NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 27 • Issue 10 • July 2023

ISSN 1366-5278

Diagnostic accuracy and clinical impact of MRI-based technologies for patients with non-alcoholic fatty liver disease: systematic review and economic evaluation

Rebecca Bresnahan, Rui Duarte, James Mahon, Sophie Beale, Marty Chaplin, Devarshi Bhattacharyya, Rachel Houten, Katherine Edwards, Sarah Nevitt, Michelle Maden and Angela Boland

DOI 10.3310/KGJU3398

Diagnostic accuracy and clinical impact of MRI-based technologies for patients with non-alcoholic fatty liver disease: systematic review and economic evaluation

Rebecca Bresnahan,[1*] Rui Duarte,[1] James Mahon,[2] Sophie Beale,[3] Marty Chaplin,[1] Devarshi Bhattacharyya,[1] Rachel Houten,[1] Katherine Edwards,[1] Sarah Nevitt,[1] Michelle Maden[1] and Angela Boland[1]

[1]LRiG, Department of Health Data Science, University of Liverpool, Liverpool, UK
[2]Coldingham Analytical Services, Berwickshire, UK
[3]Hare Research, North Yorkshire, UK

*Corresponding author

Disclosure of interests of authors

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/KGJU3398.

Primary conflicts of interest: Rui Duarte is a member of the HTA Commissioning Committee (May 2021 to May 2025).

Published July 2023
DOI: 10.3310/KGJU3398

This report should be referenced as follows:

Bresnahan R, Duarte R, Mahon J, Beale S, Chaplin M, Bhattacharyya D, *et al*. Diagnostic accuracy and clinical impact of MRI-based technologies for patients with non-alcoholic fatty liver disease: systematic review and economic evaluation. *Health Technol Assess* 2023;**27**(10). https://doi.org/10.3310/KGJU3398

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4.014 and is ranked 27th (out of 108 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2021 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), Embase (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded (Clarivate , Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number NIHR135067. The contractual start date was in October 2021. The draft report began editorial review in March 2022 and was accepted for publication in July 2022. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

NIHR Journals Library Editor-in-Chief

NIHR Journals Library Editors

Abstract

Diagnostic accuracy and clinical impact of MRI-based technologies for patients with non-alcoholic fatty liver disease: systematic review and economic evaluation

Rebecca Bresnahan,[1*] Rui Duarte,[1] James Mahon,[2] Sophie Beale,[3] Marty Chaplin,[1] Devarshi Bhattacharyya,[1] Rachel Houten,[1] Katherine Edwards,[1] Sarah Nevitt,[1] Michelle Maden[1] and Angela Boland[1]

[1]LRiG, Department of Health Data Science, University of Liverpool, Liverpool, UK
[2]Coldingham Analytical Services, Berwickshire, UK
[3]Hare Research, North Yorkshire, UK

[*]Corresponding author rebecca.bresnahan@liverpool.ac.uk

Background: Magnetic resonance imaging-based technologies are non-invasive diagnostic tests that can be used to assess non-alcoholic fatty liver disease.

Objectives: The study objectives were to assess the diagnostic test accuracy, clinical impact and cost-effectiveness of two magnetic resonance imaging-based technologies (LiverMultiScan and magnetic resonance elastography) for patients with non-alcoholic fatty liver disease for whom advanced fibrosis or cirrhosis had not been diagnosed and who had indeterminate results from fibrosis testing, or for whom transient elastography or acoustic radiation force impulse was unsuitable, or who had 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: A systematic review was conducted using established methods. Diagnostic test accuracy estimates were calculated using bivariate models and a summary receiver operating characteristic curve was calculated using a hierarchical model. A simple decision-tree model was developed to generate cost-effectiveness results.

Results: The diagnostic test accuracy review (13 studies) and the clinical impact review (11 studies) only included one study that provided evidence for patients who had indeterminate or discordant results from fibrosis testing. No studies of patients for whom transient elastography or acoustic radiation force impulse were unsuitable were identified.

Depending on fibrosis level, relevant published LiverMultiScan diagnostic test accuracy results ranged from 50% to 88% (sensitivity) and from 42% to 75% (specificity). No magnetic resonance elastography diagnostic test accuracy data were available for the specific population of interest.

Results from the clinical impact review suggested that acceptability of LiverMultiScan was generally positive.

To explore how the decision to proceed to biopsy is influenced by magnetic resonance imaging-based technologies, the External Assessment Group presented cost-effectiveness analyses for LiverMultiScan plus biopsy versus biopsy only. Base-case incremental cost-effectiveness ratio per quality-adjusted life year gained results for seven of the eight diagnostic test strategies considered showed that LiverMultiScan plus biopsy was dominated by biopsy only; for the remaining strategy (Brunt grade ≥2), the incremental cost-effectiveness ratio per quality-adjusted life year gained was £1,266,511. Results from threshold and scenario analyses demonstrated that External Assessment Group base-case results were robust to plausible variations in the magnitude of key parameters.

Limitations: Diagnostic test accuracy, clinical impact and cost-effectiveness data for magnetic resonance imaging-based technologies for the population that is the focus of this assessment were limited.

Conclusions: Magnetic resonance imaging-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 diagnostic test accuracy and clinical impact data for patients who have indeterminate results from fibrosis testing, for whom transient elastography or acoustic radiation force impulse are unsuitable or who had discordant results from fibrosis testing.

Given the External Assessment Group cost-effectiveness analyses assumptions, the use of LiverMultiScan and magnetic resonance elastography for assessing non-alcoholic fatty liver disease for patients with inconclusive results from previous fibrosis testing is unlikely to be a cost-effective use of National Health Service resources compared with liver biopsy only.

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 and Care Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 27, No. 10. See the NIHR Journals Library website for further project information.

Contents

List of tables

List of figures

List of supplementary material

Supplementary material 1

Supplementary material 2

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/KGJU3398).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

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
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
T1	longitudinal relaxation time
TE	transient elastography

TN	true negative	2 × 2 DATA	numbers of true positive, false positive, true negative and false negative test results
TP	true positive		

Plain language summary

Non-alcoholic fatty liver disease 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. It is important to identify patients with fibrosis because severe fibrosis can cause permanent liver damage (cirrhosis), which can lead to liver failure and liver cancer.

In the National Health Service, patients with non-alcoholic fatty liver disease 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 have a very high body mass index.

In the National Health Service, 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-based testing could be used as an extra test to help clinicians assess non-alcoholic fatty liver disease and identify patients who may need a liver biopsy.

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

We reviewed all studies examining how well LiverMultiScan and magnetic resonance elastography assess patients with non-alcoholic fatty liver disease. 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 National Health Service.

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 National Health Service 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 4 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 was 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 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 4 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.

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 iron-corrected T1 (cT1) outputs, 2 × 2 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: magnetic resonance elastography

For the MRE test, 2 × 2 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.

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 reduce the number of unnecessary biopsies for patients with non-NAFLD and NAFLD to diagnose non-alcoholic steatohepatitis (NASH) and fibrosis unrelated to NAFLD [EAG calculated odds ratio (OR) = 0.65, 95% CI 0.22 to 1.96] and for patients with no to mild fibrosis (F0 to F1) to diagnose significant fibrosis to cirrhosis (F2 to F4; EAG calculated OR = 0.59, 95% CI 0.18 to 1.89) 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: magnetic resonance elastography

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.

Cost-effectiveness

The EAG base-case incremental cost-effectiveness ratio (ICER) per quality-adjusted life year (QALY) gained results for seven of the eight diagnostic test strategies considered, and showed that the LiverMultiScan plus biopsy pathway was dominated by the biopsy only pathway. For Brunt grade ≥2, the ICER per QALY gained was £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 is 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 was 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 and Care Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 27, No. 10. See the NIHR Journals Library website for further project information.

Background

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 National Health Service (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.

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 pain (12.9%; 1 in 8), minor bleeding (0.19%; 1 in 500), major pain (0.48%; 1 in 200), major bleeding (0.48%; 1 in 200) and death (0.01%; 1 in 10,000).[5] Liver biopsy complications lead to hospitalisation for 0.65% (1 in 150) of patients.[5]

It is estimated that between 20%[1] and 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]

TABLE 1 Non-alcoholic steatohepatitis CRN histological scoring system

NAFLD activity score (NAS)					
Steatosis (Brunt grade)		**Hepatocyte ballooning**		**Lobular inflammation (foci per 200× field)**	
Score	**Definition**	**Score**	**Definition**	**Score**	**Definition**
0	<5%	0	None	0	None
1	5–33%	1	Few	1	<2
2	34–66%	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				

Source: Kleiner *et al.* 2005.[17]

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 ≥4 indicates a diagnosis of NASH and a NAS ≥4 plus fibrosis ≥F2 indicates a diagnosis of advanced NASH.[18] 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]

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].[19] 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).[19]

Current National Health Service 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,[20] or following an incidental ultrasound finding.[9,21] 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[8,9,22,23] and expert advice to NICE.[24]

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.

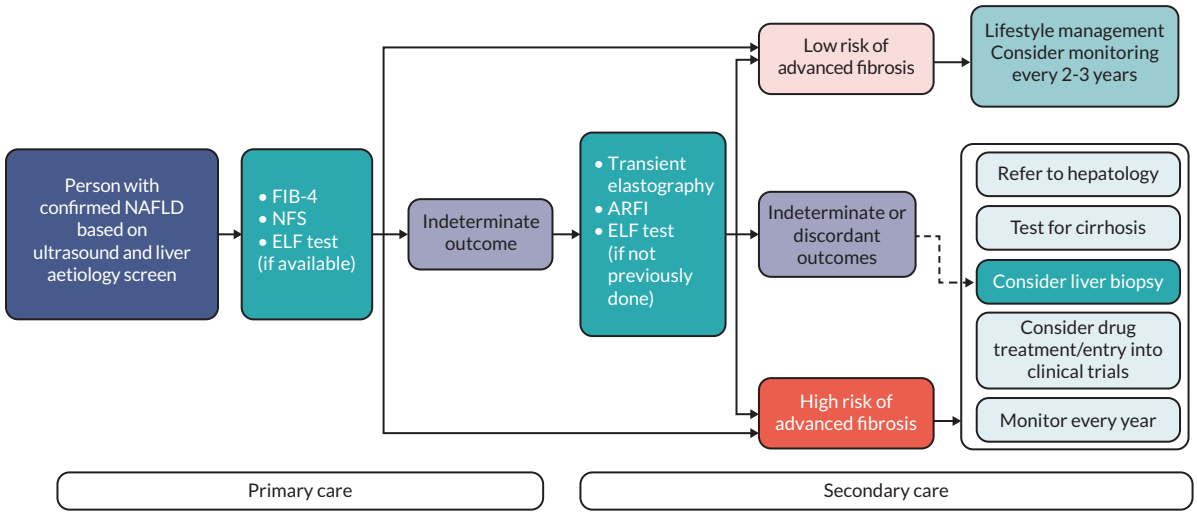

FIGURE 1 Overview of current diagnostic pathway for assessment of fibrosis in the NHS, based on guidelines and expert advice. FIB-4 = fibrosis-4; NFS = NAFLD fibrosis score. Source: Final scope[24] issued by NICE.

In the British Society of Gastroenterology (BSG) national guidelines,[22] 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,[22] 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[22] and the Lancet Commission into liver disease in the UK[25] 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.[25]

However, Byrne 2018[23] 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[22] 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.[26] 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 *Liver biopsy*.

Findings from a cross-sectional survey[26] 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[22] and used FIB-4 and NFS to assess liver fibrosis.

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.[22] However, novel therapies are in clinical development, such as glucagon-like peptide 1 agonists and sodium-glucose co-transporter 2 (SGLT2) inhibitors.[27]

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[28] 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[24] 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.

Population

In line with the final scope[24] 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 (e.g., based on prior tests for fibrosis, children or young people).

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',[29] 'intermediate risk'[22] and 'inconclusive results'[30]).

In the BSG guidelines,[22] it is recommended that clinicians should consider liver biopsy for patients with a TE score between 7.9 kPa and 9.6 kPa (intermediate risk of advanced fibrosis), and for patients with a TE score > 9.6 kPa (high risk of advanced fibrosis). In the European Association for the Study of the Liver (EASL) guidelines,[18] it is recommended that a TE score < 8 kPa rules out advanced fibrosis and that a TE score ≥ 8 kPa represents an intermediate to high risk of advanced fibrosis. Clinical advice to NICE[24] 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[24] is that indeterminate results are possible from the ELF test. ELF test scores between 7.8 and 10.5[23] or 7.7 and 9.7 are considered to be indeterminate results.[31] In the EASL guidelines,[18] 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.

Patients for whom transient elastography or acoustic radiation force impulse is unsuitable

TE and ARFI may not be suitable tests for people with a very high 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.[24] 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.

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[18] it is recommended that patients with discordant results, that is, patients for whom one non-invasive test indicates low risk of advanced fibrosis (e.g. TE < 8 kPa or ELF < 9.8) but another indicates intermediate to high risk of advanced fibrosis (e.g. TE ≥ 8 kPa or ELF ≥ 9.8), 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.

Interventions/index tests

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.[24]

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 cT1 score. 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 do not require any contrast agent or additional hardware in addition to the MRI scanner.[24] A 15 minute scan acquisition time is typically required to obtain the MR images

for analysis by LiverMultiScan software.[24] Training on how to use the LiverMultiScan protocol takes approximately 3 hours.[24] 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[24] 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:

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

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.[24] However, in the EASL guidelines[18] liver biopsy is recommended as the reference standard for the diagnosis of NASH for patients with NAFLD.

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 12–15 seconds, and is typically repeated four times.[24] The total acquisition time can last approximately 1 minute.[24] 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[24] 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.

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.

Comparator

In NHS clinical practice, the populations specified in the final scope[24] issued by NICE would not undergo any further investigation prior to deciding whether a biopsy was required. Clinical experts to NICE[24] 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).

Reference standard

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

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 (0.3–10.9%) and major intraperitoneal bleeding (0.1–4.6%).[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[24] 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 (5–10%) do not wish to proceed with liver biopsy, or are treated at centres without access to liver biopsy.

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 (see *Patients who have indeterminate results from fibrosis testing*)
- patients for whom TE or ARFI is unsuitable (see *Patients for whom TE or ARFI is unsuitable*)
- patients who have discordant results from fibrosis testing (see *Patients who have discordant results from fibrosis testing*).

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 DAP 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.

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 4 October 2021. Details of individual database searches are provided in *Appendix 1*; 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 CRD)
- 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).

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) were searched, and details of the additional searches are provided in *Appendix 2*.

Eligibility criteria

The review inclusion criteria are presented in *Table 2*.

TABLE 2 Review inclusion criteria

Parameter	Final scope[24] 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 vs. MRE or vs. no comparator MRE vs. 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 or 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 and uncontrolled single-arm studies

RCT = randomised controlled trial.
Source: Final scope[24] issued by NICE.

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.

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.

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[24] issued by NICE.

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 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).

Methods of analysis/synthesis of diagnostic test accuracy 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 3*.

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.

Subgroup analyses and sensitivity analyses

No subgroup analyses or sensitivity analyses were performed by the EAG (see *Appendix 3* for further details).

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 3*.

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.

Results of the assessment of diagnostic test accuracy and clinical impact

External Assessment Group 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 *Supplementary material 1*.

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 *Supplementary material 1* with reasons for exclusion.

Studies included in the External Assessment Group review

Thirteen studies[30,53–64] reported in 15 publications[30,31,53–65] were included in the DTA review. Two studies[30,59] reported in four publications[30,31,59,65] 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[30,53,54,57,59,62,64,66–69] reported in 14 publications[30,31,33,53,54,57,59,62,64–69] were included in the clinical impact review of MRI-based technologies. Five studies[30,59,66,68,69] reported in eight publications[30,31,33,59,65,66,68,69] 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[30,53–64] and ten of the 11 studies included in the clinical impact review[30,53,54,57,59,62,64,66–68] considered patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, only one study[30] 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.

Assessment of diagnostic test accuracy

Quality assessment

The included studies that provided DTA[30,53–64] 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 *Supplementary material 2*.

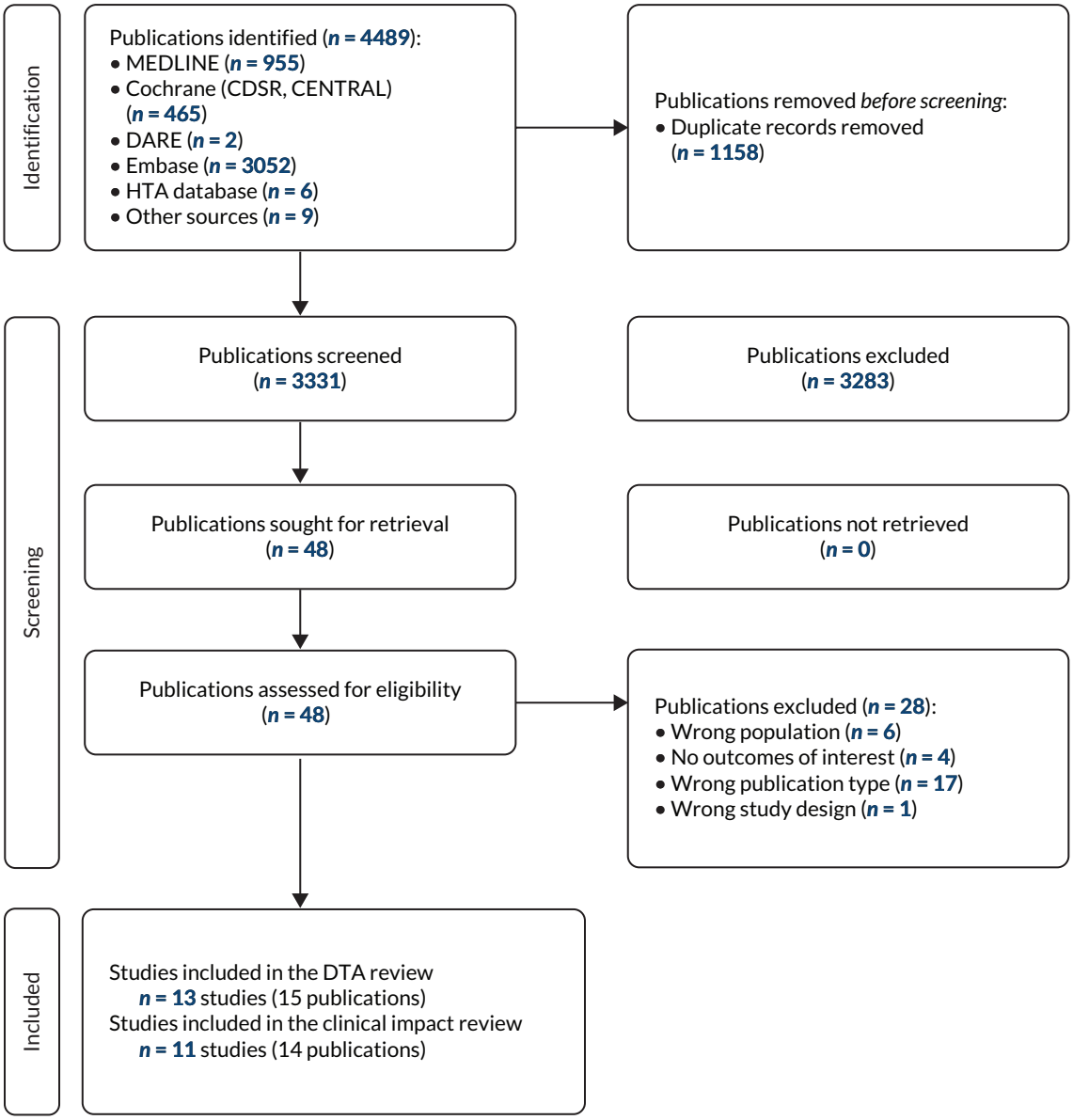

FIGURE 2 PRISMA flow diagram. Total number of studies included in the review *n* = 17 studies (20 publications).

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[30,55,57–64] 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.

TABLE 3 QUality Assessment of Diagnostic Accuracy Studies-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[30]	☺	?	☺	☺	☺	☺	☺
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.

Applicability concerns

Only one study[30] raised no concerns regarding the applicability of the study population or the index test to the review. The Eddowes 2018[30] 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[30] 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 risks 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.

Characteristics of the included studies

The characteristics of the 13 studies[30,53–64] included in the DTA review are presented in *Table 4*.

In line with the final scope[24] issued by NICE, all the studies[30,53–64] included patients with NAFLD for whom advanced fibrosis or cirrhosis had not yet been diagnosed. However, only the Eddowes study[30] 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[30] 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[30] 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[30,59] 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[30,59] 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 the 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[30,53,57] 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 (*Table 5*). As a result, the EAG quantitative synthesis includes data from only six of the identified studies.[30,56–59,62]

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. 2 × 2 data) were not presented in any of the included studies. We contacted the authors of all included studies to request these data.

Perspectum Ltd provided 2 × 2 data in response to the EAG request for information for the three LiverMultiScan studies[30,56,59] included in the DTA review. The authors of the Troelstra 2021[62] study of MRE provided 2 × 2 data in response to the EAG request. Data from the Kim 2020[58] study were obtained from a systematic review, and 2 × 2 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*.

The EAG's quantitative synthesis therefore included data from six[30,56–59,62] (out of 13) identified studies for which 2 × 2 data were available.

TABLE 4 Characteristics of studies included in the DTA 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[30]	Prospective cross-sectional; UK; NR; February 2014 to September 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 March 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): [a] 32.7 (28.1 to 38.1) T2D: [a] 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:** March 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): [a] 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 September 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): [a] 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

continued

TABLE 4 Characteristics of studies included in the DTA review (*continued*)

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
Kim 2013[57]	Retrospective cross-sectional; USA; tertiary care; January 2007 to September 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; October 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 60 Hz and 80 Hz vibration frequencies and liver biopsy within 2 months	Mean age ± SD (range): [a] 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; October 2013 to January 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: [a] 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; September 2018 to October 2020	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed (*N* = 37); recruited from patients with an incidental finding of hepatic steatosis on abdominal ultrasound	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

TABLE 4 Characteristics of studies included in the DTA review (*continued*)

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
Trout 2018[63]	Prospective cross-sectional; USA; NR; January 2012 to September 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 [a] (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; August 2011 to December 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 [a] (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; January 2019 to February 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:[b] 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

FLIP = Fatty Liver Inhibition of Progression; IQR = interquartile range; METAVIR = meta-analysis of histological data in viral hepatitis; mpMRI = multiparametric magnetic resonance imaging; NR = not reported; ROI = region of interest; SAF = steatosis, activity, fibrosis score; SD = standard deviation; T2D = type 2 diabetes; UCSD = University of California at San Diego.

a The statistics reported are based on the entire study population and not for the subpopulation.

b Does not specify type of diabetes.

TABLE 5 Data sources for 2 × 2 DTA data

Study	Data source for 2 × 2 data	Data provided for population in scope[a]
Eddowes 2018[30]	Perspectum Ltd submission[71b] included 2 × 2 data	Yes
Imajo 2021[56]	2 × 2 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 Ltd, 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 2 × 2 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 2 × 2 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]	2 × 2 data were provided in Figure S7, S10 and S14 from the Selvaraj systematic review[72]	No
Pavlides 2017[59]	2 × 2 data (*n* = 28) were provided in the Perspectum Ltd 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 2 × 2 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]	2 × 2 data were made available after personal communication with study authors (Marian Troelstra, Amsterdam University Medical Centers, 24 November 2022)	No

IPD = individual patient data.

a In line with the final scope[24] 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.

b In this EAG report, references to the Perspectum Ltd submission[71] are to the evidence submission received by the EAG from Perspectum Ltd in response to the EAG request for information.

Where studies reported 2 × 2 data (i.e. the number of TP, FP, TN and FN test results), data from individual studies were summarised in forest plots (*Figures 3–6*) 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 4* (*Tables 21* and *22*).

Individual study results: LiverMultiScan

For the LiverMultiScan PDFF and LiverMultiScan cT1 outputs (see *Interventions/index tests*), 2 × 2 data were available from three studies[30,56,59] as shown in *Figures 3–5*. 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 *Characteristics of the included studies* of this EAG report, the EAG considers that the Eddowes 2018[30] 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[30] 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

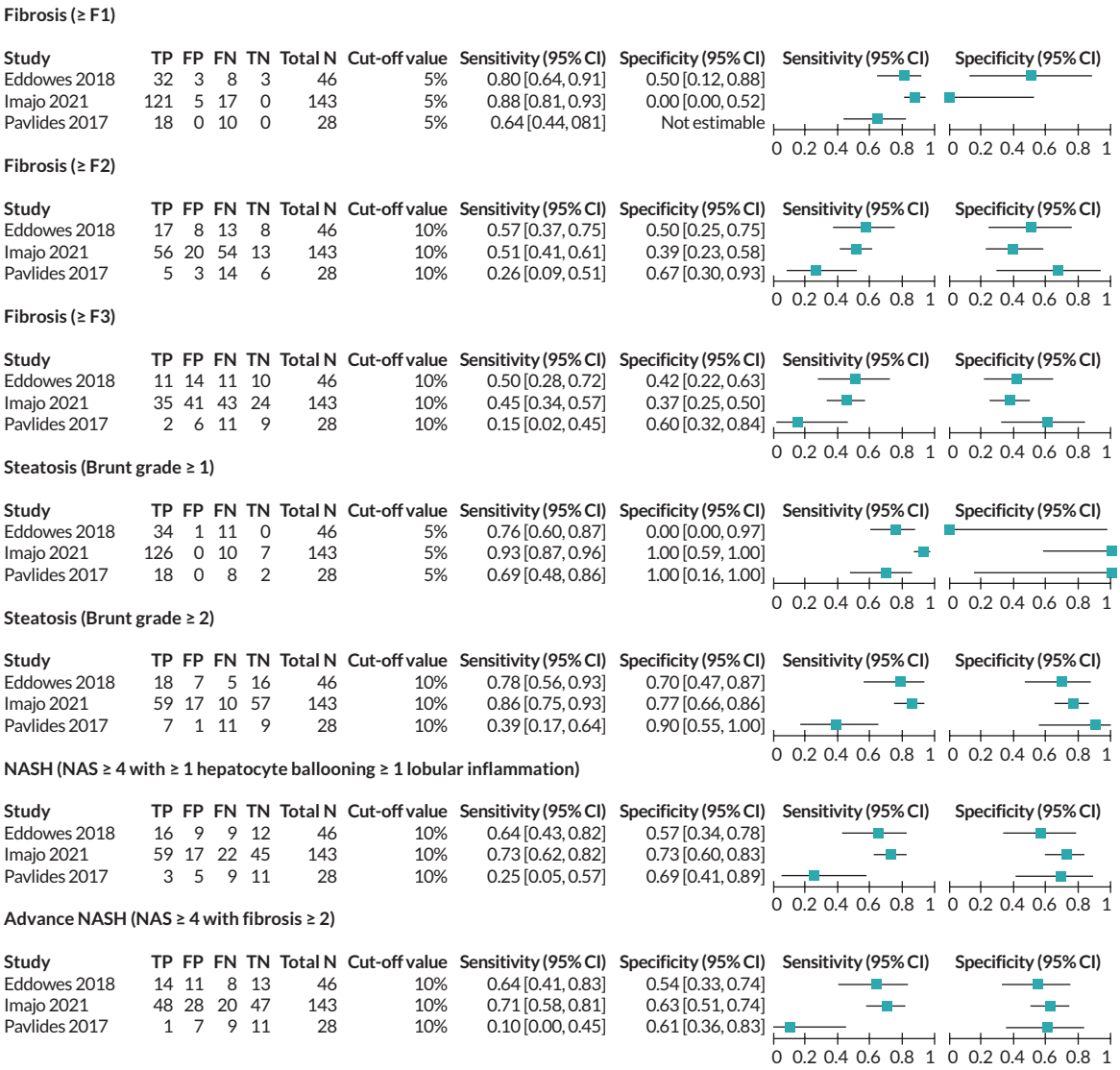

FIGURE 3 Forest plot displaying 2 × 2 data, sensitivity and specificity for LiverMultiScan PDFF from the included studies. Source: see Table 5.

(≥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[30] 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[30] study for both LiverMultiScan PDFF and LiverMultiScan cT1, that is, 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] that is, 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[30,56,59] that assessed the DTA of LiverMultiScan.

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]

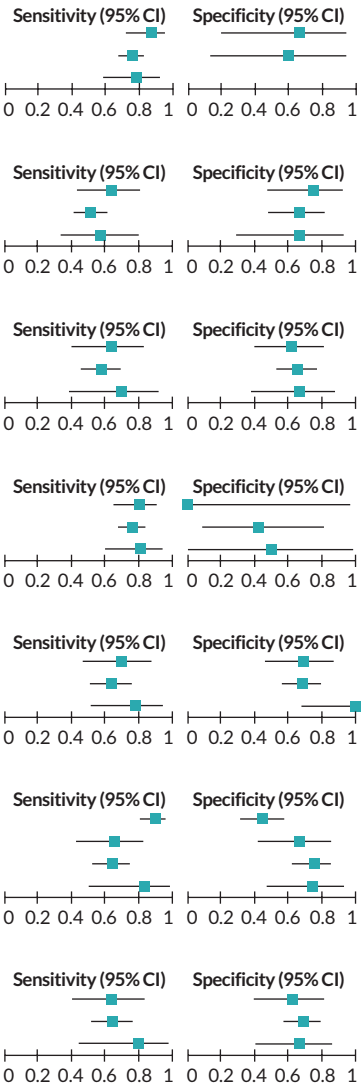

FIGURE 4 Forest plot displaying 2 × 2 data, sensitivity and specificity for LiverMultiScan cT1 from the included studies. *Data for NASH was available from the Imajo 2021[56] study for two cut-off values, 800 ms and 875 ms. All other studies reported data for the 875 ms cut-off value only. cT1 = iron corrected longitudinal relaxation time. Source: see Table 5.

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	56	9	25	53	143	800ms + 10%	0.69 [0.58, 0.79]	0.85[0.74, 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)
Imajo 2021	38	11	30	64	143	875ms + 10%	0.56 [0.43, 0.68]	0.85 [0.75, 0.92]

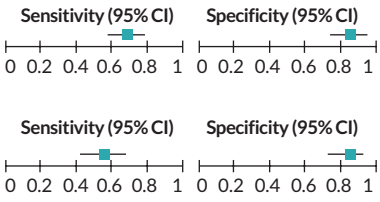

FIGURE 5 Forest plot displaying 2 × 2 data, sensitivity and specificity for LiverMultiScan PDFF and cT1 combined from the included studies. cT1 = iron corrected longitudinal relaxation time. Source: see Table 5.

For the diagnosis of NASH and advanced NASH, sensitivity was estimated to be 64% in the Eddowes 2018[30] 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%).

Individual study results: magnetic resonance elastography

For MRE, 2 × 2 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

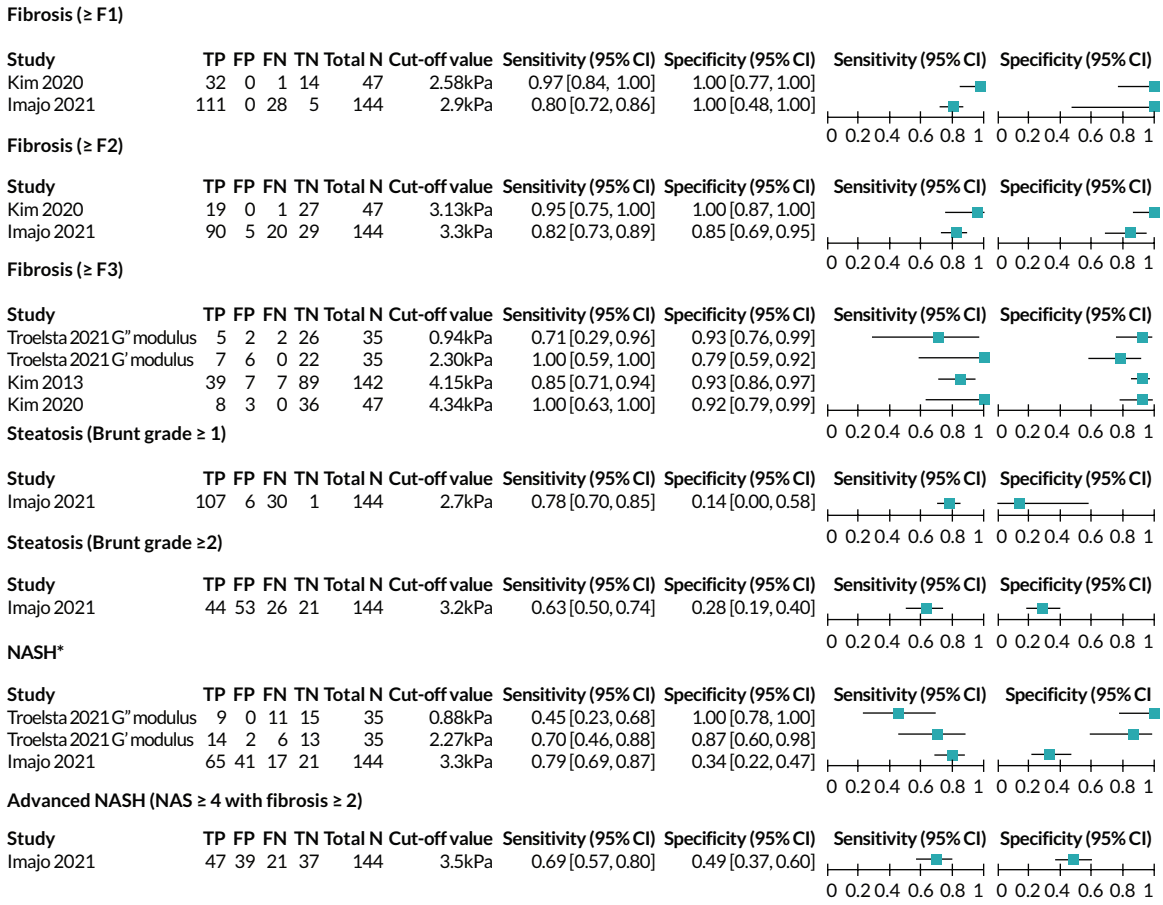

FIGURE 6 Forest plot displaying 2 × 2 data, sensitivity and specificity for MRE from the included studies. *NASH was defined in the Imajo 2021[56] study as NAS ≥4 with ≥1 hepatocyte ballooning and ≥1 lobular inflammation, and in the Troelstra 2021[62] study as ≥1 steatosis, ≥1 hepatocyte ballooning and ≥1 lobular inflammation. Source: see Table 5.

values used. The EAG notes that the Imajo 2021[56] study used the cut-off values that Resoundant, Inc. suggested to NICE[24] should be used to stage fibrosis (see *Magnetic resonance elastography*). 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.[24]

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 (2.30 kPa) was lower than the value that Resoundant, Inc. suggested to NICE[24] should be used to stage advanced fibrosis (>3.9 kPa) whereas the cut-off values used by the Kim 2013[57] study (4.15 kPa) and the Kim 2020[58] study (4.34 kPa) were greater. Sensitivity values were 100% for both the study which used the lowest cut-off value (Troelstra 2021,[62] cut-off value = 2.30 kPa) and the study that used the highest cut-off value (Kim 2020,[58] cut-off value = 4.34 kPa). 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 = 4.15 kPa; Kim 2020:[58] specificity = 92%, cut-off value = 4.34 kPa), but a lower specificity value (79%) was observed for the Troelstra 2021[62] study, which applied a lower cut-off value (2.30 kPa).

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 CI (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% vs. 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% vs. 69%), but less specific (34% vs. 49%).

Results from External Assessment Group 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[30,56,59] was very large (specificity was reported to be 0% for one study[30] 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 the 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%).

Results from External Assessment Group meta-analyses: magnetic resonance elastography

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 were 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

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) [a]	Specificity (%, 95% CI) [a]
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 one 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	800 ms	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	875 ms	3	217	54.1 (46.3 to 61.7)	69.0 (56.0 to 79.5)
Fibrosis	≥F3	875 ms	3	217	60.2 (50.9 to 68.8)	65.4 (55.8 to 73.9)
Steatosis	Brunt grade ≥1	800 ms	3	217	77.3 (71.1 to 82.5)	40.0 (15.8 to 70.3)
Steatosis	Brunt grade ≥2	875 ms	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	800 ms	1	143	Insufficient number of studies to perform meta-analysis	
NASH		875 ms	3	217	66.1 (57.1 to 74.1)	73.7 (64.2 to 81.5)
Advanced NASH	NAS ≥4 + fibrosis ≥2	875 ms	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	800 ms + 10%	1	143	Insufficient number of studies to perform meta-analysis	
Advanced NASH	NAS ≥4 + fibrosis ≥2	875 ms + 10%	1	143	Insufficient number of studies to perform meta-analysis	

cT1 = iron corrected longitudinal relaxation time.
a Where no meta-analysis was performed, justification is provided instead of estimates of sensitivity and specificity.
Source: Bivariate random-effects meta-analysis performed using 2 × 2 data reported in *Figure 3* and *Figure 4*.

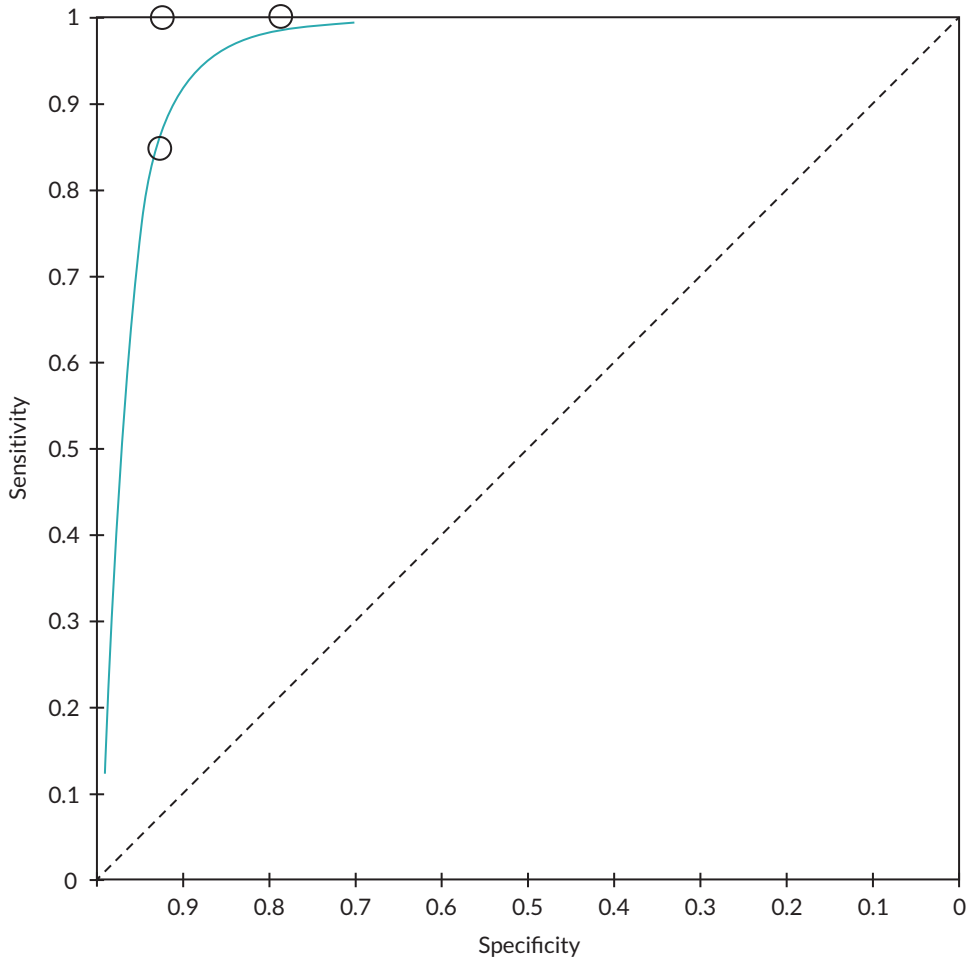

FIGURE 7 Summary ROC plot for fibrosis (≥F3) data from the MRE test. The solid line is the summary ROC curve. The dashed line indicates sensitivity = 1 − specificity (i.e. an uninformative test). The circles represent individual study results. The EAG notes that the Troelstra 2021[62] study used an investigational 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.

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.

Assessment of clinical impact

Eleven studies[30,53,54,57,59,62,64,66–69] reported in 14 publications[30,31,33,53,54,57,59,62,64–69] were included in the clinical impact review of MRI-based technologies. Five studies[30,59,66,68,69] reported in eight publications[30,31,33,59,65,66,68,69] 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.

Quality assessment

Seven[30,53,54,57,59,62,64] of the 13[30,53–64] DTA studies were also included in the clinical impact review. The EAG reassessed the methodological quality of the seven DTA studies[30,53,54,57,59,62,64] 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 was an RCT 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[30,53,54,57,59,62,64] and the two cohort studies,[66,67] the full assessment and summary of the risk of bias assessment for the included RCT[68,74] and the full assessment using the CASP qualitative-studies checklist[52] for the included qualitative study[69] are presented in *Supplementary material 2*.

Cross-sectional studies included in the diagnostic test accuracy review (*n* = 7)

Five studies[30,53,59,62,64] 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[30,53,54,57,59,62,64] 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[30,53,54,57,59,62,64] 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 30 December 2020) and a Clinical Study Report (CSR)[68] provided by Perspectum Ltd, 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 55 of the 802 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 were 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 undergone 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.

Characteristics of the included studies

Only one study[30] provided clinical impact results for a population of patients with NAFLD who had indeterminate or discordant results from fibrosis testing. Seven studies[30,53,54,57,59,62,64] 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[30,53,54,57,59,62,64] are presented in *Table 4*. In addition to these seven studies,[30,53,54,57,59,62,64] 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 USA, 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*.

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 1 ms increase in cT1. When a predefined cut-off of cT1 > 825 ms[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 >825 ms.

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 and 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 1 kPa 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 fewer patients with non-NAFLD and NAFLD underwent unnecessary biopsies in the LiverMultiScan

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 four countries (Germany, Netherlands, Portugal and UK) • 5 year study (1 year study setup; 3 year recruitment phase; 12 months follow-up)
Patient population	• Patients (18–75 years old) with suspected NAFLD • Dosage of eculizumab stable for ≥3 months prior to screening • Within SoC: ◦ 1.5 × ULN ≤ ALT and AST ≤5 × ULN and GGT ≥ 1.5 × ULN 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 (≥130/85 mmHg) ◦ elevated triglycerides (≥1.7 mmol/l) ◦ low HDL-cholesterol (<1.05 mmol/l for men or <1.25 mmol/l for women)
Intervention	• Patients (n = 403) were treated according to LiverMultiScan results. Further diagnostic evaluation was recommended when LiverMultiScan cT1 ≥ 800 ms or PDFF ≥ 10%. This was not a mandatory study requirement and was left at the discretion of the clinician and patient
Comparator	• SoC (n = 399)
Primary outcome	• Proportion of patients with suspected NAFLD incurring 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 Ltd 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; cT1 = iron-corrected longitudinal relaxation time; GGT = gamma-glutamyl transferase; 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/m²); 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):[a] 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):[a] 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; September 2017 to December 2020	Patients with NAFLD for whom advanced fibrosis or cirrhosis has not been diagnosed[b] (*n* = 55/802); recruited patients with suspected or known fatty liver disease. Patients recruited from seven UK sites (*n* = 253)	Median age:[c] 55 Male: 453 (56) Median BMI: 31 T2D: 334 (42)	NR	NR
MRE					
Gidener 2022[67]	Retrospective cohort study; USA; NR; retrospective 10 year follow-up of patients who underwent MRE; January 2007 to December 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):[a] 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[d]

IQR = interquartile range; NA = not applicable; NR = not reported; T2D = type 2 diabetes.

a 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.

b Only 55/802 patients had complete histology scores from liver biopsy to confirm stage of fibrosis.

c The statistics reported are based on the entire study population and not for the subpopulation of patients with complete histology scores from liver biopsy.

d 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.

arm (*n* = 9/22, 41%) compared to the SoC arm [*n* = 16/31, 52%, EAG calculated odds ratio (OR) = 0.65, 95% CI 0.22 to 1.96]. The RADIcAL trial[68] also reported that fewer patients with no to mild fibrosis (F0 to F1) in the LiverMultiScan arm underwent unnecessary biopsies (*n* = 9/22, 41%) compared to the SoC arm (*n* = 13/24, 54%, EAG calculated OR = 0.59, 95% CI 0.18 to 1.89).

A similar proportion of patients with non-NAFLD and NAFLD underwent unnecessary biopsies with elastography (22/48, 46%) and without elastography (3/6, 50%) prior to biopsy. A similar proportion of patients with no to mild fibrosis (F0 to F1) underwent unnecessary biopsies with elastography (20/41, 43%) and without elastography (2/6, 33%) prior to biopsy.

The RADIcAL trial[68] reported correlations between patients' histology scores and LiverMultiScan cT1 outputs (*Appendix 5*, *Figure 11*) and between patients' histology scores and LiverMultiScan PDFF outputs (*Appendix 5*, *Figure 12*).

Test failure rate

Three studies[30,59,66] 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%.[30] 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).[30]

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 6* (*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, for example 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 (see *Additional searches (clinical impact review)*), 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[24] 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[24] 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.

Summary of External Assessment Group diagnostic test accuracy and clinical impact review, and External Assessment Group quantitative analysis

External Assessment Group diagnostic test accuracy and clinical impact review

The EAG DTA review identified 13 studies[30,53–64] reported in 15 publications.[30,31,53–65] The EAG clinical impact review identified 11 studies[30,53,54,57,59,62,64,66–69] reported in 14 publications.[30,31,33,53,54,57,59,62,64–69] However, the EAG was only confident that one study (the Eddowes 2018[30] study) was carried out in the population described in the final scope[24] 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, fewer patients with non-NAFLD, fewer with NAFLD and fewer patients with no-mild fibrosis (F0 to F1) 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.

External Assessment Group quantitative analysis

The only relevant study[30] (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). The Eddowes study[30] categorised patients according to low or high risk of progressive liver disease. However, it was also possible to interpret the DTA data[71] generated by LiverMultiScan as follows: 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.[30,56–59,62] 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[30] 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.

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 had 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.

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.

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 1*. To identify published economic evaluations, the EAG appended an economic evaluation-specific search filter to the clinical search strategies (*Appendix 7*). In addition, two databases of economic publications [EconLit (EBSCO) and the CEA registry] were searched, using the search strategies presented in *Appendix 7*, from inception until 4 October 2021. The results of the searches were entered into an Endnote X9 library and de-duplicated (MM) before being exported into Covidence.

Study selection and inclusion criteria

The review inclusion and exclusion criteria (*Table 9*) reflected the decision problem outlined in the final scope[24] 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.

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.

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

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. CEA, 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 (e.g. reviews, editorials and opinion papers)
Language	English only	Non-English-language studies

LY = life year; QALY = quality-adjusted life year.

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.

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[30] was identified as being relevant. The PRISMA flow diagram[48] provides an illustration of the screening and selection process (*Figure 8*). A list of the studies excluded at the full-text stage, along with reasons for exclusion, is provided in *Supplementary material 1*.

Quality of the included evidence

The quality of the included study[30] was assessed using the Drummond checklist[76] (*Table 10*) and the CHEERS checklist[77] (see *Supplementary material 2*).

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[30] 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

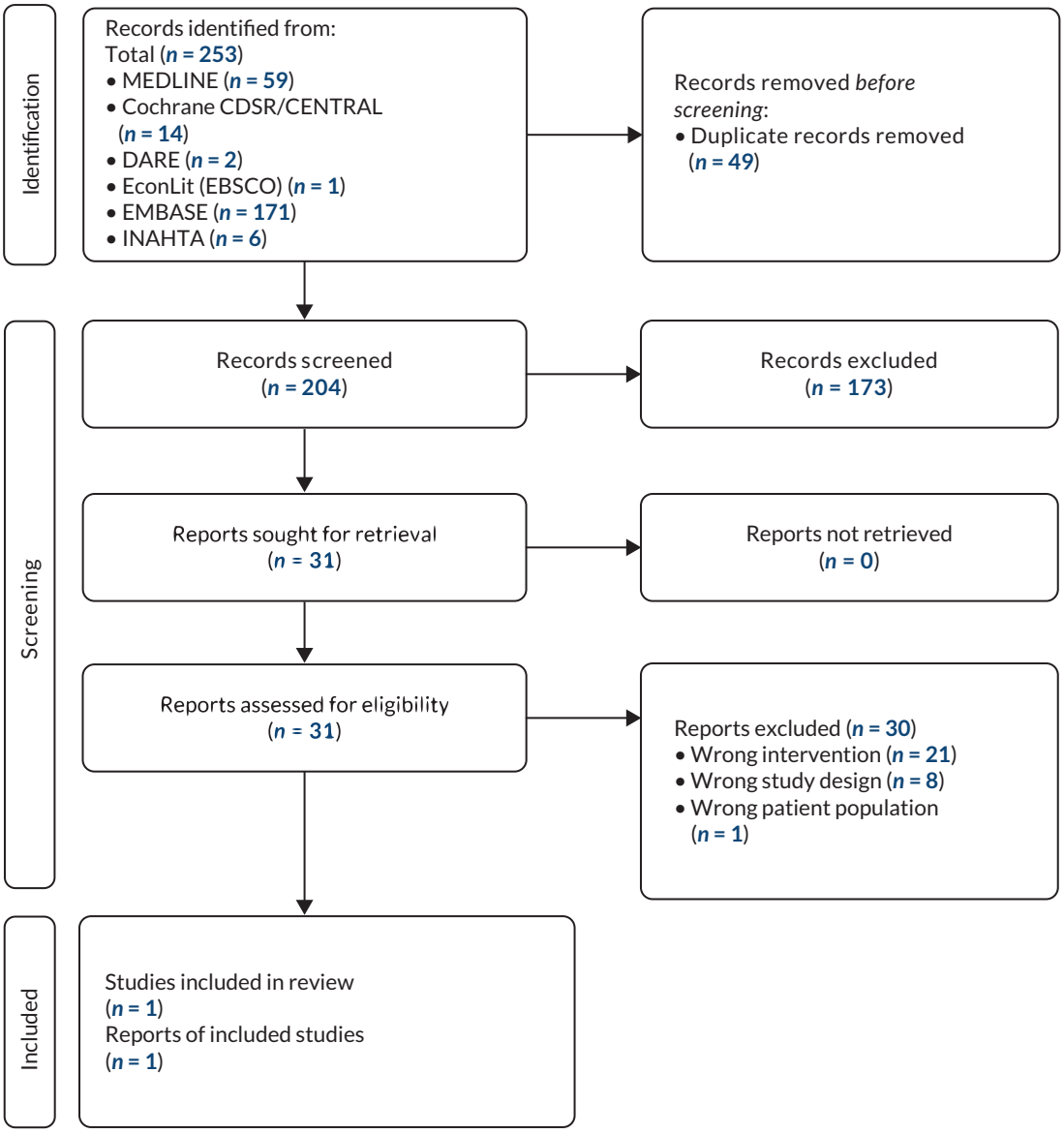

FIGURE 8 PRISMA flow diagram for the cost-effectiveness review.

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
Were 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).

inconclusive results from previous fibrosis testing and therefore the Eddowes 2018[30] cost-saving results are not relevant to this appraisal.

The information provided in the published paper[30] 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 CEA, nor the generalisability of results.

Characteristics of the included study

The characteristics of the included study[30] are summarised in *Table 4*. This study[30] was also included in the EAG DTA and clinical impact review.

The included study, Eddowes 2018,[30] 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: 822 ms and 875 ms), TE (two cut-offs: 5.8 kPa and 7.0 kPa), 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.

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 (875 ms) plus TE (7.0 kPa) 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).

Study conclusions

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

External Assessment Group cost-effectiveness review conclusions

The EAG searches for published economic evaluations that assessed the cost-effectiveness of LiverMultiScan and MRE only identified one study.[30] The included study[30] 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[30] study, clinical effectiveness evidence was collected from a population with inconclusive results from previous fibrosis testing. To generate cost-effectiveness results, Eddowes 2018[30] 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

TABLE 11 Cost-effectiveness results

Intervention	Biopsies avoided Per 1000 patients	Total costs	Cost savings vs. biopsy	Total costs per correct diagnosis
LMS cT1 822 ms[a]	381.9	£538,345	£101,265	£649.57
LMS cT1 875 ms[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	£1138.43
ELF 9.8[b]	858.9	£201,322	£438,288	£363.97
LMS cT1 822 ms+TE 5.8 kPa[a]	734.6	£338,260	£301,359	£415.37
LMS cT1 875 ms+TE 5.8 kPa	722.7	£345,851	£293,759	£414.60
LMS cT1 822 ms+TE 7.0 kPa	841.1	£242,309	£397,301	£315.60
LMS cT1 875 ms+TE 7.0 kPa[b]	848.7	£237,488	£402,122	£307.92

cT1 = iron-corrected longitudinal relaxation time; LMS = LiverMultiScan.
a Patients with simple steatosis and ≤F1 fibrosis.
b Patients with either NASH or >F1 fibrosis.
Source: Eddowes 2018[30] study.

results from previous fibrosis testing. Therefore, Eddowes 2018[30] study cost-effectiveness results are not relevant to this appraisal.

Development of a de novo model

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[24] 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: that is, 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[24] issued by NICE and are generalisable to patients treated in NHS clinical practice.

The EAG only identified one study (Eddowes 2018[30]) 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[30] study population represents a subset of the population described in the final scope[24] 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[24] issued by NICE.

Further, LiverMultiScan data are not available for patients for whom TE or ARFI was unsuitable. In addition, no DTA or population prevalence data were available for any of the population described in the final scope[24] issued by NICE for patients who had had an MRE.

The EAG cautions that the data presented in the Eddowes 2018[30] 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, that is, as described in the Eddowes 2018[30] publication, and are referred to as Eddowes 2018/Perspectum Ltd.[30,71]

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.

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[30] chose to categorise patients according to low or 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: that is, 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 875 ms; however, the EAG has assumed that patients with cT1 scores less than 800 ms 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*.

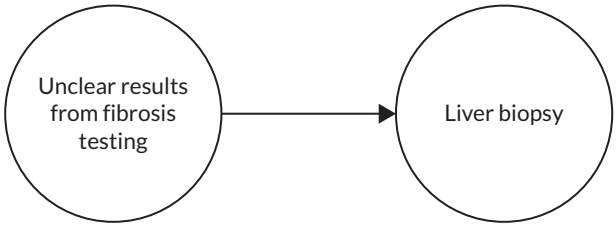

FIGURE 9 Current NHS diagnostic test pathway for patients with inconclusive results from previous fibrosis testing.

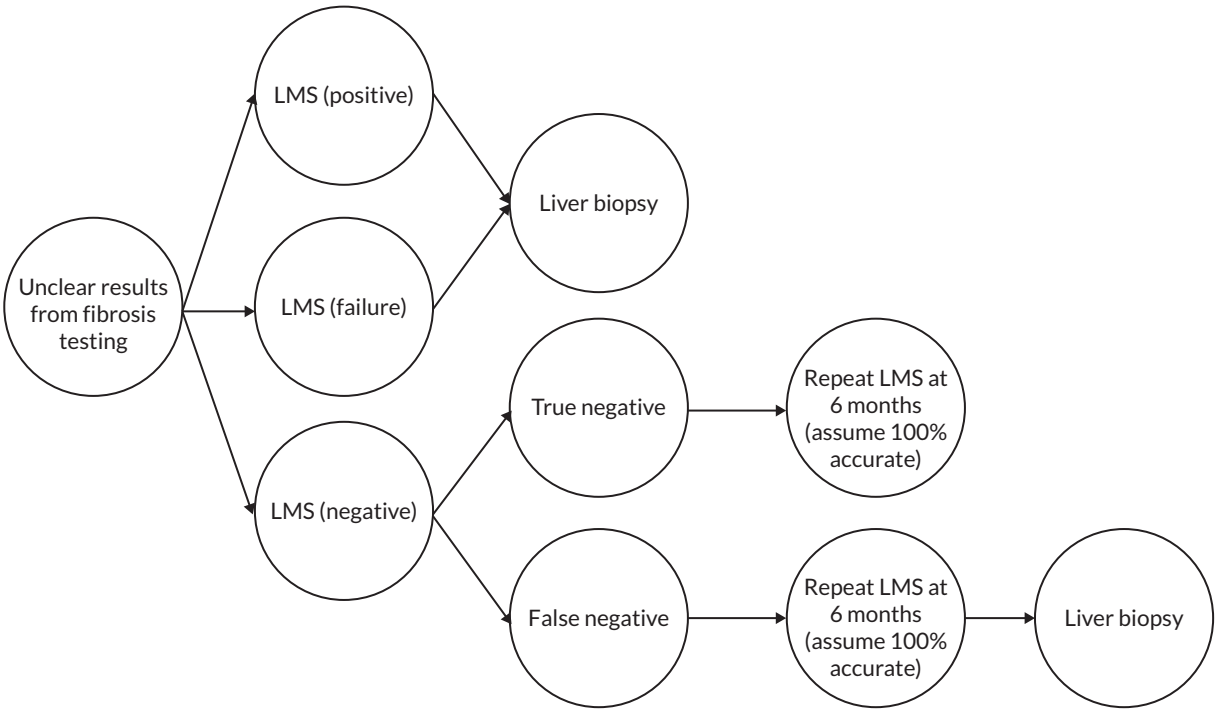

FIGURE 10 Proposed LiverMultiScan plus biopsy diagnostic test pathway for patients with inconclusive tests from fibrosis testing. LMS, LiverMultiScan

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[30] 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.

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.[30] No MRE DTA data and population prevalence data were available for the population described in the final scope[24] 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 800 ms: 'Fatty liver'
  - Reassure as no inflammation present
  - Reassess with MRI in 3 years

- 800–875 ms: 'Non-alcoholic steatohepatitis (NASH)'
  - Lifestyle modification
  - Management of type 2 diabetes and cardiovascular disease
  - Monitor disease status with MRI after 6 months

- More than 875 ms: 'High-risk NASH'
  - Reassess with MRI every 6 months
  - Consider liver biopsy if cirrhosis is suspected
  - Cancer surveillance
  - Consider inclusion in NASH therapeutic trials.

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 CEA is, therefore, populated with LiverMultiScan cT1 scores. For completeness, the cost-effectiveness results generated using PDFF values are presented in *Appendix 8* (*Tables 23–27*).

Comparator

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

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.

External Assessment Group 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.[30,71] 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[30] was 5.5%. In the EAG model, any patient who had a test failure result was referred for a biopsy.

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, whichever test comes first in the pathway, are not included in the EAG analysis. Intervention costs are displayed in *Table 13*.

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).

TABLE 12 LiverMultiScan DTA strategies and values (per 1000 successful tests)

Diagnostic test strategy		cT1 cut-off value (ms)	Population prevalence (%)	True positive	True negative	False positive	False negative	Sensitivity	Specificity
T1	Any fibrosis (≥F1)	800	87.0	761	87	43	109	0.88	0.67
T2	Significant fibrosis (≥F2)	875	65.2	413	261	87	239	0.63	0.75
T3	Advanced fibrosis (≥F3)	875	47.8	304	326	196	174	0.64	0.63
T4	Brunt grade ≥1	800	97.8	782	0	22	196	0.8	0
T5	Brunt grade ≥2	875	50.0	348	348	152	152	0.7	0.7
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	54.4	348	304	152	196	0.64	0.67
T7	Advanced NASH (NAS ≥4 plus ≥F2)	875	47.8	304	326	196	174	0.64	0.62
T8[a]	High risk (NASH or >F1)	875	82.6	478	152	22	348	0.58	0.88

cT1 = iron-corrected longitudinal relaxation time; F = fibrosis stage.
a Only sensitivity and specificity values were available from the Eddowes 2018[30] study; the other values were calculated by the EAG.
Source: Eddowes 2018 study/Perspectum Ltd.[30,71]

TABLE 13 Intervention costs

Intervention	Cost	Description	Source
Biopsy	£1513	YG10Z Percutaneous transvascular[a] biopsy of lesion of liver	NHS Reference Costs 2019/20[79]
	£770	YG11A Percutaneous punch[b] 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[30,71]

a Transjugular.
b Standard biopsy procedure.

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).

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.

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.

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[30] 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.

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.

Summary of base-case assumptions

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

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.

TABLE 14 External Assessment Group 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[30,71]
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)	82.6%	
LiverMultiScan test accuracy		
Sensitivity		
Any fibrosis (≥F1)	0.88	Eddowes 2018/Perspectum Ltd[30,71]
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)	0.58	
Specificity		
Any fibrosis (≥F1)	0.67	Eddowes 2018/Perspectum Ltd[30,71]
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)	0.88	

TABLE 14 External Assessment Group base-case model assumptions (*continued*)

Parameter	Assumption	Source/justification
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; ICER = incremental cost-effectiveness ratio; NA = not applicable; NG = NICE guideline.

External Assessment Group base-case cost-effectiveness analysis results

The EAG has generated base-case analysis cost-effectiveness results for a hypothetical cohort of 1000 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 CEA 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 in the pathway increases costs per patient; range: £244 (Brunt grade ≥2) to £412 (Brunt grade ≥1).

For seven 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), advanced NASH (NAS ≥4 plus ≥F2), and high risk (NASH or >F1)], QALY losses were greater for the

LiverMultiScan plus biopsy pathway than for the biopsy only pathway. For the remaining diagnostic test strategy (Brunt grade ≥2), the QALY loss was greater for the biopsy only pathway.

For seven 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), advanced NASH (NAS ≥4 plus ≥F2) and high risk (NASH or >F1)], the base-case ICERs per QALY gained show that the LiverMultiScan plus biopsy pathway is dominated by the biopsy only pathway, that is, the biopsy only pathway is less expensive and leads to fewer QALY losses than the LiverMultiScan plus biopsy pathway.

The most cost-effective diagnostic test strategy is Brunt grade ≥2. The incremental cost-effectiveness ratio (ICER), for this strategy, for the comparison of the LiverMultiScan plus biopsy pathway versus the biopsy only pathway, is £1,266,511 per QALY gained. Clinicians suggested that, when considering this strategy, 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 *Tables 15–19*.

TABLE 15 Initial LiverMultiScan outcomes generated by the EAG model (per 1000 tests)

Diagnostic test strategy	cT1 cut-off value (ms)	True positive	True negative	False positive	False negative	Failed tests
T1: Any fibrosis (≥F1)	800	719.1	82.2	40.6	103.0	55.0
T2: Significant fibrosis (≥F2)	875	390.3	246.6	82.2	225.9	55.0
T3: Advanced fibrosis (≥F3)	875	287.6	308.2	184.9	164.3	55.0
T4: Brunt grade ≥1	800	739.9	0.0	20.8	185.2	55.0
T5: Brunt grade ≥2	875	328.9	328.9	143.6	143.6	55.0
T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	328.9	287.3	143.6	185.2	55.0
T7: Advanced NASH (NAS ≥4 plus ≥F2)	875	287.3	308.1	185.2	164.4	55.0
T8: High Risk (NASH or >F1)	875	452.0	143.8	20.5	328.7	55.0

cT1 = iron-corrected longitudinal relaxation time; F = stage of fibrosis.

TABLE 16 LiverMultiScan plus biopsy pathway: biopsies performed and averted (per 1000 patients)

Diagnostic test strategy	cT1 cut-off value (ms)	Total number of biopsies, including those following a repeated LiverMultiScan at 6 months	Biopsies averted
T1: Any fibrosis (≥F1)	800	917.8	82.2
T2: Significant fibrosis (≥F2)	875	753.4	246.6
T3: Advanced fibrosis (≥F3)	875	691.8	308.2
T4: Brunt grade ≥1	800	1000.0	0.0
T5: Brunt grade ≥2	875	671.1	328.9
T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	712.7	287.3
T7: Advanced NASH (NAS ≥4 plus ≥F2)	875	691.9	308.1
T8: High Risk (NASH or >F1)	875	898.9	143.8

cT1 = iron-corrected longitudinal relaxation time; F = stage of fibrosis.

TABLE 17 Pathway diagnostic test strategy costs (per 1000 patients)

Diagnostic test strategy	cT1 cut-off value (ms)	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)	800	£738,817	£7838	£411,556	£1,158,211	£805,000	£8540	£813,540	£344,671
T2: Significant fibrosis (≥F2)	875	£606,451	£6434	£511,311	£1,124,195	£805,000	£8540	£813,540	£310,655
T3: Advanced fibrosis (≥F3)	875	£556,938	£5908	£511,311	£1,074,157	£805,000	£8540	£813,540	£260,617
T4: Brunt grade ≥1	800	£805,000	£8540	£411,556	£1,225,096	£805,000	£8540	£813,540	£411,556
T5: Brunt grade ≥2	875	£540,268	£5732	£511,311	£1,057,310	£805,000	£8540	£813,540	£243,770
T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	£573,740	£6087	£511,311	£1,091,137	£805,000	£8540	£813,540	£277,597
T7: Advanced NASH (NAS ≥4 plus ≥F2)	875	£557,004	£5909	£511,311	£1,074,224	£805,000	£8540	£813,540	£260,684
T8: High risk (NASH or >F1)	875	£689,238	£7312	£511,311	£1,207,860	£805,000	£8540	£813,540	£394,320

cT1 = iron-corrected longitudinal relaxation time; F = stage of fibrosis; LMS = LiverMultiScan.

TABLE 18 Quality-adjusted life year analyses for the two diagnostic pathways (per 1000 patients)

		LiverMultiScan plus biopsy pathway					Biopsy only pathway				Incremental QALYs (LMS+biopsy pathway)[a]
Diagnostic test strategy	cT1 cut-off value (ms)	Biopsy procedure	Biopsy complications	Biopsy death	False negatives	Total QALY losses	Biopsy procedure	Biopsy complications	Biopsy death	Total QALY losses	
T1: Any fibrosis (≥F1)	800	5.12	0.13	1.29	1.55	8.10	5.58	0.15	1.41	7.14	−0.96
T2: Significant fibrosis (≥F2)	875	4.20	0.11	1.06	3.39	8.76	5.58	0.15	1.41	7.14	−1.63
T3: Advanced fibrosis (≥F3)	875	3.86	0.10	0.98	2.47	7.40	5.58	0.15	1.41	7.14	−0.27
T4: Brunt grade ≥1	800	5.58	0.15	1.41	2.78	9.92	5.58	0.15	1.41	7.14	−2.78
T5: Brunt grade ≥2	875	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)	875	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)	875	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)	875	4.78	0.13	1.21	4.93	11.04	5.58	0.15	1.41	7.14	−3.90

cT1 = iron corrected longitudinal relaxation time; F = stage of fibrosis; LMS = LiverMultiScan.

a 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.

TABLE 19 Incremental analyses for LiverMultiScan plus biopsy vs. biopsy (1000 patients)

Diagnostic test strategy	cT1 cut-off value (ms)	Incremental		ICER per QALY gained (vs. biopsy)
		Costs	QALYs	
T1: Any fibrosis (≥F1)	800	£344,671	−0.96	LMS+biopsy dominated by biopsy
T2: Significant fibrosis (≥F2)	875	£310,655	−1.63	LMS+biopsy dominated by biopsy
T3: Advanced fibrosis (≥F3)	875	£260,617	−0.27	LMS+biopsy dominated by biopsy
T4: Brunt grade ≥1	800	£411,556	−2.78	LMS+biopsy dominated by biopsy
T5: Brunt grade ≥2	875	£243,770	0.19	£1,266,511
T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	875	£277,597	−0.73	LMS+biopsy dominated by biopsy
T7: Advanced NASH (NAS ≥4 plus ≥F2)	875	£260,684	−0.27	LMS+biopsy dominated by biopsy
T8: High risk (NASH or >F1)	875	£394,320	−3.90	LMS+biopsy dominated by biopsy

cT1 = iron-corrected longitudinal relaxation time; F = stage of fibrosis. LMS, LiverMultiScan

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[30] 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 Brunt grade ≥2. 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 Brunt grade ≥2 test strategy (50.0%) was lower than the threshold values required for this strategy to be considered cost-effective at thresholds of £20,000 (9.1%) or £30,000 (14.8%) per QALY gained; the accuracy of the LiverMultiScan test for this strategy was not close to 100% (sensitivity = 0.70; specificity = 0.70)

Quality-adjusted life year losses associated with each biopsy

The values that QALY losses associated with a biopsy would need to be for the most cost-effective diagnostic test strategy, 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*.

Cost analysis

The EAG threshold cost analysis focused on Brunt grade 2, which was the most cost-effective diagnostic test strategy (£1,266,511 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 £184.31 (£185.61) per patient, then the ICER per QALY gained for this comparison would fall to £20,000 (£30,000).

TABLE 20 Quality-adjusted life year loss associated with biopsy: results from threshold analyses

	Threshold: £20,000 per QALY			Threshold: £30,000 per QALY		
Diagnostic test strategy	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%

External Assessment Group 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), advanced NASH (NAS ≥4 plus ≥F2) and high risk (NASH or >F1) patients. Brunt grade ≥2 remained the 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), advanced NASH (NAS ≥4 plus ≥F2) and high-risk (NASH or >F1) patients. Brunt grade ≥2 remained the most cost-effective diagnostic strategy, with the ICER increasing from £1,266,511 to £1,356,715 per QALY gained.

Removal of quality-adjusted life year 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.

Cost-effectiveness of magnetic resonance elastography plus biopsy versus biopsy only

The EAG carried out cost-effectiveness analyses using published MRE sensitivity and specificity estimates. To undertake this analysis, the EAG used the MRE 2 × 2 data provided by Perspectum Ltd (14 December 2021)[71] from the trial reported in the Imajo 2021[56] publication. In this trial, LiverMultiScan cT1 scores and MRE were used to diagnose NASH in Japanese patients with a diagnosis or suspicion of NAFLD who were also suspected to have NASH; however, the data used in these analyses were not derived from the population described in the final scope[24] issued by NICE and therefore results can only be considered as illustrative. Results from these analyses are presented in *Appendix 9* (*Tables 29–35*) for information only.

External Assessment Group analyses of uncertainty considered and rejected

Probabilistic sensitivity analysis

The EAG model is linear (single-node decision tree). The EAG confirmed model non-linearity by increasing and decreasing parameters by 20%, averaging the ICERs per QALY gained from these analyses and comparing them with the deterministic base-case ICERs per QALY gained. The results showed that, depending on the test strategy, the difference between the ICERs per QALY gained generated from averaging results from the ±20% analyses and the deterministic ICERs per QALY gained was between 0.01% and 0.02%. Therefore, using probabilistic sensitivity analysis (PSA) to explore the impact of model

non-linearity on cost-effectiveness results is not required. Further, due to the uncertainty around the validity of point estimates, especially the covariance between sensitivity and specificity, and as the distributions around most of the model inputs are unknown, any PSA would largely be populated with arbitrary data, 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.

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[30] 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[24] 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, the 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 is more suitable than the value from the Eddowes 2018[30] study used in the EAG model as they do not specifically relate to the patients described in the final scope[24] 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[30] study data; other studies carried out in the same population may lead to substantially different population prevalence estimates.

Alternative sources of diagnostic test accuracy 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[24] 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. However, the EAG has carried out cost-effectiveness analyses using published MRE sensitivity and specificity estimates. To undertake this analysis, the EAG has used the MRE 2 × 2 data provided by Perspectum Ltd (14 December 2021)[71] from the trial reported in the Imajo 2021[56] publication; however, the ERG reiterates that the data used in these analyses were not derived from the population described in the final scope[24] issued by NICE. Results from these analyses are presented in *Appendix 9* (*Tables 29–35*).

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.

External Assessment Group 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[30] 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 seven 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[30] 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.[30,71] 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).[30] This is of concern as, in a different population to that described in the final scope[24] 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[24] population).

Discussion

Statement of principal findings

Diagnostic test accuracy

In line with the final scope[24] issued by NICE, the 13 studies[30,53–64] 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[30,53–64] that were included in the DTA review, the EAG was confident that only one study[30] 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[30] 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.

Clinical impact

Eleven studies[30,53,54,57,59,62,64,66–69] 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[30] 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 reduce the number of unnecessary biopsies for patients with non-NAFLD, NAFLD and no to mild fibrosis (F0 to F1) when compared to standard care.

Test failure rate in a population of patients with NAFLD was reported in four studies.[30,53,57,62] The test failure rate of the index tests for patients with NAFLD was 5.6%[30] 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[24] issued by NICE (see *Table 2*).

Cost-effectiveness

Eddowes 2018[30] 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[30] 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 seven diagnostic pathways showed that LiverMultiScan plus biopsy was dominated by biopsy only and for Brunt grade ≥2 the ICER per QALY gained was £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.

The EAG also carried out MRE analyses using sensitivity and specificity data from a population that differed from the population described in the final scope[24] issued by NICE and, therefore, results should only be considered as illustrative.

Strengths and limitations of the assessment

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.

Limitations of the assessment

The DTA and population prevalence data available from Eddowes 2018/Perspectum Ltd[30,71] 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[24] 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.[30,56–59,62] Furthermore, the meta-analyses were populated with data from small numbers of studies and only one[30] 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 were 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[24] issued by NICE.

Eddowes 2018/Perspectum Ltd[30,71] 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, that is, MRE is unlikely to be cost-effective in the population described in the final scope[24] issued by NICE (using data from Eddowes 2018/Perspectum Ltd[30,71]) even if test accuracy was 100%.

In the EAG model, LiverMultiScan is positioned as a triage test, that is, 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[24] issued by NICE are 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.

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 is 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] that 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[30] 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.

Reporting equality, diversity and inclusion

The EAG elicited the views of the Diagnostic Assessment Specialist Committee members during the review process. The EAG took into account the views of the Committee (which was made up of professional and lay members) when developing the EAG cost-effectiveness model. In addition, the EAG considered all the comments submitted by British Association for the Study of the Liver as part of the consultation process.

Conclusions

Clinical effectiveness

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 a population that may benefit from implementation of this technology, namely patients with indeterminate or discordant results from previous fibrosis testing or patients for whom TE and ARFI are not suitable.

Cost-effectiveness

Only one small LiverMultiScan study[29] provided DTA and population prevalence data for patients described in the final scope[24] 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 was unsuitable. Further, no MRE DTA data were available for the population described in the final scope[24] issued by NICE. The

EAG considers that even if MRE was 100% accurate, due to high population prevalence estimates it is unlikely that MRE would be cost-effective at current prices.

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.

Suggested research priorities

Only Eddowes 2018/Perspectum Ltd[30,71] 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.

Qualitative studies are required to investigate the impact of non-invasive 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.

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 DTA and Chris Hyde (Professor of Public Health and Clinical Epidemiology, University of Exeter) for comments on a draft version of the EAG 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.

Data-sharing statement

Data in this review were extracted from publications freely available in the public domain for Kim 2013[57] and Kim 2020.[58] Further details about data sources are provided in *Table 5*. Data for the three LiverMultiScan studies[30,56,59] were provided upon request from Perspectum Ltd and are not publicly available. Data from the Troelstra 2021[62] study were provided by the study authors in response to the EAG request and are not publicly available. Requests for access to the data should be addressed to Perspectum Ltd and to the corresponding author of the Troelsta 2021[62] study respectively. If you have any queries, please contact the corresponding author.

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 (https://orcid.org/0000-0001-5578-1535) (Deputy Director, LRiG, 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 (https://orcid.org/0000-0001-5578-1535) (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 (https://orcid.org/0000-0002-2187-1003) (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 (https://orcid.org/0000-0003-0164-103X) (Director, Hare Research, Health Economics and Modelling) provided input to the health economic model and provided senior advice to the project.

Angela Boland (https://orcid.org/0000-0002-7097-8704) (Director, LRiG, Health Economics and Modelling) provided input to the health economic model and provided senior advice to the project.

Marty Chaplin (https://orcid.org/0000-0003-1694-8106) (Research Associate, Statistician) contributed to the statistical analysis methods and performed the statistical analysis for the systematic review of diagnostic test accuracy and clinical impact.

Devarshi Bhattacharyya (https://orcid.org/0000-0002-4315-7732) (Health Economic Modeller, Health Economics and Modelling) conducted the review of cost-effectiveness evidence.

Rachel Houten (https://orcid.org/0000-0002-1092-0092) (Health Economic Modeller, Health Economics and Modelling) contributed to the review of cost-effectiveness evidence.

Katherine Edwards (https://orcid.org/0000-0001-9988-2709) (Senior Research Fellow, Systematic Reviewer) acted as the second reviewer in the systematic review.

Sarah Nevitt (https://orcid.org/0000-0003-4419-6343) (Research Associate, Statistician) contributed to the statistical analysis for the diagnostic test accuracy review.

Michelle Maden (https://orcid.org/0000-0002-5435-8644) (Research Associate, Information Specialist) devised and performed the literature searches.

References

1. National Health Service. *Non-alcoholic fatty liver disease (NAFLD)*. Published 13 January 2022. URL: 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. Thomaides-Brears HB, Alkhouri N, Allende D, Harisinghani M, Noureddin M, Reau NS, *et al.* Incidence of complications from percutaneous biopsy in chronic liver disease: a systematic review and meta-analysis. *Dig Dis Sci* 2021;**67**(7).
6. British Liver Trust. *The alarming impact of liver disease in the UK. Facts and statistics*. Published June 2019. URL: 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. URL: 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. URL: 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. 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.
19. 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.
20. de Alwis NMW, Day CP. Non-alcoholic fatty liver disease: the mist gradually clears. *J Hepatol* 2008;**48**:S104–S12.
21. 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.
22. Horn P, Newsome PN. *NAFLD – diagnosis, assessment and management*. Published 17 November 2021. URL: www.bsg.org.uk/clinical-articles-list/nafld-diagnosis-assessment-and-management/ (accessed 8 December 2021).
23. Byrne CD, Patel J, Scorletti E, Targher G. Tests for diagnosing and monitoring non-alcoholic fatty liver disease in adults. *BMJ* 2018;**362**:k2734.
24. National Institute for Health and Care Excellence. *MRI-based technologies for the assessment of non-alcoholic fatty liver disease: final scope*. Published September 2021. URL: www.nice.org.uk/guidance/gid-dg10045/documents/final-scope (accessed 22 November 2021).
25. 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.
26. 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.
27. Smeuninx B, Boslem E, Febbraio MA. Current and future treatments in the fight against non-alcoholic fatty liver disease. *Cancers* 2020;**12**:1714.
28. National Institute for Health and Care Excellence. *Obesity: identification, assessment and management*. Published 6 July 2016. URL: www.nice.org.uk/guidance/cg189 (accessed 14 December 2021).
29. 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.
30. 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.
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. URL: 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. URL: 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. URL: 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. URL: 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. URL: 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. URL: 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. URL: 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. URL: 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. URL: 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;**54**(6).
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 LiverMultiScanTM: 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 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. URL: 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**;PWE-038.

Appendix 1 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 2 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 3 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 homogeneous 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 CIs where appropriate), and pooled these data in 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 4 Area under the receiver operating characteristic curve results reported in the included studies

TABLE 21 Area under the receiver operating characteristic curve 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[30]	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[30]	50	0.63 (0.45 to 0.81)
	≥F3	Eddowes 2018[30]	50	0.62 (0.46 to 0.78)
Steatosis	Simple steatosis with no significant fibrosis[a]	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[30]	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[30]	50	0.74 (0.59 to 0.88)
Risk of progressive disease	High risk (NASH or >F1) vs. low risk (simple steatosis and ≤F1)	Eddowes 2018[30]	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)

cT1 = iron-corrected longitudinal relaxation time; F = fibrosis stage; NR = not reported.
a No further definition given. ≤F1 was assumed as no significant fibrosis because significant fibrosis was defined as >F1.

TABLE 22 Area under the receiver operating characteristic curve 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 <35 kg/m²**: 0.89 (0.82 to 0.96) **Patients with BMI ≥35 kg/m²**: 0.93 (0.84 to 1.00)
		Caussy 2018[53]: Mayo clinic cohort	75	**Patients with BMI <40 kg/m²**: 0.97 (0.93 to 1.00) **Patients with BMI ≥40 kg/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)

F = fibrosis stage; G' = shear modulus; G' = loss modulus; NR = not reported; UCSD = University of California, San Diego.

Appendix 5 Correlations between individual histology scores and LiverMultiScan outputs from the RADIcAL1 trial

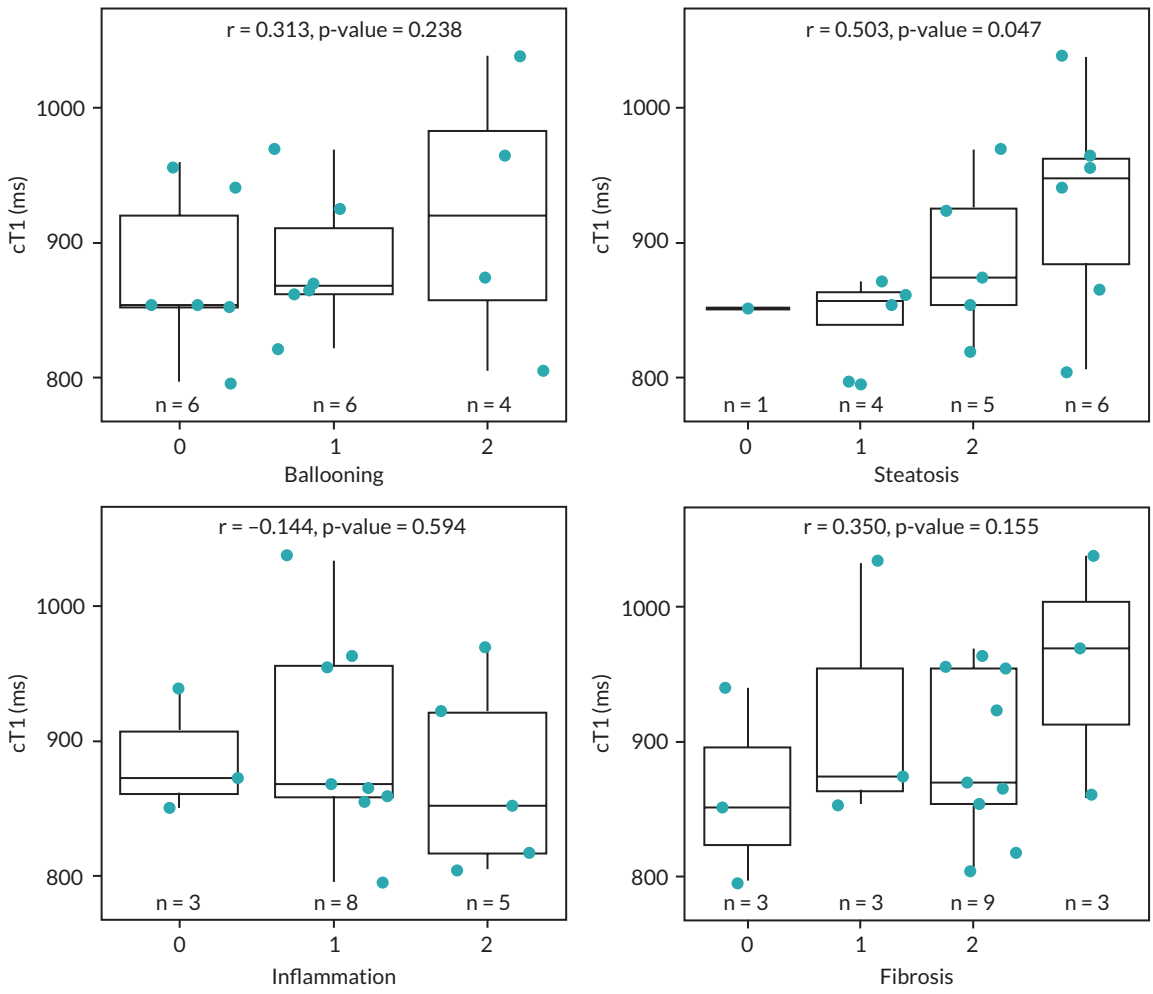

FIGURE 11 Correlations between LiverMultiScan cT1 and histology scores. cT1 = iron-corrected longitudinal relaxation time; *r* = Spearman's rank correlation coefficient.

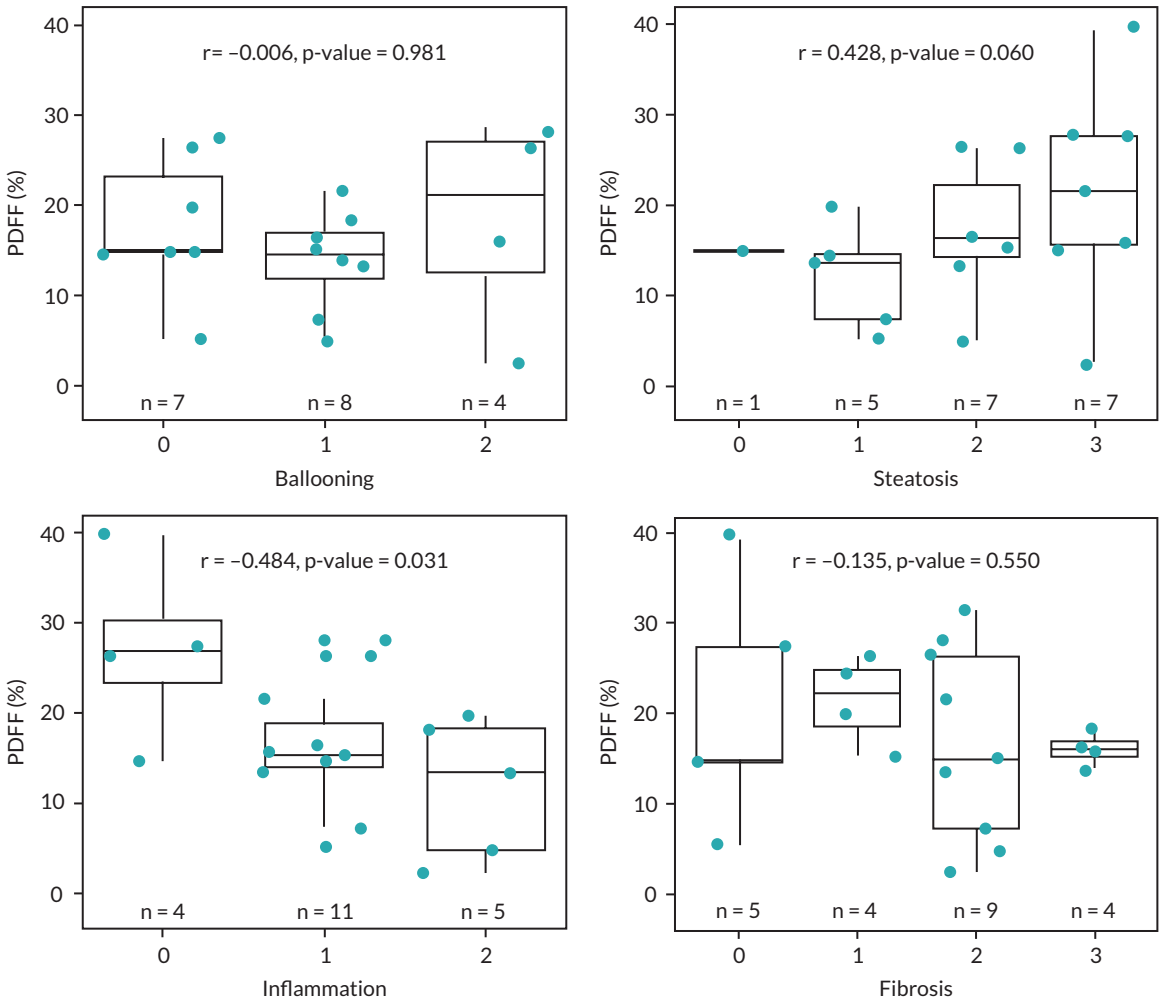

FIGURE 12 Correlations between LiverMultiScan PDFF and histology scores. *r* = Spearman's rank correlation coefficient.

Appendix 6 Results from the External Assessment Group meta-analysis for test failure rate

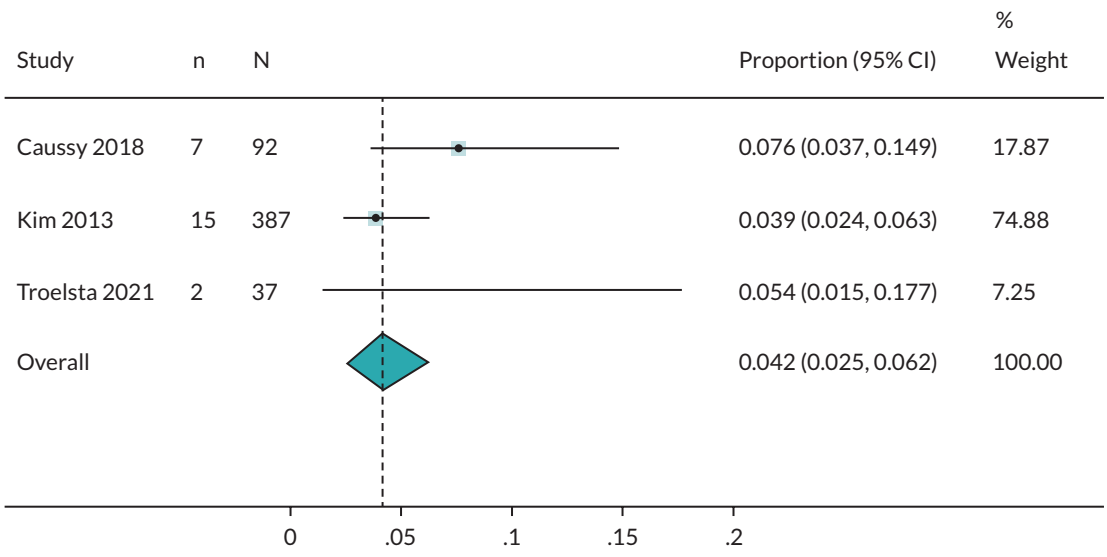

FIGURE 13 Forest plot displaying the EAG meta-analysis for test failure rate of MRE. *n* = number of test failures; *N* = total number of tests.

Appendix 7 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 8 LiverMultiScan PDFF results

TABLE 23 Initial LiverMultiScan outcomes generated by the EAG model (per 1000 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

F = stage of fibrosis.

TABLE 24 LiverMultiScan plus biopsy pathway: biopsies performed and averted (per 1000 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.

TABLE 25 Pathway diagnostic test strategy costs (per 1000 patients)

Diagnostic test strategy	PDFF cut-off value (%)	LMS 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	£8014	£425,709	£1,189,110	£805,000	£8540	£813,540	£375,570
T2: Significant fibrosis (≥F2)	>10	£672,700	£7136	£497,044	£1,176,880	£805,000	£8540	£813,540	£363,340
T3: Advanced fibrosis (≥F3)	>10	£573,475	£6084	£525,578	£1,105,137	£805,000	£8540	£813,540	£291,597
T4: Brunt grade ≥1	>5	£805,000	£8540	£425,709	£1,239,249	£805,000	£8540	£813,540	£425,709
T5: Brunt grade ≥2	>10	£540,400	£5733	£497,044	£1,043,177	£805,000	£8540	£813,540	£229,637
T6: NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	>10	£606,550	£6435	£497,044	£1,110,029	£805,000	£8540	£813,540	£296,489
T7: Advanced NASH (NAS ≥4 plus ≥F2)	>10	£590,013	£6259	£497,044	£1,093,316	£805,000	£8540	£813,540	£279,776

F = stage of fibrosis; LMS = LiverMultiScan.

TABLE 26 Quality-adjusted life year analyses for the two diagnostic pathways (per 1000 patients)

		LMS plus biopsy pathway					Biopsy only pathway				
Diagnostic test strategy	PDFF cut-off value (%)	Biopsy procedure	Biopsy complications	Biopsy death	False negatives	Total QALY losses	Biopsy procedure	Biopsy complications	Biopsy death	Total QALY losses	Difference in QALY losses (LMS+biopsy pathway)[a]
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

LMS = LiverMultiScan.

a A positive value means that the biopsy only pathway is preferred; a negative value means that the LiverMultiScan plus biopsy pathway is preferred.

TABLE 27 Incremental analyses for LiverMultiScan plus biopsy vs. biopsy (1000 patients)

Diagnostic test strategy Fibrosis	PDFF cut-off value (%)	Incremental		ICER per QALY gained (vs. 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. LMS, LiverMultiScan.

Appendix 9 Magnetic resonance elastography analyses

The EAG carried out cost-effectiveness analyses to compare MRE plus biopsy versus biopsy only using sensitivity and specificity data from a population that differed from the population described in the final scope[24] issued by NICE. Therefore, results should only be considered as illustrative. Sensitivity and specificity data are presented in *Table 28*.

Methods and key results

For the costs of MRE, Resoundant, Inc. provided information to the EAG that the approximate cost of adding MRE to an existing MRI machine would be in the region of £35,000, although new machines may add MRE for no additional cost and some centres in the UK already have MRE. The EAG has therefore estimated two costs for MRE – one assuming the MRI device already has MRE capabilities (i.e. the cost of MRE is the same as the cost of MRI alone) and the second assuming that MRE would have to be installed onto the MRI device. To estimate the cost per MRE scan if MRE has to be installed, the EAG divided the £35,000 installation cost by the estimated number of MRE scans that would be undertaken in the NICE scope population over the lifetime of the MRI machine in which MRE was installed. Currently, MRE is only used for the diagnosis of liver disease and so the use of the machine for other diseases does not need to be considered.

To estimate the number of MRE scans in the target population that would be performed over the lifetime of an MRI machine, the EAG required estimates of the:

1. number of patients with NAFLD and indeterminate results from fibrosis testing in England each year
2. number of MRI machines where MRE would be installed
3. average lifespan of existing MRI machines in the UK.

An estimate of the number of people with NAFLD and indeterminate results from fibrosis testing in England each year is difficult to establish. The number of liver biopsies performed each year in England has been estimated to be 7000–8000 liver biopsies per year, with the majority being undertaken for the investigation of liver disease (West 2010).[3] Not all these biopsies are for people with NAFLD with indeterminate results and include biopsies for liver cancer, hepatitis and alcoholic liver disease. The EAG has assumed that half the biopsies were carried out in patients with NAFLD and that half of these patients had indeterminate results from fibrosis testing. Taking the upper bound of 8000 biopsies per year, this means that 2000 per year could be due to patients with NAFLD and indeterminate results from fibrosis testing.

The number of MRI machines in the UK was estimated in 2017 to be 6.1 per million population (Clinical Imaging Board 2017). Applying this to the population in England of 56.5 million (Census 2021) suggests there were approximately 345 MRI machines in England in 2017. Not all MRI machines in the UK would need to be modified for MRE to meet the demand for MRE. The EAG has assumed that with only 2000 patients per year requiring an MRE due to indeterminate results from fibrosis testing, this demand could be met if 10% of the MRI machines available were modified to perform MRE.

Results from a Royal College of Radiographers (RCR) survey (Clinical Imaging Board 2017) showed that the median age of MRI scanners in England was 7 years. The RCR quotes the European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR) that no more than 10% of MRI machines available in a healthcare system should be aged over 10 years old. Taking these factors into account, the average remaining lifespan of MRI machines in England was estimated by the

TABLE 28 Sensitivity and specificity of MRE

		MRE			LiverMultiScan				
			Sensitivity	Specificity		Sensitivity	Specificity	Sensitivity	Specificity
Diagnostic test strategy		Cut-off (kPa)	Perspectum Ltd/Imajo 2021[84]		cT1 cut-off (ms)	Perspectum Ltd/Imajo 2021[84]		Perspectum Ltd/Eddowes 2018[40]	
T1	Any fibrosis (≥F1)	2.9	0.79	1.0	800	0.76	0.60	0.87	0.67
T2	Significant fibrosis (≥F2)	3.3	0.82	0.83	875	0.51	0.65	0.63	0.75
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	0.71	0.41	875	0.65	0.76	0.64	0.67
T7	Advanced NASH (NAS ≥4, ≥F2)	3.5	0.69	0.50	875	0.65	0.68	0.64	0.62

cT1 = iron-corrected longitudinal relaxation time; F = stage of fibrosis. MRE, magnetic resonance elastography.

EAG as 5 years. However, if only 10% of machines were modified to perform MRE then it is reasonable to assume that only the newest machines would be modified. Thus, the EAG has assumed that the effective lifespan for an MRE modified MRI is 10 years.

These estimates can be used to generate the following costs:

- the total cost of adapting 34 MRI machines so that they include MRE is £1,190,000
- the total number of patients with NAFLD and indeterminate results from testing who, over 10 years, have an MRE is 20,000
- the additional cost of MRE is £59.50, making a total cost of MRE of £207.74 (the cost of a standard MRI of £148.24 + the additional cost of MRE of £59.50).

As has been detailed, this cost is built on several assumptions, some of which are not evidenced. Therefore, as was the case for the EAG analysis of LiverMultiScan, the EAG has carried out threshold analyses to determine the price of MRE at which MRE would be cost-effective at WTP thresholds of £20,000 and £30,000 per QALY gained.

The proportion of failed MRE tests was assumed to be identical to the proportion of failed LiverMultiScan tests. The EAG has also used the assumption that was used to generate LiverMultiScan base-case results, that is, all patients with a negative result from a MRE are recalled at 6 months for a second MRE, at which point a correct diagnosis is made.

TABLE 29 Initial MRE outcomes generated by the EAG model (per 1000 tests)

Diagnostic test strategy		Cut-off score (kPa)	True positive	True negative	False positive	False negative	Failed tests
T1	Any fibrosis (≥F1)	2.9	649.5	122.9	0.0	172.7	55.0
T2	Significant fibrosis (≥F2)	3.3	505.2	273.0	55.9	110.9	55.0
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.0	365.0	176.7	254.2	149.1	55.0
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	311.7	246.6	246.6	140.0	55.0

F = stage of fibrosis.

TABLE 30 MRE plus biopsy pathway: biopsies performed and averted (per 1000 patients)

Diagnostic test strategy		Cut-off score (kPa)	Total number of biopsies, including those following a repeated MRE at 6 months	Biopsies averted	Unnecessary biopsies
T1	Any fibrosis (≥F1)	2.9	877.2	122.9	7.2
T2	Significant fibrosis (≥F2)	3.3	727.0	273.0	75.0
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	823.3	176.7	279.3
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	753.4	246.6	275.4

F = stage of fibrosis.

TABLE 31 Pathway diagnostic test strategy costs (per 1000 patients) – MRE cost of £59.50 on top of MRI cost

Diagnostic test strategy		MRE cut-off score (kPa)	MRE plus biopsy pathway costs				Biopsy only pathway costs			Additional cost for the MRE pathway
			Biopsy procedures	Biopsy complications	MRE	Total costs	Biopsy procedures	Biopsy complications	Total costs	
T1	Any fibrosis (≥F1)	2.9	£706,106	£7491	£269,127	£982,724	£805,000	£8540	£813,540	£169,184
T2	Significant fibrosis (≥F2)	3.3	£585,272	£6209	£287,483	£878,964	£805,000	£8540	£813,540	£65,424
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	£662,775	£7031	£275,413	£945,219	£805,000	£8540	£813,540	£131,679
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	£606,451	£6434	£288,068	£900,952	£805,000	£8540	£813,540	£87,412

F = stage of fibrosis; MRE = magnetic resonance elastography; MRI, magnetic resonance imaging; NAS = non-alcoholic fatty liver disease (NAFLD) activity score; NASH = non-alcoholic steatohepatitis.

TABLE 32 Pathway diagnostic test strategy costs (per 1000 patients) – no MRE cost in addition to MRI cost

Diagnostic test strategy		MRE cut-off score (kPa)	MRE plus biopsy pathway costs				Biopsy only pathway costs			Additional cost for the MRE pathway
			Biopsy procedures	Biopsy complications	MRE	Total costs	Biopsy procedures	Biopsy complications	Total costs	
T1	Any fibrosis (≥F1)	2.9	£706,106	£7491	£192,045	£905,642	£805,000	£8540	£813,540	£92,102
T2	Significant fibrosis (≥F2)	3.3	£585,272	£6209	£205,143	£796,624	£805,000	£8540	£813,540	−£16,916
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	£662,775	£7031	£196,531	£866,337	£805,000	£8540	£813,540	£52,797
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	£606,451	£6434	£205,561	£818,445	£805,000	£8540	£813,540	£4905

F = stage of fibrosis. MRI, magnetic resonance imaging.

TABLE 33 Quality-adjusted life year analyses for the two diagnostic pathways (per 1000 patients)

Diagnostic test strategy	MRE cut-off score (kPa)	MRE plus biopsy pathway					Biopsy only pathway				Incremental QALYs (MRE+biopsy pathway)[a]
		Biopsy procedure	Biopsy complications	Biopsy death	False negatives	Total QALY losses	Biopsy procedure	Biopsy complications	Biopsy death	Total QALY losses	
T1 Any fibrosis (≥F1)	2.9	4.89	0.13	1.24	2.59	8.85	5.58	0.15	1.41	7.14	−1.71
T2 Significant fibrosis (≥F2)	3.3	4.06	0.11	1.03	1.66	6.85	5.58	0.15	1.41	7.14	0.28
T6 NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	4.59	0.12	1.16	2.24	8.11	5.58	0.15	1.41	7.14	−0.98
T7 Advanced NASH (NAS ≥4 plus ≥F2)	3.5	4.20	0.11	1.06	2.10	7.48	5.58	0.15	1.41	7.14	−0.34

F = stage of fibrosis.

a A negative value means that the biopsy only pathway generates more QALYs than the MRE+biopsy pathway; a positive value means that the MRE plus biopsy pathway generates more QALYs than the biopsy only pathway.

TABLE 34 Incremental analyses for MRE plus biopsy vs. biopsy (1000 patients) – MRE cost of £59.50 on top of MRI cost

Diagnostic test strategy		MRE cut-off score (kPa)	QALY loss from false negatives			No QALY loss from false negatives		
			Incremental			Incremental		
			Costs	QALYs	ICER per QALY gained (vs. biopsy)	Costs	QALYs	ICER per QALY gained (vs. biopsy)
T1	Any fibrosis (≥F1)	2.9	£169,184	−1.71	MRE+biopsy dominated by biopsy	£169,184	0.88	£192,961
T2	Significant fibrosis (≥F2)	3.3	£65,424	0.28	£229,967	£65,424	1.95	£33,584
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	£131,679	−0.98	MRE+biopsy dominated by biopsy	£131,679	1.26	£104,429
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	£87,412	−0.34	MRE+biopsy dominated by biopsy	£87,412	1.76	£49,657

F = stage of fibrosis. MRI, magnetic resonance imaging.

TABLE 35 Incremental analyses for MRE plus biopsy vs. biopsy (1000 patients) – no MRE cost on top of MRI cost

Diagnostic test strategy		MRE cut-off score (kPa)	QALY loss from false negatives			No QALY loss from false negatives		
			Incremental			Incremental		
			Costs	QALYs	ICER per QALY gained (vs. biopsy)	Costs	QALYs	ICER per QALY gained (vs. biopsy)
T1	Any fibrosis (≥F1)	2.9	£92,102	−1.71	MRE+biopsy dominated by biopsy	£92,102	0.88	£105,045
T2	Significant fibrosis (≥F2)	3.3	−£16,916	0.28	MRE+biopsy **dominates** biopsy	−£16,916	1.95	MRE+biopsy dominates biopsy
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	£52,797	−0.98	MRE+biopsy dominated by biopsy	£52,797	1.26	£41,871
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	£4905	−0.34	MRE+biopsy dominated by biopsy	£4905	1.76	£2787

F = stage of fibrosis. MRI, magnetic resonance imaging.

Threshold analysis

In addition to base-case analyses, the EAG undertook threshold analysis to determine at what prevalence and total cost the different MRE testing strategies would become cost-effective at £20,000 and £30,000 (*Table 36*). Results without any additional cost of MRE over a standard MRI are provided in *Table 37*.

TABLE 36 Magnetic resonance elastography plus biopsy vs. biopsy (1000 patients) – prevalence and total MRE cost at which MRE becomes cost-effective at different QALY WTP thresholds

Diagnostic test strategy		MRE cut-off score (kPa)	Base-case prevalence from CALM trial (%)	£20,000/QALY			£30,000/QALY		
				Prevalence (QALY loss from false negative) (%)	Prevalence (no QALY loss from false negative) (%)	Price of MRE at which it becomes cost-effective	Prevalence (QALY loss from false negative) (%)	Prevalence (no QALY loss from false negative) (%)	Price of MRE at which it becomes cost-effective
T1	Any fibrosis (≥F1)	2.9	87	62	67	£50.70*	63	69	£37.48[a]
T2	Significant fibrosis (≥F2)	3.3	65	56	61	£164.58	58	64	£166.63
T6	NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	54	19	24	£93.70*	22	29	£86.35[a]
T7	Advanced NASH (NAS ≥4 plus ≥F2)	3.5	48	29	35	£139.80*	31	40	£137.34[a]

F = stage of fibrosis.

a The ICERs in these scenarios are in the south-west quadrant and as such lower costs for MRE are required to make the QALY loss associated with each strategy compared to no MRE cost-effective as the WTP threshold increases from £20,000 to £30,000 per QALY.

TABLE 37 Magnetic resonance elastography plus biopsy vs. biopsy (1000 patients) – prevalence at which MRE becomes cost-effective at different QALY WTP thresholds with no additional cost per MRE over a standard MRI

Diagnostic test strategy	MRE cut-off score (kPa)	Base-case prevalence from CALM trial (%)	£20,000/QALY		£30,000/QALY	
			Prevalence (QALY loss from false negative)	Prevalence (no QALY loss from false negative) (%)	Prevalence (QALY loss from false negative) (%)	Prevalence (no QALY loss from false negative) (%)
T1 Any fibrosis (≥F1)	2.9	87	72%	78	72	79
T2 Significant fibrosis (≥F2)	3.3	65	MRE+biopsy dominates biopsy			
T6 NASH (NAS ≥4, ≥1 for lobular inflammation and hepatocyte ballooning)	3.3	54	19%	47	22	50
T7 Advanced NASH (NAS ≥4 plus ≥F2)	3.5	48	46%	55	45	58

F = stage of fibrosis. MRI, magnetic resonance imaging.

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library