index test.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Two months interval between index test and reference standard		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Toguchi 2017[61]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Consecutive patients with chronic liver disease recruited between October 2013 and January 2015.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Unclear
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.		
Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Results were interpreted by a radiologist who was blinded to the patient's clinical history. MRE was performed prior to the reference standard. Applicability concerns were judged to be high because the techniques for drawing regions of interest to calculate liver stiffness may not be representative of MRE in clinical practice.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**HIGH**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Results were interpreted by liver pathologists who were blinded to the patients' characteristics and results of the index test.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

The interval between index test and reference standard was less than 90 days.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Troelstra 2021[62]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Included the first 37 patients recruited to a separate study. Unclear how those patients were recruited.

Was a consecutive or random sample of patients enrolled?		Unclear
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study recruited patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed but it was unclear if these were patients who had indeterminate results from fibrosis testing, for whom TE or ARFI was unsuitable or who had discordant results from fibrosis testing.

Is there concern that the included patients do not match the review question?	**Concerns**	**UNCLEAR**

DOMAIN 2: INDEX TEST

A. Risk of Bias

The results were interpreted by a single observer blinded to the histopathology results. Applicability concerns were judged to be high because the study used an investigational MRE design and not the Resoundant MRE platform that is commercially-available.

Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**HIGH**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

The results were interpreted by a liver pathologist who was blinded to all other data.

Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Liver biopsy was performed within one week of the index test, with the exception of one participant whose biopsy was performed two months after the index test.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Trout 2018[63]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Patients who had undergone MRE between January 2012 and September 2016.		
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Yes
Could the selection of patients have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.		
Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Index test results were interpreted by a single observer who was blinded to the histologic data. Thresholds were not pre-specified.		
Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Liver biopsy results were interpreted by a single pathologist who was blinded to the index test results.		
Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

MRE and liver biopsy performed within three months of one another.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Xanthakos 2014[64]

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

The study included 35 children and adolescents who were evaluated with MRE and liver biopsy as part of their clinical evaluation for chronic liver disease from August 2011 to December 2012.

Was a consecutive or random sample of patients enrolled?		Unclear
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		Unclear
Could the selection of patients have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

The study included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed and patients with other liver disease aetiologies. Results were not presented separately for the population of interest to this assessment.

Is there concern that the included patients do not match the review question?	**Concerns**	**HIGH**

DOMAIN 2: INDEX TEST

A. Risk of Bias

Interpreter of index test was blinded to the results of the reference standard. Thresholds were not pre-specified.

Were the index test results interpreted without knowledge of the results of the reference standard?		Yes
If a threshold was used, was it pre-specified?		No
Could the conduct or interpretation of the index test have introduced bias?	**Risk**	**UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Concerns**	**LOW**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Interpreter of the liver biopsy results was blinded to the results of the index test.

Is the reference standard likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index test?		Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	**Risk**	**LOW**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?	**Concerns**	**LOW**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Median of 1.5 months interval between index test and reference standard.		
Was there an appropriate interval between index test and reference standard?		Yes
Did all patients receive a reference standard?		Yes
Did patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
Could the patient flow have introduced bias?	**Risk**	**LOW**

Appendix 9 AUROC results reported in the included studies

Table 21 AUROC results reported for LiverMultiScan

Diagnosis	Definition	Study	No. of patients	AUROC (95% CI)
LiverMultiScan PDFF				
Fibrosis	≥F1	Imajo 2021[56]	143	0.68 (0.44 to 0.92)
	≥F2	Imajo 2021[56]	143	0.60 (0.48 to 0.72)
Steatosis	Brunt grade ≥1	Eddowes 2018[29]	38	1.00 (1.00 to 1.00)
		Imajo 2021[56]	143	0.92 (0.87 to 0.98)
	Brunt grade ≥2	Imajo 2021[56]	143	0.86 (0.80 to 0.93)
	Brunt grade ≥3	Imajo 2020[82]	143	0.83 (NR)
NASH	NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation	Imajo 2021[56]	143	0.80 (0.73 to 0.87)
Advanced NASH	NAS ≥4 with fibrosis ≥F2	Imajo 2021[56]	143	0.71 (0.63 to 0.80)
LiverMultiScan cT1				
Fibrosis	≥F1	Imajo 2021[56]	143	0.63 (0.30 to 0.97)
	≥F2	Imajo 2021[56]	143	0.62 (0.49 to 0.74)
		Eddowes 2018[29]	50	0.63 (0.45 to 0.81)
	≥F3	Eddowes 2018[29]	50	0.62 (0.46 to 0.78)
Steatosis	Simple steatosis with no significant fibrosis*	Eddowes 2018[83]	50	0.75 (0.56 to 0.93)
	Brunt grade ≥1	Imajo 2021[56]	143	0.64 (0.46 to 0.82)
	Brunt grade ≥2	Imajo 2021[56]	143	NR
NASH	NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation	Imajo 2021[56]	143	0.75 (0.67 to 0.84)
	≥1 hepatocyte ballooning and ≥1 lobular inflammation	Eddowes 2018[29]	50	0.69 (0.50 to 0.88)
Advanced NASH	NAS ≥4 with fibrosis ≥2	Imajo 2021[56]	143	0.74 (0.66 to 0.82)
Disease activity	NAS ≥5	Eddowes 2018[29]	50	0.74 (0.59 to 0.88)
Risk of progressive disease	High risk (NASH or >F1) versus low risk (simple steatosis and ≤F1)	Eddowes 2018[29]	50	0.73 (0.53 to 0.93)
LiverMultiScan PDFF and cT1 combined				
NASH	NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation	Imajo 2021[56]	143	0.83 (0.76 to 0.90)
Advanced NASH	NAS ≥4 with fibrosis ≥F2	Imajo 2021[56]	143	0.76 (0.69 to 0.84)

* No further definition given. ≤F1 was assumed as no significant fibrosis because significant fibrosis was defined as >F1
AUROC=area under the receiver operating characteristic curve; CI=confidence interval; cT1=iron corrected longitudinal relaxation time; F=fibrosis stage; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; NR=not reported; PDFF=proton density fat fraction

Table 22 AUROC results reported for MRE

Diagnosis	Definition	Study	No. of patients	AUROC (95% CI)
Fibrosis	≥F1	Kim 2020[58]	47	0.99 (95% CI NR)
		Imajo 2021[56]	144	0.97 (0.94 to 1.00)
	≥F2	Kim 2020[58]	47	0.88 (95% CI NR)
		Imajo 2021[56]	144	0.92 (0.87 to 0.97)
		Caussy 2018[53]: UCSD cohort	119	**Patients with BMI <35kg/m²:** 0.89 (0.82 to 0.96) **Patients with BMI ≥35kg/m²:** 0.93 (0.84 to 1.00)
		Caussy 2018[53]: Mayo clinic cohort	75	**Patients with BMI <40kg/m²:** 0.97 (0.93 to 1.00) **Patients with BMI ≥40kg/m²:** 0.84 (0.69 to 0.98)
	≥F3	Kim 2020[58]	47	0.98 (95% CI NR)
		Kim 2013[57]	142	0.95 (0.91 to 0.98)
		Troelstra 2021[62] **G" modulus**	35	0.74 (0.48 to 1.00)
		Troelstra 2021[62] **G' modulus**	35	0.92 (0.83 to 1.00)
Lobular inflammation	≥2	Kim 2020[58]	47	0.77 (95% CI NR)
Steatosis	Brunt grade ≥1	Imajo 2021[56]	144	0.53 (0.33 to 0.72)
NASH	≥1 steatosis, ≥1 hepatocyte ballooning and ≥1 lobular inflammation	Troelstra 2021[62] **G" modulus**	35	0.69 (No CI)
		Troelstra 2021[62] **G' modulus**	35	0.79 (No CI)
	NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation	Imajo 2021[56]	144	0.57 (0.47 to 0.67)
Advanced NASH	NAS ≥4 with fibrosis ≥F2	Imajo 2021[56]	144	0.66 (0.57 to 0.75)
Hepatocyte ballooning	≥1	Kim 2020[58]	47	0.90 (95% CI NR)
	≥2	Kim 2020[58]	47	0.81 (95% CI NR)

AUROC=area under the receiver operating characteristic curve; BMI=body mass index; CIs=confidence intervals; F=fibrosis stage; G'=shear modulus; G"=loss modulus; MRE=magnetic resonance imaging; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; NR=not reported; UCSD=University of California, San Diego

Appendix 10 NIH quality assessment of clinical impact studies

Table 23 NIH quality assessment of cross-sectional studies

Criteria	Caussy 2018[53]	Eddowes 2018[29]	Forsgren 2020[54]	Kim 2013[57]	Pavlides 2017[59]	Troelstra 2021[62]	Xanthakos 2014[64]
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Was the study population clearly specified and defined?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3. Was the participation rate of eligible persons at least 50%?	CD	CD	Yes	Yes	CD	CD	CD
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5. Was a sample size justification, power description, or variance and effect estimates provided?	No	No	Yes	No	No	No	No
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	No	No	No	No	No	No	No
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	No	No	No	No	No	No	No
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10. Was the exposure(s) assessed more than once over time?	NA	NA	NA	NA	NA	NA	NA

Criteria	Caussy 2018[53]	Eddowes 2018[29]	Forsgren 2020 [54]	Kim 2013[57]	Pavlides 2017[59]	Troelstra 2021[62]	Xanthakos 2014[64]
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	Yes	Yes	CD	Yes	Yes	Yes	Yes
13. Was loss to follow-up after baseline 20% or less?	NA	NA	NA	NA	NA	NA	NA
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	No	No	No	No	No	No	No

CD=cannot determine; NA=not applicable

Table 24 NIH quality assessment of cohort studies

Criteria	Jayaswal 2020[66]	Gidener 2022[67]
1. Was the research question or objective in this paper clearly stated?	Yes	Yes
2. Was the study population clearly specified and defined?	Yes	Yes
3. Was the participation rate of eligible persons at least 50%?	CD	Yes
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	Yes	Yes
5. Was a sample size justification, power description, or variance and effect estimates provided?	No	No
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	Yes	Yes
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	Yes	Yes
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?	Yes	Yes
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes
10. Was the exposure(s) assessed more than once over time?	No	No
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	Yes	No
13. Was loss to follow-up after baseline 20% or less?	Yes	NA
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	No	Yes

CD=cannot determine; NA=not applicable

Appendix 11 Risk of bias assessment of randomised controlled trials

Study details

Reference	Tonev 2020,[74] Perspectum 2021[68]

Study design

X Individually-randomised parallel-group trial
☐ Cluster-randomised parallel-group trial
☐ Individually randomised cross-over (or other matched) trial

For the purposes of this assessment, the interventions being compared are defined as

Experimental: LiverMultiScan Comparator: Standard of care

Specify which outcome is being assessed for risk of bias	Number of unnecessary liver biopsies
Specify the numerical result being assessed. In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.	Liver biopsies Proportion of patients in each arm for which liver biopsies could have been avoided

Is the review team's aim for this result…?

X to assess the effect of *assignment to intervention* (the 'intention-to-treat' effect)

☐ to assess the effect of *adhering to intervention* (the 'per-protocol' effect)

If the aim is to assess the effect of *adhering to intervention*, select the deviations from intended intervention that should be addressed (at least one must be checked):

☐ occurrence of non-protocol interventions
☐ failures in implementing the intervention that could have affected the outcome
☐ non-adherence to their assigned intervention by trial participants

Which of the following sources were <u>obtained</u> to help inform the risk-of-bias assessment? (tick as many as apply)

☐ Journal article(s) with results of the trial
X Trial protocol
☐ Statistical analysis plan (SAP)
☐ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
☐ Company-owned trial registry record (e.g. GSK Clinical Study Register record)
☐ "Grey literature" (e.g. unpublished thesis)
☐ Conference abstract(s) about the trial
X Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
☐ Research ethics application
☐ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)
☐ Personal communication with trialist
☐ Personal communication with the sponsor

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Domain 1: Risk of bias arising from the randomisation process

Signalling questions	Comments	Response options
1.1 Was the allocation sequence random?	Patients were be randomised using a 1:1 allocation, without blinding, to the LiverMultiScan (intervention) arm and the standard of care (control) arm. Randomisation was automatically calculated using a random number generator for patients who had been already stratified based on inclusion criteria and the recruitment site	Y
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?		N
1.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?	Patient characteristics were not reported for the two treatment arms, only for the whole study population	NI
Risk-of-bias judgement		High

N=no; NI=no information; Y=yes

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

Signalling questions	Comments	Response options
2.1. Were participants aware of their assigned intervention during the trial?	Open-label trial	Y
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?		PY
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?		N
2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?		NA
2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?		NA
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	The protocol reported that intention-to-treat analysis would be used but did not report any additional statistical analyses to estimate effect of assignment to the intervention. No details for the statistical analysis were provided in the CSR[68]	NI

2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?		PN
Risk-of-bias judgement		Some concerns

CSR=clinical study report; N=no; NA=not applicable; NI=no information; PN=probably no; PY=probably yes; Y=yes

Domain 3: Missing outcome data

Signalling questions	Comments	Response options
3.1 Were data for this outcome available for all, or nearly all, participants randomised?	Only [redacted] patients underwent liver biopsy	N
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?		N
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	Patients who were not suspected to have NASH/significant fibrosis would not be scheduled for liver biopsy. This means that authors could not confirm the true negative rate and false negative rate	Y
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?		Y
Risk-of-bias judgement		High

N=no; NA=not applicable; NASH=non-alcoholic steatohepatitis; NI=no information; PN=probably no; PY=probably yes; Y=yes

Domain 4: Risk of bias in measurement of the outcome

Signalling questions	Comments	Response options
4.1 Was the method of measuring the outcome inappropriate?	Histological score using the NAS CRN scoring system was appropriate to determine whether patient should have undergone liver biopsy	N
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	The liver biopsy procedure is standardised and should not differ between sites or patients	PN
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Open-label trial	PY
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	Although it is possible that knowledge of the MRI data or SoC data could have influenced the NAS CRN score from liver biopsy, liver biopsy is a standard procedure which is done with prior knowledge in clinical practice	PY
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?		N

Risk-of-bias judgement		Some concerns

N=no; NA=not applicable; NAS CRN=NAFLD activity score; Clinical Research Network; NI=no information; PN=probably no; PY=probably yes; Y=yes

Domain 5: Risk of bias in selection of the reported result

Signalling questions	Comments	Response options
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	It is unclear whether a pre-specified analysis plan was finalised before data were available for analysis	Y
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...		
5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	All eligible reported results for the outcome domain correspond to all intended outcome measurements	N
5.3 ... multiple eligible analyses of the data?		N
Risk-of-bias judgement		Low

N=no; Y=yes

Overall risk of bias

Risk-of-bias judgement		High

Summary of the risk of bias assessment of randomised controlled trials

Author	Outcome	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall Bias
Tonev 2020,[74] Perspectum 2021[68]	Number of unnecessary liver biopsies	High	Some concerns	High	Some concerns	Low	High

Appendix 12 CASP checklist assessment of the qualitative study

Table 25 CASP qualitative studies checklist

Item	McKay 2021[69]
Section A. Are the results valid?	
1. Was there a clear statement of the aims of the research	Yes
2. Is a qualitative methodology appropriate?	Yes
3. Was the research design appropriate to address the aims of the research?	Yes
4. Was the recruitment strategy appropriate to the aims of the research?	Yes
5. Was the data collected in a way that addressed the research issue?	Yes
6. Has the relationship between the researcher and the participant been adequately considered?	Yes
Section B. What are the results?	
7. Have ethical issues been taken into consideration?	Yes
8. Was the data analysis sufficiently rigorous?	Yes
9. Is there a clear statement of findings?	Yes
Section C. Will the results help locally?	
10. How valuable is the research?	
The authors discuss the implications of the study findings for clinical practice	

Appendix 13 [redacted]

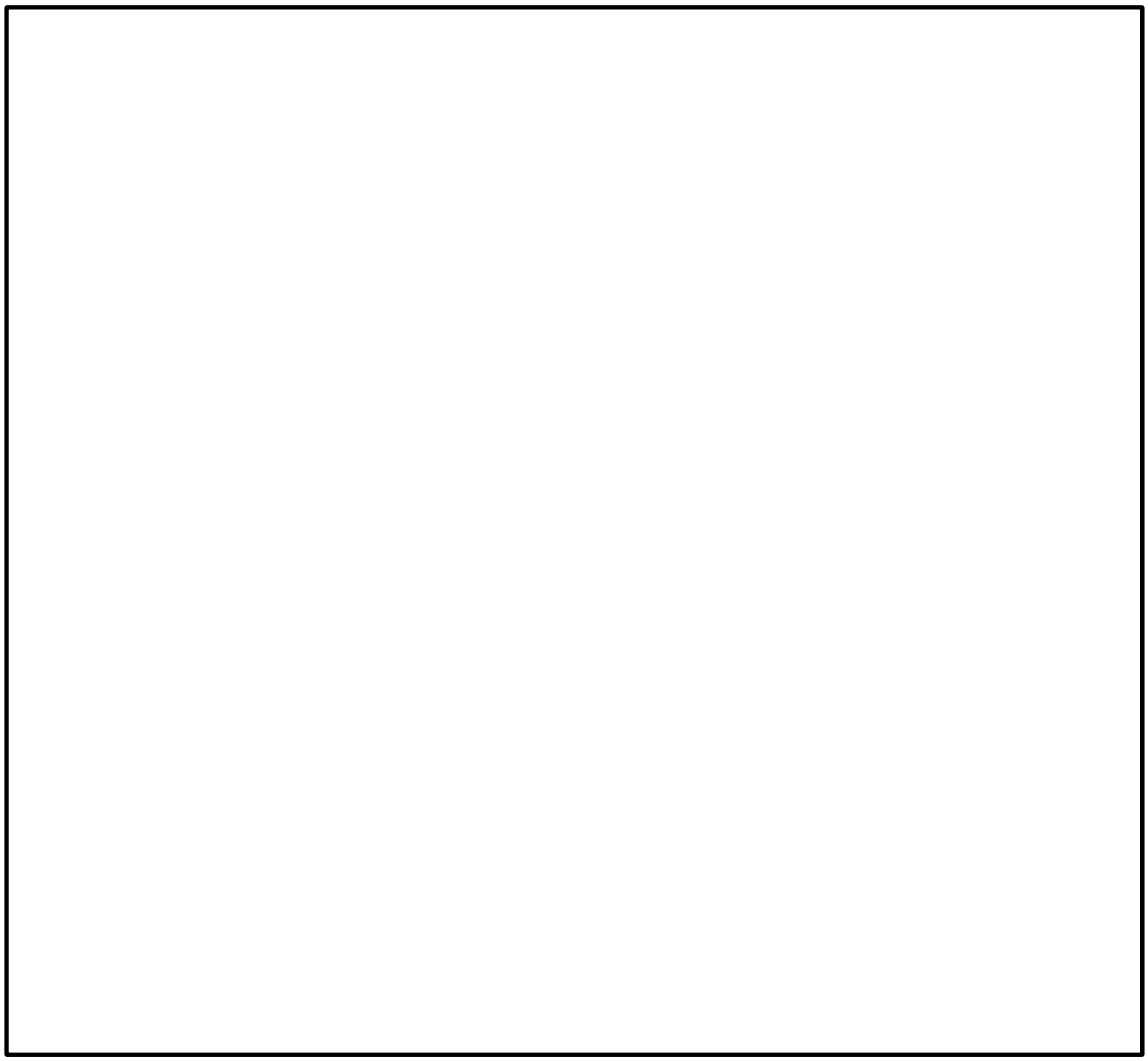

Figure 11 [redacted]

cT1=iron corrected longitudinal relaxation time; r=Spearman's rank correlation coefficient

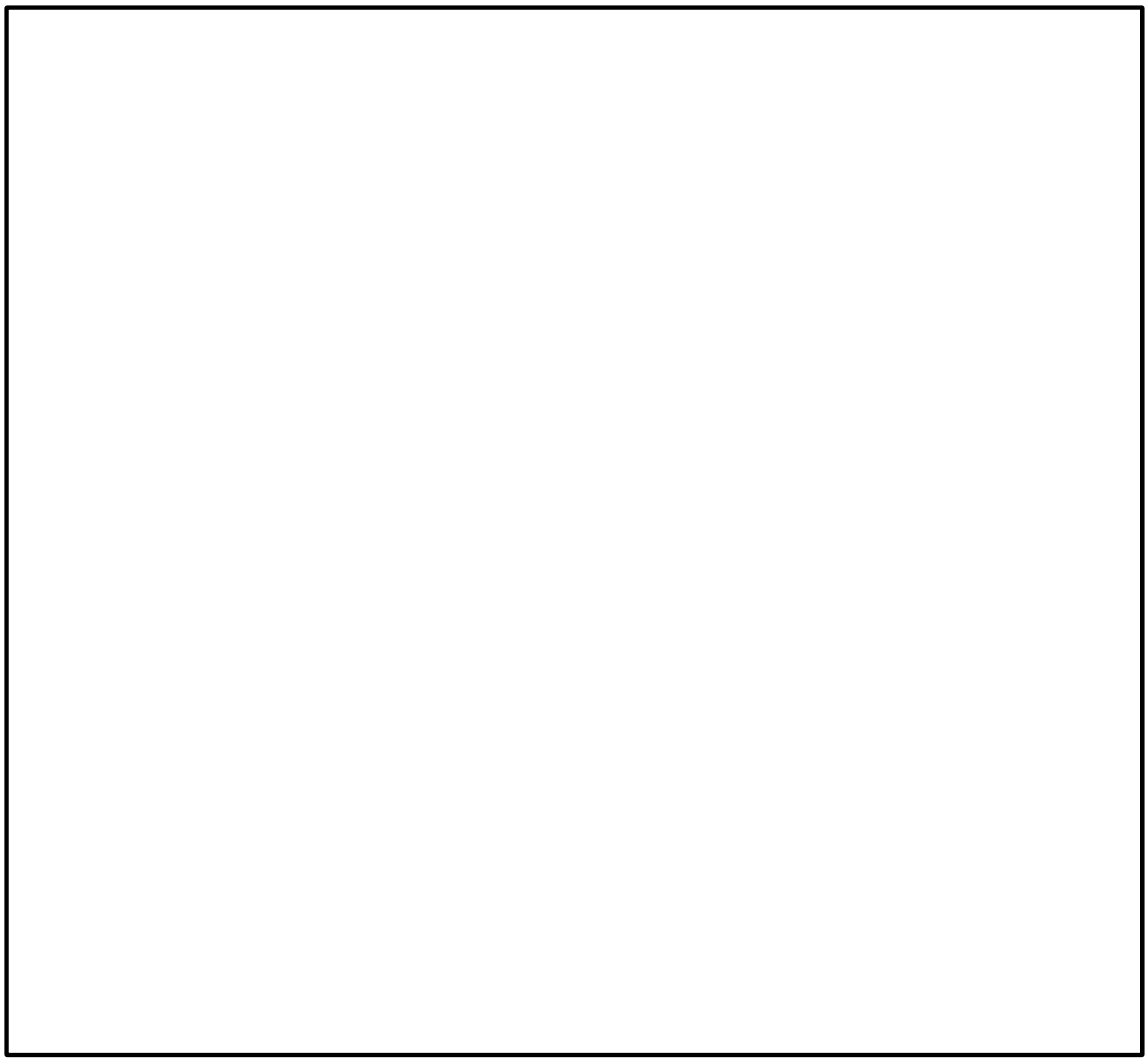

Figure 12 [redacted]

PDFF=proton density fat fraction; r=Spearman's rank correlation coefficient

Appendix 14 Results from the EAG meta-analysis for test failure rate

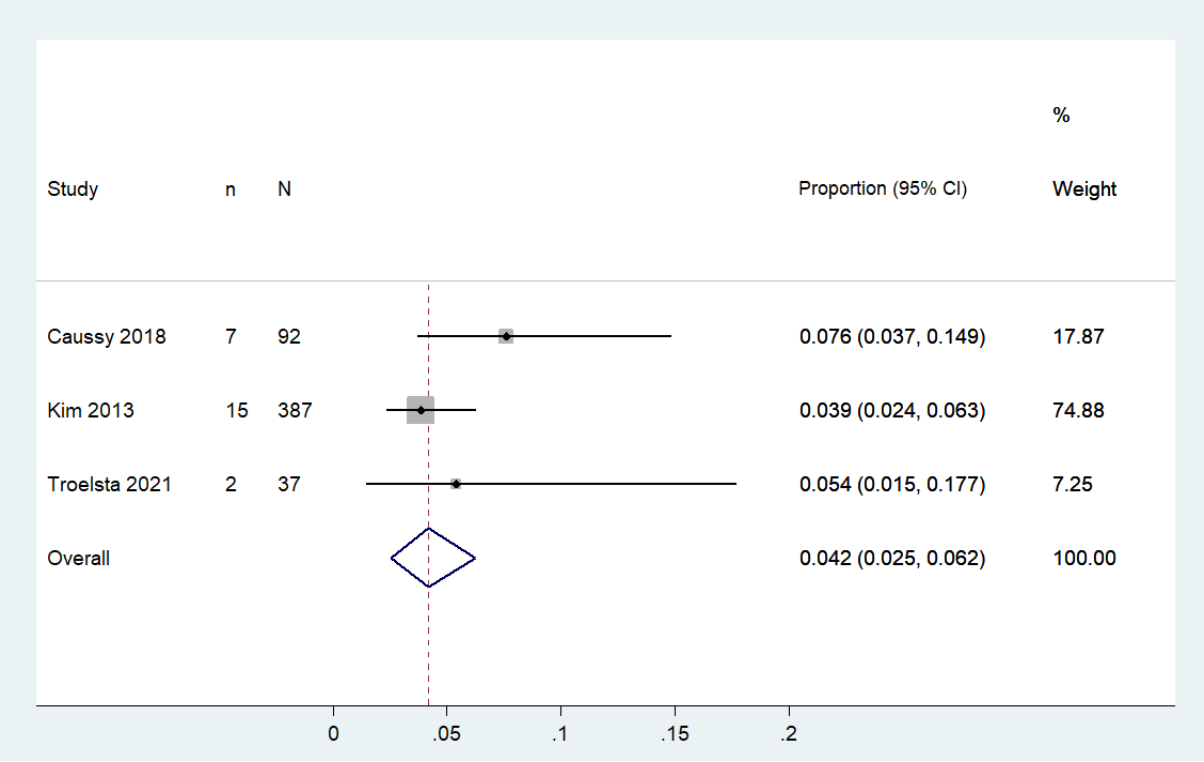

Figure 13 Forest plot displaying the EAG meta-analysis for test failure rate of MRE

CI=confidence interval; n=number of test failures; N=total number of tests

Appendix 15 Search strategies cost effectiveness

MEDLINE (via Ovid)

1 exp Non-alcoholic Fatty Liver Disease/
2 non-alcoholic fatty liver disease.tw,kw.
3 NAFLD.tw,kw.
4 non-alcoholic steatohepatitis.tw,kw.
5 NASH.tw,kw.
6 metabolic dysfunction associated fatty liver disease.tw,kw.
7 MAFLD.tw,kw.
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 exp Magnetic Resonance Imaging/
10 MRI.tw,kw.
11 magnetic resonance imag*.tw,kw.
12 LiverMultiScan.tw,kw.
13 Magnetic resonance elastograph*.tw,kw.
14 MRE.tw,kw.
15 9 or 10 or 11 or 12 or 13 or 14
16 8 and 15
17 Economics/
18 exp "Costs and Cost Analysis"/
19 Economics, Nursing/
20 Economics, Medical/
21 Economics, Pharmaceutical/
22 exp Economics, Hospital/
23 Economics, Dental/
24 exp "Fees and Charges"/
25 exp Budgets/
26 budget*.ti,ab,kf.
27 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kf.
28 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab.
29 (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf.
30 (value adj2 (money or monetary)).ti,ab,kf.

31 exp models, economic/
32 economic model*.ab,kf.
33 markov chains/
34 markov.ti,ab,kf.
35 monte carlo method/
36 monte carlo.ti,ab,kf.
37 exp Decision Theory/
38 (decision* adj2 (tree* or analy* or model*)).ti,ab,kf.
39 or/17-38
40 16 and 39
41 limit 40 to english language

Embase (via Ovid)

1 exp nonalcoholic fatty liver/
2 non-alcoholic fatty liver disease.tw,kw.
3 NAFLD.tw,kw.
4 exp nonalcoholic steatohepatitis/
5 non-alcoholic steatohepatitis.tw,kw.
6 NASH.tw,kw.
7 exp metabolic fatty liver/
8 metabolic dysfunction associated fatty liver disease.tw,kw.
9 MAFLD.tw,kw.
10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
11 exp nuclear magnetic resonance imaging/
12 MRI.tw,kw.
13 magnetic resonance imag*.tw,kw.
14 LiverMultiScan.tw,kw.
15 exp magnetic resonance elastography/
16 Magnetic resonance elastograph*.tw,kw.
17 MRE.tw,kw.
18 11 or 12 or 13 or 14 or 15 or 16 or 17
19 10 and 18
20 Economics/
21 Cost/
22 exp Health Economics/
23 Budget/
24 budget*.ti,ab,kw.

25 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ti,kw.
26 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed).ab.
27 (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kw.
28 (value adj2 (money or monetary)).ti,ab,kw.
29 Statistical Model/
30 economic model*.ab,kw.
31 Probability/
32 markov.ti,ab,kw.
33 monte carlo method/
34 monte carlo.ti,ab,kw.
35 Decision Theory/
36 Decision Tree/15762
37 (decision* adj2 (tree* or analy* or model*)).ti,ab,kw.
38 or/20-37
39 19 and 38
40 limit 39 to english language
41 limit 40 to embase

Cochrane Central Database of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) (via The Cochrane Library)

1 MeSH descriptor: [Non-alcoholic Fatty Liver Disease] explode all trees
2 ("non-alcoholic fatty liver disease"):ti,ab,kw
3 (NAFLD):ti,ab,kw
4 ("non-alcoholic steatohepatitis"):ti,ab,kw
5 (NASH):ti,ab,kw
6 ("metabolic dysfunction associated fatty liver disease"):ti,ab,kw
7 (MAFLD):ti,ab,kw
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 MeSH descriptor: [Magnetic Resonance Imaging] explode all trees
10 (MRI):ti,ab,kw
11 (magnetic NEXT resonance NEXT imag*):ti,ab,kw
12 (LiverMultiScan):ti,ab,kw

13 (Magnetic resonance elastograph*):ti,ab,kw
14 (MRE):ti,ab,kw
15 9 or 10 or 11 or 12 or 13 or 14
16 #8 AND #15
17 MeSH descriptor: [Economics] this term only
18 MeSH descriptor: [Costs and Cost Analysis] explode all trees
19 MeSH descriptor: [Economics, Nursing] this term only
20 MeSH descriptor: [Economics, Medical] this term only
21 MeSH descriptor: [Economics, Pharmaceutical] this term only
22 MeSH descriptor: [Economics, Hospital] explode all trees
23 MeSH descriptor: [Economics, Dental] this term only
24 MeSH descriptor: [Fees and Charges] explode all trees
25 MeSH descriptor: [Budgets] explode all trees
26 (budget*):ti,ab,kw
27 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed):ti,kw
28 (economic* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic* or pharmaco-economic* or expenditure or expenditures or expense or expenses or financial or finance or finances or financed):ab
29 (cost* NEAR/2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)):ab,kw
30 ((value NEAR/2 (money or monetary))):ti,ab,kw
31 MeSH descriptor: [Models, Economic] explode all trees
32 (economic NEXT model*):ab,kw
33 MeSH descriptor: [Markov Chains] this term only
34 (markov):ti,ab,kw
35 MeSH descriptor: [Monte Carlo Method] this term only
36 ("monte carlo"):ti,ab,kw
37 MeSH descriptor: [Decision Theory] explode all trees
38 ((decision* NEAR/2 (tree* or analy* or model*))):ti,ab,kw
39 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38
40 #16 AND #39

Database of Abstracts of Reviews of Effects (DARE) (via Centre for Reviews and Dissemination)

1 MeSH DESCRIPTOR Non-alcoholic Fatty Liver Disease EXPLODE ALL TREES
2 ("non-alcoholic fatty liver disease")
3 (NAFLD)
4 (non-alcoholic steatohepatitis)
5 (NASH)
6 ("metabolic dysfunction associated fatty liver disease")
7 (MAFLD)
8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
9 MeSH DESCRIPTOR Magnetic Resonance Imaging EXPLODE ALL TREES
10 (MRI)
11 ("magnetic resonance imag*")
12 (LiverMultiScan)
13 ("Magnetic resonance elastograph*")
14 (MRE)
15 #9 OR #10 OR #11 OR #12 OR #13 OR #14
16 #8 AND #15

Health Technology Assessment Database (HTA) (via International HTA Database)

(MAFLD) OR ("metabolic dysfunction associated fatty liver disease") OR (NASH) OR ("non-alcoholic steatohepatitis") OR (NAFLD) OR ("non-alcoholic fatty liver disease") OR ("Non-alcoholic Fatty Liver Disease"[mhe])

EconLit (via EBSCO)

S1 TI "non-alcoholic fatty liver disease" OR AB "non-alcoholic fatty liver disease" OR SU "non-alcoholic fatty liver disease")
S2 TI NAFLD OR AB NAFLD OR SU NAFLD
S3 TI "non-alcoholic steatohepatitis" OR AB "non-alcoholic steatohepatitis" OR SU "non-alcoholic steatohepatitis"
S4 TI NASH OR AB NASH OR SU NASH
S5 TI "metabolic dysfunction associated fatty liver disease" OR AB "metabolic dysfunction associated fatty liver disease" OR SU "metabolic dysfunction associated fatty liver disease"
S6 TI MAFLD OR AB MAFLD OR SU MAFLD
S7 TI MRI OR AB MRI OR SU MRI
S8 TI "magnetic resonance imag*" OR AB "magnetic resonance imag*" OR SU "magnetic resonance imag*"
S9 TI LiverMultiScan OR AB LiverMultiScan OR SU LiverMultiScan

S10 TI "Magnetic resonance elastograph* OR AB "Magnetic resonance elastograph* OR SU "Magnetic resonance elastograph*

S11 TI MRE OR AB MRE OR SU MRE

S12 S1 OR S2 OR S3 OR S4 OR S5 OR S6

S13 S7 OR S8 OR S9 OR S10 OR S11

S14 S12 AND S13

Cost Effectiveness Analysis (CEA) registry

non-alcoholic fatty liver disease

NAFLD

non-alcoholic steatohepatitis

NASH

metabolic dysfunction associated fatty liver disease

MAFLD

Appendix 16 Excluded studies for the cost effectiveness review

1. Alisi A, Nobili V. Sensitive non-invasive circulating markers in paediatric non-alcoholic fatty liver disease. Pediatric Obesity 2012; 7:89–91. **Wrong intervention**

2. Ando Y, Jou JH. Nonalcoholic Fatty Liver Disease and Recent Guideline Updates. Clinical Liver Disease 2021;17(1):23-28. **Wrong intervention**

3. Blake L, Duarte RV, Cummins C. Decision analytic model of the diagnostic pathways for patients with suspected non-alcoholic fatty liver disease using non-invasive transient elastography and multiparametric magnetic resonance imaging. BMJ Open 2016;6(9):e010507. **Wrong study design**

4. Boursier J, Cales P. Controlled attenuation parameter (CAP): A new device for fast evaluation of liver fat? Liver International 2012; 32(6):875-877. **Wrong intervention**

5. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2019;156(5):1264. **Wrong study design**

6. Chen J, Yin M, Glaser KJ, Talwalkar JA, Ehman RL. MR elastography of liver disease: State of the art. Applied Radiology 2013;42(4):5-12. **Wrong intervention**

7. Cleveland E, Bandy A, VanWagner LB. Diagnostic challenges of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. Clinical Liver Disease 2018;11(4):98-104. **Wrong intervention**

8. Crossan C, Longworth L, Tsochatzis EA, Rodriguez-Peralvarez M, Mantzoukis K, O'Brien J 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 Technology Assessment 2015;19(9):1-458. **Wrong intervention**

9. de Alwis NMW, Anstee QM, Day CP. How to Diagnose Nonalcoholic Fatty Liver Disease. Digestive diseases 2016;34 Suppl 1(dds, 8701186):19-26.**Wrong intervention**

10. Degnan AJ, Serai SD, Anupindi SA, Panganiban J, Dhyani M. Imaging Modalities in Pediatric NAFLD. Clinical Liver Disease 2021;17(3):200-208. **Wrong intervention**

11. Jiang ZG, Tapper EB. Cost Saving or Cost Effective? Unanswered Questions in the Screening of Patients With Nonalcoholic Fatty Liver Disease. Hepatology Communications 2019;3(10):1293-1295. **Wrong study design**

12. Kayadibi H, Sertoglu E, Uyanik M. Biochemical Markers, Liver Biopsy, or Magnetic Resonance Elastography to Detect or Exclude Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Hepatology 2015;62(1):324-325. **Wrong intervention**

13. Lee SS, Park SH. Radiologic evaluation of nonalcoholic fatty liver disease. World journal of gastroenterology 2014;20(23):7392-402. **Wrong study design**

14. Martinez SM, Crespo G, Navasa M, Forns X. Noninvasive assessment of liver fibrosis. Hepatology 2011;53(1):325-335. **Wrong intervention**

15. Mishra A, Younossi ZM, Bush H, Henry L. Clinical and Economic Burden of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Clinics in Liver Disease 2018;22(1):1-10. **Wrong intervention**

16. Nathan R, Jain D, Rossi S. CON: This Patient Should Have a Noninvasive Assessment of Liver Staging. Clinical Liver Disease / 2019;14(3):116-120. **Wrong intervention**

17. NCT03289897. Non-invasive Rapid Assessment of NAFLD Using Magnetic Resonance Imaging With LiverMultiScan 2017. **Wrong study design**

18. Noureddin M, Khoyilar C, Palmer SL. MRI, CT scan, and ultrasound in the diagnosis of nonalcoholic fatty liver disease. Journal of Clinical Gastroenterology 2015;49(4):351-352. **Wrong intervention**

19. Paul S, Davis AM. Diagnosis and Management of Nonalcoholic Fatty Liver Disease. JAMA 2018; 320:23:2474-2475. **Wrong intervention**

20. Ronot M, Vilgrain V. Multiparametric magnetic resonance imaging in patients with chronic liver disease: Are we there yet? Liver International 2016;36(5):631-633. **Wrong intervention**

21. Shiha G, Ibrahim A, Sarin S, Kumar M, Omata M, Hemy A, et al. Asian-Pacific Association for the Study of the Liver (APASL) consensus guidelines on invasive and non-invasive assessment of hepatic fibrosis: a 2016 update. Hepatology International 2017;11(1). **Wrong study design**

22. Stoopen-Rometti M, Ramirez-Carmona CR, Kimura-Hayama E, Saavedra-Abril JA, Encinas-Escobar ER, Wolpert-Barraza E, et al. Diagnosis and quantification of fibrosis, steatosis, and hepatic siderosis through multiparametric magnetic resonance imaging. Revista de Gastroenterologia de Mexico 2017;82(1):32-45. **Wrong intervention**

23. Taouli B, Serfaty L. Magnetic Resonance Imaging/Elastography Is Superior to Transient Elastography for Detection of Liver Fibrosis and Fat in Nonalcoholic Fatty Liver Disease. Gastroenterology / 2016;150(3):553-556. **Wrong intervention**

24. Tonev D, Shumbayawonda E, Tetlow LA, Herdman L, French M, Rymell S et al. The Effect of Multi-Parametric Magnetic Resonance Imaging in Standard of Care for Nonalcoholic Fatty Liver Disease: Protocol for a Randomized Control Trial. JMIR research protocols 2020;9(10):e19189. **Wrong study design**

25. Vilar-Gomez E, Vuppalanchi R, Chalasani N, Lou Z, Kong N, Imperiale TF. Cost Effectiveness of Different Strategies for Detecting Cirrhosis in Patients With Nonalcoholic Fatty Liver Disease Based on United States Health Care System. Clinical Gastroenterology and Hepatology 2020;18(10):2305. **Wrong intervention**

26. Vuppalanchi R, Chalasani N. Screening Strategies for Nonalcoholic Steatohepatitis in High-Risk Individuals: Trimming Away the Fat. Digestive Diseases and Sciences 2016;61(7):1790-1792. **Wrong study design**

27. Zhang E, Wartelle-Bladou C, Lepanto L, Lachaine J, Cloutier G,Tang A. Cost-utility analysis of nonalcoholic steatohepatitis screening. European Radiology 2015;25(11):3282-3294. **Wrong patient population**

28. Zhang J, Cai J-J, Yu Y, She Z-G; Li H. Nonalcoholic fatty liver disease: An update on the diagnosis. Gene Expression The Journal of Liver Research 2019;19(3):187-198. **Wrong intervention**

29. Zhou J-H, She Z-G. Li H-L, Cai J-J. Noninvasive evaluation of nonalcoholic fatty liver disease: Current evidence and practice. World Journal of Gastroenterology 2019;25(11):1307-1326. **Wrong intervention**

Appendix 17 CHEERS checklist[77] summary of the included study in the EAG's review of economic evidence

Section	Recommendation	Eddowes 2018[29]
Title and abstract		
Title	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Yes, page 631
Abstract	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Yes, page 631
Introduction		
Background and objectives	Provide an explicit statement of the broader context for the study	Yes, page 632
	Present the study question and its relevance for health policy or practice decisions.	Yes, page 631
Methods		
Target population and subgroup	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Yes, pages 632 and 634
Setting and location	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Yes, pages 632, 634
Study perspective	Describe the perspective of the study and relate this to the costs being evaluated.	Yes, page 634 (Decision analytic model)
Comparators	Describe the interventions or strategies being compared and state why they were chosen.	Yes, page 634 (Decision analytic model)
Time horizon	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Not reported but assumed to be short
Discount rate	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Not applied
Choice of health outcomes	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Yes, page 634 (Decision Analytic Model)
Measurement of effectiveness	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not reported
Measurement and valuation of preference-based outcomes	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not used

Estimating resources and costs	Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Total included including cost per diagnosis
Currency, price date, and conversion	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Currency is stated but price data and any conversion necessary not reported
Choice of model	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	Yes, page 634
Assumptions	Describe all structural or other assumptions underpinning the decision-analytical model.	Yes, page 634
Analytical methods	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Yes, Supplement 1
Results		
Study parameters	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Yes, page 634 (Decision Analytic Model)
Incremental costs and outcomes	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Not reported
Characterising uncertainty	Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Not reported
Characterising heterogeneity	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information	Not reported
Discussion		
Study findings, limitations, generalisability, and current knowledge	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge	Not reported in terms of economic evaluation
Other		
Source of funding	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Yes, page 642

Conflicts of interest	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Yes, page 642

Appendix 18 LiverMultiScan PDFF results

Table 26 Initial LiverMultiScan outcomes generated by the EAG model (per 1,000 tests)

Diagnostic test strategy	PDFF cut-off value	True Positive	True Negative	False Positive	False Negative	Failed tests
T1: Any fibrosis (≥F1)	>5%	657.4	61.6	61.6	164.3	55.0
T2: Significant fibrosis (≥F2)	>10%	349.2	164.3	164.3	267.1	55.0
T3: Advanced fibrosis (≥F3)	>10%	226.0	287.6	205.4	226.0	55.0
T4: Brunt Grade ≥1	>5%	698.5	0.0	20.5	226.0	55.0
T5: Brunt Grade ≥2	>10%	369.8	328.7	143.8	102.7	55.0
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10%	328.7	246.5	184.9	184.9	55.0
T7: Advanced NASH (NAS≥4 plus ≥F2)	>10%	287.6	267.1	226.0	164.3	55.0

EAG=External Assessment Group; F=stage of fibrosis; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction

Table 27 LiverMultiScan plus biopsy pathway: biopsies performed and averted (per 1,000 patients)

Diagnostic test strategy	PDFF cut-off value	Total number of biopsies, including those following a repeated LiverMultiScan at 6 months	Biopsies averted
T1: Any fibrosis (≥F1)	>5%	938.4	61.6
T2: Significant fibrosis (≥F2)	>10%	835.7	164.3
T3: Advanced fibrosis (≥F3)	>10%	712.4	287.6
T4: Brunt Grade ≥1	>5%	1000.0	0.0
T5: Brunt Grade ≥2	>10%	671.3	328.7
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10%	753.5	246.5
T7: Advanced NASH (NAS≥4 plus ≥F2)	>10%	732.9	267.1

F=stage of fibrosis; MRI=magnetic resonance imaging; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction

Table 28 Pathway diagnostic test strategy costs (per 1,000 patients)

Diagnostic test strategy	PDFF cut-off value	LiverMultiScan plus biopsy pathway costs				Biopsy only pathway costs			Additional cost for the LMS pathway
		Biopsy procedures	Biopsy complications	LiverMultiScan test	Total costs	Biopsy procedures	Biopsy complications	Total costs	
T1: Any fibrosis (≥F1)	>5%	£755,388	£8,014	£425,709	£1,189,110	£805,000	£8,540	£813,540	£375,570
T2: Significant fibrosis (≥F2)	>10%	£672,700	£7,136	£497,044	£1,176,880	£805,000	£8,540	£813,540	£363,340
T3: Advanced fibrosis (≥F3)	>10%	£573,475	£6,084	£525,578	£1,105,137	£805,000	£8,540	£813,540	£291,597
T4: Brunt Grade ≥1	>5%	£805,000	£8,540	£425,709	£1,239,249	£805,000	£8,540	£813,540	£425,709
T5: Brunt Grade ≥2	>10%	£540,400	£5,733	£497,044	£1,043,177	£805,000	£8,540	£813,540	£229,637
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10%	£606,550	£6,435	£497,044	£1,110,029	£805,000	£8,540	£813,540	£296,489
T7: Advanced NASH (NAS≥4 plus ≥F2)	>10%	£590,013	£6,259	£497,044	£1,093,316	£805,000	£8,540	£813,540	£279,776

F=stage of fibrosis; LMS=LiverMultiScan; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction

Table 29 QALY analyses for the two diagnostic pathways (per 1,000 patients)

Diagnostic test strategy	PDFF cut-off value	LiverMultiScan plus biopsy pathway					Biopsy only pathway				Difference in QALY losses (LMS+biopsy pathway)*
		Biopsy procedure	Biopsy complications	Biopsy death	False negatives	Total QALY losses	Biopsy procedure	Biopsy complications	Biopsy death	Total QALY losses	
T1: Any fibrosis (≥F1)	>5%	5.2	0.1	1.3	2.5	9.2	5.6	0.1	1.4	7.1	-2.0
T2: Significant fibrosis (≥F2)	>10%	4.7	0.1	1.2	4.0	10.0	5.6	0.1	1.4	7.1	-2.8
T3: Advanced fibrosis (≥F3)	>10%	4.0	0.1	1.0	3.4	8.5	5.6	0.1	1.4	7.1	-1.3
T4: Brunt Grade ≥1	>5%	5.6	0.1	1.4	3.4	10.5	5.6	0.1	1.4	7.1	-3.4
T5: Brunt Grade ≥2	>10%	3.7	0.1	0.9	1.5	6.3	5.6	0.1	1.4	7.1	0.8
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10%	4.2	0.1	1.1	2.8	8.2	5.6	0.1	1.4	7.1	-1.0
T7: Advanced NASH (NAS≥4 plus ≥F2)	>10%	4.1	0.1	1.0	2.5	7.7	5.6	0.1	1.4	7.1	-0.6

* A positive value means that the biopsy only pathway is preferred; a negative value means that the LiverMultiScan plus biopsy pathway is preferred
LMS=LiverMultiScan; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction QALY=quality adjusted life year

Table 30 Incremental analyses for LiverMultiScan plus biopsy versus biopsy (1,000 patients)

Diagnostic test strategy Fibrosis	PDFF cut-off value	Incremental		ICER per QALY gained (versus biopsy)
		Costs	QALYs	
T1: Any fibrosis (≥F1)	>5%	£375,570	-2.0	LMS+biopsy dominated by biopsy
T2: Significant fibrosis (≥F2)	>10%	£363,340	-2.8	LMS+biopsy dominated by biopsy
T3: Advanced fibrosis (≥F3)	>10%	£291,597	-1.3	LMS+biopsy dominated by biopsy
T4: Brunt Grade ≥1	>5%	£425,709	-3.4	LMS+biopsy dominated by biopsy
T5: Brunt Grade ≥2	>10%	£229,637	0.8	£285,214
T6: NASH (NAS≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10%	£296,489	-1.0	LMS+biopsy dominated by biopsy
T7: Advanced NASH (NAS≥4 plus ≥F2)	>10%	£279,776	-0.6	LMS+biopsy dominated by biopsy

F=stage of fibrosis; ICER=incremental cost per QALY gained; NAS=non-alcoholic fatty liver disease (NAFLD) activity score; NASH=non-alcoholic steatohepatitis; PDFF=proton density fat fraction QALY=quality adjusted life year