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

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

Scientific summary

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

Health Technology Assessment 2023; Vol. 27: No. 10
DOI: 10.3310/KGJU3398

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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.

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).