Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease

Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease

Running head: Evaluation of multiparametric MRI in NAFLD

Peter J Eddowes*, Natasha McDonald*, Nigel Davies, Scott IK Semple, Timothy J Kendall, James Hodson, Phillip N Newsome, Robert B Flintham, Roman Wesolowski, Laurence Blake, Rui V Duarte, Catherine J Kelly, Amy H Herlihy, Matthew D Kelly, Simon P Olliff, Stefan G Hübscher, Jonathan A Fallowfield†, Gideon M Hirschfield[1†].

*Joint 1st authors †Joint senior authors

1. Centre for Liver Research and National Institute for Health Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK and Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK.

Corresponding author

Prof. GM Hirschfield, Centre for Liver Research, University of Birmingham, Birmingham, B15 2TH, UK. E-mail: g.hirschfield@bham.ac.uk

Financial Support

This work is supported by a grant from Innovate-UK (101679). The study sponsor was the University of Birmingham.

Peter J Eddowes, Phillip N Newsome and Gideon M Hirschfield are supported by the National Institute of Health Research Birmingham Biomedical Research Centre.

Jonathan A Fallowfield is supported by a NHS Research Scotland/Universities Senior Clinical Fellowship.

This paper presents independent research supported by the National Institute for Health Research Birmingham Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Disclosures

This is an academic led and reported study, with industry engagement. Funding for the project was from an Innovate-UK grant (101679). The role of Perspectum Diagnostics Ltd. was the provision of access to Liver*Multiscan*™ and blinded analysis of raw MRI data. All study investigations and data analysis was performed by the academic centres. Peter J Eddowes, Natasha McDonald, Nigel Davies, Scott IK Semple, Robert B Flintham, Roman Wesolowski, Laurence Blake, Rui V Duarte, Timothy J Kendall, James Hodson, Simon P Olliff, Stefan G Hübscher, Jonathan A Fallowfield and Gideon M Hirschfield have no relevant conflict of interest to declare. Philip N Newsome has consulted for EchoSens and was a member of the National Institute for Health and Clinical Excellence NAFLD guidelines committee. Catherine J Kelly, Amy H Herlihy and Matthew D Kelly are employees of Perspectum Diagnostics Limited, the developer of Liver*Multiscan*™.

Acknowledgements

Professor Gideon Hirschfield is the guarantor for this work

Peter J Eddowes wrote the study protocol, collected data, analysed data and wrote the manuscript. Natasha McDonald was involved in the design of the study, collected data and provided critical review of the manuscript. Nigel Davies, Scott IK Semple, Robert B

Flintham and Roman Wesolowski were involved in study design and development of the MRI protocol. Timothy J Kendall and Stefan G Hübscher were involved in study design and analysed pathology specimens. James Hodson was the study statistician. Phillip N Newsome was involved in protocol design and critical review of the manuscript. Laurence Blake and Rui V Duarte performed the cost analysis. Catherine J Kelly, Amy H Herlihy and Matthew D Kelly analysed MRI data. Simon P Olliff provided MRI safety reporting. Jonathan A Fallowfield and Gideon M Hirschfield devised the study, were involved in study design and critical review of the manuscript.

All authors have reviewed and approved the final version of the manuscript.

We are grateful to Professor Stefan Neubauer (University of Oxford and Perspectum Diagnostics Limited) for critical review of the manuscript and contribution to the award, and delivery of the Innovate-UK grant (101679).

Abstract

Background: Validated diagnostic tools that are accurate, cost effective and acceptable to patients are required for disease stratification and monitoring in NAFLD.

Aims: We investigated the performance and cost of multiparametric MRI alongside existing biomarkers in the assessment of NAFLD.

Methods: Adult patients undergoing standard of care liver biopsy for NAFLD were prospectively recruited at two UK liver centres and underwent multiparametric MRI, blood sampling and transient elastography withing 2 weeks of liver biopsy. Non-invasive markers were compared to histology as the gold standard.

Results: Data were obtained in 50 patients and 6 healthy volunteers. Corrected T1 (cT1) correlated with NAFLD activity score (Rho=0.514, p<0.001). cT1, enhanced liver fibrosis (ELF) test and liver stiffness differentiated patients with simple steatosis and NASH with AUROC (95% CI) of 0.69 (0.50-0.88), 0.87 (0.77-0.79) and 0.82 (0.70-0.94) respectively and healthy volunteers from patients with AUROC (95% CI) of 0.93 (0.86-1.00), 0.81 (0.69-0.92) and 0.89 (0.77-1.00) respectively. For the risk stratification of NAFLD, multiparametric MRI could save £150,218.00 per 1000 patients compared to biopsy. Multiparametric MRI did not discriminate between individual histological fibrosis stages in this population (p=0.068).

Conclusions: Multiparametric MRI accurately identified patients with steatosis, stratifies those with NASH or simple steatosis and reliably excludes clinically significant liver disease with superior negative predictive value (83.3%) to liver stiffness (42.9%) and ELF (57.1%).

For the risk stratification of NAFLD multiparametric MRI was cost effective and, combined with transient elastography, had the lowest cost per correct diagnosis.

Keywords: Non-alcoholic fatty liver disease; non-alcoholic steatohepatitis; hepatic fibrosis; magnetic resonance T1 mapping, cost effectiveness.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is an important and growing clinical concern associated with the increasing prevalence of obesity, type 2 diabetes, hypertension and dyslipidaemia. With the global prevalence of NAFLD estimated at 25.24%,[1] effective strategies are required to ensure optimal risk stratification of patients in order that appropriate treatment is both offered and developed. As an umbrella term, NAFLD encompasses two main clinico-pathological entities: simple steatosis and non-alcoholic steatohepatitis (NASH). These conditions have distinct clinical significance and prognosis. Simple steatosis indicates hepatic steatosis without inflammation and whilst simple steatosis has been shown to increase mortality from cardiovascular disease, natural history studies indicate that simple steatosis leads to little or no progression of fibrosis and no increase in liver related mortality.[2, 3] In contrast, NASH is characterised histologically by hepatocyte ballooning and lobular inflammation, culminates in cirrhosis in up to 20%[4] and carries a hepatocellular carcinoma risk of 5.29 cases per 1,000 person years.[1]

The majority of liver related morbidity and mortality in NAFLD is secondary to the complications of advanced fibrosis and fibrosis stage predicts clinical outcome.[5-7] This offers valuable prognostic information to patients and their clinicians. In addition to fibrosis staging, distinguishing between simple steatosis and NASH is highly relevant for risk stratification in clinical practice and the enrichment of interventional trial populations.

Presently, the staging of NAFLD is largely dependent on interpretation of liver biopsy specimens, with the inherent challenges of procedural complications, cost, patient acceptability and sampling error.[8] Alternative approaches to biopsy have been sought across liver disease aetiologies and include serum markers of fibrosis such as the Fibrosis-4 score, NAFLD fibrosis score, Enhanced Liver Fibrosis (ELF) test (iQur Limited, London, UK)[9] and imaging technologies based on transient elastography[10] and magnetic resonance imaging (MRI).[11] Many of these non-invasive techniques have acquired a role in routine clinical practice. To date for example, National Institute for Health and Clinical Excellence guidance for NAFLD recommends ELF as a risk stratification tool, and in the National Institute for Health and Clinical Excellence guidance on cirrhosis; transient elastography is recommended for staging of fibrosis.[12, 13]

As diagnostic tests evolve and clinical and trial strategies change, comparative evaluation of novel technologies is critically important for users. We sought to determine how a novel quantitative liver MRI technology performed in terms of utility and comparative effectiveness, in the assessment of a prospective cohort of patients with NAFLD having routine liver biopsy as standard of care. We subsequently used the data to assess the potential cost effectiveness of this technique to understand whether this added investigation would reduce the cost burden of liver biopsy.

Our study was an academic led, funded, and delivered evaluation of a proprietary algorithm - LiverMultiscan™ (Perspectum Diagnostics Ltd. Oxford, UK), a multiparametric MRI

technology used to quantify liver fat, iron and fibro-inflammatory injury by proton density fat fraction, T2* mapping and corrected T1 (cT1)[14] mapping, respectively. T1 as a biomarker of hepatic fibrosis is confounded by hepatic siderosis. cT1 uses a measurement of iron content (T2*) and a patented algorithm to correct for this confounding.[11] cT1 of the liver has previously been reported to stage hepatic fibrosis in an unselected population of patients undergoing liver biopsy.[11] The proton density fat fraction measured using a modified Dixon sequence is a well-established and accurate technique for the assessment of hepatic fat content.[15-17] Iron concentration was estimated from T2* according to a previously determined model.[18] We report our comparative evaluation of the ability of this quantitative MRI technology to identify patients with hepatic steatosis, differentiate those with simple steatosis from those with NASH, grade disease activity and stage hepatic fibrosis. We demonstrate maximal utility in the ability of multiparametric MRI to distinguish between patients at low risk and those at high risk of progressive disease.

Methods

Study Participants

Our prospective study was undertaken at the Queen Elizabeth Hospital Birmingham and Royal Infirmary of Edinburgh between February 2014 and September 2015. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki, and was approved by the National Research Ethics Service (West Midlands – The Black Country; Ref: 14/WM/0010). The study was registered with the International Standard Randomised Controlled Trial Number registry (ISRCTN39463479) and the National Institute of Health Research portfolio (15912). The study sponsor was the University of Birmingham. Male and female adult (≥18 years of age) patients booked for non-targeted liver biopsy for any indication were prospectively recruited to a validation study of Liver*MultiScan*™ (currently being prepared for publication). Patients with a histologically confirmed diagnosis of NAFLD without secondary cause and without history of alcohol excess (men >21 UK units/week, women >14 UK units/week) were included in this sub-group analysis. Data on the indication for biopsy are included in the supplementary material. Exclusion criteria were: biopsy of a distinct focal lesion, inability to give fully informed consent and any contraindication to MRI.

Healthy volunteers were recruited from staff and students at the University of Birmingham. Exclusion criteria were obesity (Body mass index (BMI) >30kg/m^2), current or previous history of liver disease, significant medical co-morbidity, family history of liver disease, excess alcohol intake or any contraindication to MRI. All participants gave written, informed

consent and attended for a single study visit during which they underwent multiparametric MRI, transient elastography with FibroScan™ (Echosens, Paris, France), blood sampling and collection of clinical and demographic data. All study investigations were performed after a 4 hour fast. Patients undertook their study visit in the 2 weeks prior to liver biopsy and healthy volunteers did not undergo liver biopsy.

Study Investigations

All MRI scans were performed at 3.0 Tesla on Siemens Verio MRI scanners (Siemens Healthcare GMBH, Erlangen, Germany). The MRI protocol does not require intravenous contrast and has been previously described.[11] In brief, the participant lies supine with 3-lead ECG for cardiac gating. A combination of body matrix and spine matrix coil elements was used to acquire data. Following localisers and shimming, the sequences include: shortened modified Look Locker inversion (ShMOLLI) recovery sequence (T1 mapping), multi-gradient-echo sequence (T2* mapping), modified Dixon sequence and proton magnetic resonance spectroscopy with the Stimulated Echo Acquisition Mode sequence. All data were acquired during diastole with breath held in expiration to minimise movement artefact. Maps were acquired in a transverse plane through the liver hilum using the same slice position for each sequence. The voxel for magnetic resonance spectroscopy was placed in the right lobe of the liver avoiding biliary and vascular structures. Details of the MRI sequence parameters are contained in the supplementary material. Due to an error in the MRI acquisition protocol, proton density fat fraction as measured by the Dixon sequence (PDFF-Dixon) could not be calculated for the first 12 data sets. At this point the error was identified and corrected. PDFF-Dixon was calculated reliably for the remaining participants.

Transient elastography was performed by trained operators (PJE and NM) in accordance with manufacturer's guidelines and validated local clinical practice.[19] The decision on using the M probe or XL probe was made by the automated probe selection tool incorporated into the FibroScan™ machine. Examinations were regarded as 'possible' if at least 10 valid readings could be recorded and 'reliable' if they contained at least 10 valid readings and had interquartile range (IQR) to median ratio ≤30% (Boursier's criteria).[20] At the start of the study the Controlled Attenuation Parameter was not available on the FibroScan™ XL probe. Controlled Attenuation Parameter on the XL probe was enabled during study recruitment so was recorded if available in addition to median liver stiffness.

Blood samples were analysed routinely for markers of liver disease. Simple blood biomarker panels including aspartate aminotransferase/alanine aminotransferase (AST/ALT) ratio, Fibrosis-4 (FIB-4) and NAFLD fibrosis score were calculated according to published formulae.[21, 22] Serum was also analysed to determine the enhanced liver fibrosis (ELF) score.

MRI data analysis and iron correction

T1, T2* and PDFF-Dixon maps were analysed using Liver*Multiscan*™ software by a single operator (AHH) blinded to the clinical findings and biopsy results. A user defined region of interest of approximately 1.4cm^3 was placed in a representative area of the right lobe of the

liver avoiding vascular and biliary structures on the T1, T2* and PDFF-Dixon maps. Liver*Multiscan*™ software then calculates a corrected T1 value (cT1).

Magnetic resonance spectroscopy data were analysed by a single operator (RBF) blinded to the histology results. Proton density fat fraction measured with magnetic resonance spectroscopy (PDFF-MRS) was calculated from the non-water-suppressed Stimulated Echo Acquisition Mode acquisition (5 measurements of 1 signal average) using totally automatic robust quantitation in nuclear magnetic resonance (TARQUIN) software to perform automated preprocessing and fitting of the fat/liver spectrum. [23] Visual quality control of fitted spectra was performed by a medical physicist with experience of in vivo Magnetic resonance spectroscopy. Poorly fitted spectra were excluded from the analysis. PDFF-MRS was defined by the equation:

$$\text{PDFF-MRS} = \frac{\text{area under methylene (1.3ppm) peak}}{\text{area under methylene (1.3ppm) peak} + \text{area under water peaks}}$$

Histological assessment

Liver biopsy samples were taken with 16-gauge biopsy needles. Histology was assessed by experienced academic liver histopathologists blinded to the MRI, ELF and transient elastography findings. Biopsies that were less than 15 mm in length or that contained fewer than 11 portal tracts were regarded as inadequate for histological assessment and were

therefore excluded.[24, 25] Fibrosis and steatosis were staged according to the system described by Kleiner *et al.*[26, 27] and siderosis according the Scheuer grading system.[28]

Biopsies were categorised as NASH based on the presence of lobular inflammation and hepatocyte ballooning.[29] Overall disease activity was graded according to the NAFLD activity score.[26] Biopsy sections were also stained with Picro-Sirius Red and morphometry used to determine the collagen proportionate area (%) as previously described.[30]

Statistical analysis

Statistical analysis was carried out using IBM SPSS Statistics for Windows version 22 (IBM Corp, Armonk, NY). Variables are summarised with mean ± standard deviation (SD) if normally distributed and with median and range if not normally distributed. Comparisons between patients and healthy volunteers were performed using independent samples t-tests, Mann-Whitney tests, or Fisher's exact tests, as applicable. Correlation between continuous variables was determined with Spearman's correlation coefficient (Rho). Comparisons across variables with multiple groups were performed using Kruskal-Wallis tests for nominal variables, or Jonckheere–Terpstra tests for ordinal variables. Post-hoc pairwise comparisons between groups were performed using Dunn's test. Diagnostic performance was compared by calculation of the receiver operating characteristic and determination of the area under the curve (AUROC) with 95% confidence intervals (CI). For all tests, a p-value <0.05 was taken to indicate statistical significance.

Decision analytic model

A recent study investigated the potential cost savings of adding multi-parametric MRI to the NAFLD risk stratification pathway.[31] A decision tree model was developed to compare the expected outcomes and costs associated with three potential risk stratification pathways for NAFLD: using transient elastography alone, using multiparametric MRI as an adjunct to transient elastography, and using multiparametric MRI alone. We repeated this analysis, using the data collected from our study, based on the risk stratification pathways detailed in supplementary figure 1. For each risk stratification pathway, failure rates of transient elastography and MRI were used as reported in the previous study[31] and we presumed the failure rate of ELF to be negligible at 0%. Using the prevalence of NAFLD in the previous study cohort (58.8%) and the sensitivities and specificities of each test derived in this study (**Table 8**), the proportions of patients that would have been correctly and incorrectly stratified were then calculated for each pathway. The expected test results were used to estimate the number of biopsies that would have been performed for a cohort of 1,000 patients. The total cost of each pathway was then evaluated, by adding together the costs of the tests and biopsies (based on NHS tariff data contained in the recent National Institute for Health and Clinical Excellence guidelines[12]) that would have been required per 1,000 patients. This was then subtracted from the total cost that would result from performing biopsies on all patients, in order to estimate the cost savings of each pathway. The resulting value was divided by the number of correct diagnoses, in order to estimate the cost per correct diagnosis, which could then be compared to the cost of a biopsy.

Results

Patient demographics

Of 54 patients with NAFLD recruited into the study (whole study included 162 unselected liver biopsies) 50 had sufficient data for analysis. Three MRI data sets were unusable and 1 biopsy was judged too small for reliable fibrosis assessment and so these 4 patients were excluded. Seven healthy volunteers were recruited. One healthy volunteer was subsequently excluded from analysis due to the discovery of abnormal liver biochemistry. The study flow chart is shown in supplementary figure 2. The characteristics of the 50 patients and 6 healthy volunteers are outlined in **Table 1**. Comparisons between these groups found that the healthy volunteers were significantly younger (median 32 vs. 54 years, p=0.011), had significantly lower BMI and lower waist to hip ratio. Healthy volunteers were more likely to consume alcohol than patients but there was no difference in the median consumption of drinkers and no patient or healthy volunteer drank alcohol to excess.

In 49/50 (98%) patients, transient elastography was possible (≥10 valid readings) and in 47/50 (94%) was reliable by Boursier's criteria.[20] Non-reliable transient elastography examinations were excluded from further analysis. In 16/47 (34%) patients, transient elastography was measured with the M probe and the remainder with the XL probe. For all healthy volunteers, transient elastography was measured with the M probe and was reliable by Boursier's criteria.

Median (range) length of liver biopsy samples was 25 (15-50) mm. Median (range) collagen proportionate area was 5.3 (0.6-34.2) %. Collagen proportionate area correlated strongly with Kleiner fibrosis stage ($p<0.001$) (supplementary figure 3). The characteristics of the histology and distribution of fibrosis stages in the cohort are shown in **Table 2**. Twelve (24%) of the patients had simple steatosis and 38 (76%) had NASH.

Grading of steatosis by multiparametric MRI

PDFF-Dixon data was available for 38/50 (76%) patients and all healthy volunteers. Median PDFF-Dixon for healthy volunteers, grade 1, grade 2 and grade 3 steatosis were 1.8, 6.6, 15.3 and 21.4% respectively ($p<0.001$) (**Figure 1**A). PDFF-MRS was available for 43/50 (86%) patients and 5/6 (83%) healthy volunteers. Median PDFF-MRS for volunteers, grade 1, grade 2 and grade 3 steatosis were 0.3, 11.3, 23.7, and 31.5% respectively ($p<0.001$) (**Figure 1**B). AUROC (95% CI) for the identification of steatosis (Brunt grade $\geq$1) for both PDFF-Dixon and PDFF-MRS was 1.00 (1.00-1.00).

Controlled Attenuation Parameter was available in 24/50 (48%) patients and all healthy volunteers. Median Controlled Attenuation Parameter for healthy volunteers, grade 1, grade 2 and grade 3 steatosis were 246, 331, 361, 344 dB/m respectively ($p=0.002$) (**Figure 1**C). AUROC (95% CI) for Controlled Attenuation Parameter for the identification of steatosis (Brunt grade $\geq$1) was 0.95 (0.87-1.00).

Differentiation between NASH and simple steatosis by cT1 measurement

Demographic characteristics and blood results presented in **Table 3** showed no significant difference between patients with NASH and those with simple steatosis. cT1 showed a significant difference between simple steatosis and NASH and, although not validated for this purpose, liver stiffness and ELF also showed significant differences between patients with simple steatosis and those with NASH (**Table 4**). Whilst multiparametric MRI did differentiate between NASH and simple steatosis, the AUROC (95% CI) for cT1 0.69 (0.50-0.88) was inferior to ELF 0.87 (0.77-0.79) and liver stiffness 0.82 (0.70-0.94) (**Figure 2**).

Grading of disease activity by cT1 measurement

In patients with NAFLD, semi-quantitative assessment of hepatocyte ballooning showed a statistically significant correlation with cT1, liver stiffness, and ELF. Lobular inflammation was significantly correlated with liver stiffness and ELF but not cT1. Overall assessment of disease activity (as defined by the total NAFLD activity score) showed significant correlation with cT1, liver stiffness and ELF . The strength of these correlations can be seen in **Table 5** and graphically in **Figure 3** with p-values from the Jonckheere–Terpstra test.

AUROC (95% CI) to differentiate those with NAFLD activity score <5 and NAFLD activity score ≥5 was statistically significant for cT1, liver stiffness, ELF and FIB-4, 0.74 (0.59-0.88), 0.74 (0.59-0.89), 0.74 (0.59-0.89) and 0.73 (0.58-0.88) respectively. Statistical significance

was not reached by AST:ALT ratio and NAFLD fibrosis score 0.60 (0.43-0.77) and 0.63 (0.47-0.77), respectively.

Staging of liver fibrosis by cT1 measurement

Mean (±SD) cT1 for healthy volunteers was 791 (±42) ms. For patients with NAFLD with F0, F1, F2, F3 and F4 fibrosis mean (±SD) cT1 was 882 (±141), 969 (±115), 985(±93), 1016 (±97) and 997 (±86) ms respectively as can be seen in **Figure 4**. Statistically significant differences were demonstrated between healthy volunteers and F2 fibrosis (p=0.048) and F3 fibrosis (p=0.003). However, cT1 showed no significant trend across the fibrosis stages (p=0.068), with pairwise comparisons finding no evidence of significant differences between individual fibrosis stages in patients with NAFLD. As shown in **Figure 5**, there was no evidence of significant correlation between cT1 and collagen proportionate area in patients with NAFLD (Rho=0.142, p=0.324).

In NAFLD patients there was a significant association between Kleiner fibrosis stage and ELF (p<0.001), Liver stiffness (n=47) (p<0.001), NAFLD fibrosis score (p=0.003), AST/ALT ratio (p=0.002) and FIB-4 (p=0.013) (**Figure 4**). Collagen proportionate area showed significant correlation with: ELF (Rho=0.404, p=0.004), liver stiffness (n=47) (Rho=0.511, p<0.001), NAFLD fibrosis score (Rho=0.306, p=0.030), AST/ALT ratio (Rho=0.453, p=0.001) and FIB-4 (Rho=0.292, p=0.039) (**Figure 5**).

To diagnose clinically significant (defined as ≥F2) fibrosis in patients with NAFLD, the AUROC (95% CI) for ELF, liver stiffness, AST:ALT ratio, NAFLD fibrosis score and FIB-4 were statistically significant; 0.90 (0.82-0.99), 0.90 (0.81-0.99), 0.78 (0.64-0.93), 0.72 (0.54-0.89) and 0.69 (0.52-0.86) respectively. cT1 did not reach statistical significance with AUROC (95% CI) of 0.63 (0.45-0.81). To diagnose advanced (defined as ≥F3) fibrosis in patients with NAFLD, liver stiffness, ELF and NAFLD fibrosis score were statistical significant with AUROC (95% CI) of 0.88 (0.76-0.99), 0.80 (0.68-0.93) and 0.66 (0.50-0.82), respectively. AST:ALT ratio, cT1 and FIB-4 did not reach statistical significance with AUROC (95%CI) of 0.63 (0.47-0.79), 0.62 (0.46-0.78) and 0.61 (0.45-0.78), respectively.

Comparative utility of multiparametric MRI to exclude significant liver disease

10/50 (20%) of patients in the cohort were classified as being at low risk for progressive liver disease. This was defined as simple steatosis without clinically significant (>F1) fibrosis. cT1, liver stiffness, ELF, AST:ALT ratio and NAFLD fibrosis score showed statistically significant differences between healthy volunteers, low risk patients and high risk patients as shown in **Table 6**. The AUROC (95% CI) to differentiate the different groups is shown in **Table 7**, and confirmed effective utility of multiparametric MRI as well as liver stiffness and ELF to exclude liver disease, with cT1 having the highest AUROC for differentiation of NAFLD and healthy volunteers (0.93).

Taking common cut-off values for the three best performing tests, sensitivity, specificity, negative predictive value and positive predictive value for the diagnosis of high risk patients were calculated and are shown in **Table 8**. Importantly, negative predictive values,

suggesting those patients for whom biopsy could potentially be avoided, were substantially higher for cT1 (80.0-83.3%) compared to liver stiffness (39.1-42.9%) and ELF (26.3-57.1%).

Cost analysis of non-invasive tests for the staging of NAFLD

The results of applying the sensitivity and specificity described in this study to the previously published decision tree model using our cohort of patients with NAFLD are provided in **Table 9**. The risk stratification pathways considered the use of each test individually and also the combination of transient elastography followed by multiparametric MRI. For example, the use of the "cT1 only" risk stratification pathway using a 875ms cutoff was estimated to reduce the number of biopsies required by almost half (reduction of 458 per 1,000 patients). As a result, the estimated saving was £150,218 per 1,000 patients, relative to the pathway in which biopsies are performed on all patients. All of the pathways considered were found to be potentially cost saving, relative to biopsy alone, with the exception of ELF at the lower cut-off of 7.7. In addition, those pathways that combined transient elastography and multiparametric MRI provided additional cost savings over multiparametric MRI alone. For example, the estimated cost per correct diagnosis was £554.26 using cT1 with a cutoff of 875ms, which reduced to £307.92 in the pathway where this was preceded by liver stiffness (7.0kPa cutoff).

Discussion

The global burden of NAFLD is increasing inexorably and validated non-invasive diagnostic tests are important for patients, clinicians and industry. This is not only the first independent validation study to assess the diagnostic accuracy of multiparametric MRI with Liver*Multiscan*^{TM} in NAFLD, but also the first study to compare the performance and potential cost-effectiveness of this emerging methodology against more established non-invasive biomarkers of liver disease. In our prospectively recruited population we demonstrated the ability of multiparametric MRI to grade hepatic steatosis with a high degree of accuracy. Moreover, multiparametric MRI demonstrated accurate differentiation of patients with simple steatosis from those with NASH and also correlated in a highly significant manner with overall disease activity as defined by NAFLD activity score. However, in this cohort, multiparametric MRI did not predict the severity of histological liver fibrosis. Multiparametric MRI demonstrated the greatest negative predictive value for excluding significant liver disease in those with NAFLD.

Identifying those patients with NAFLD requires accurate detection of steatosis. In clinical practice, steatosis is typically assessed by visual grading of standard liver ultrasound images.[32] Although the sensitivity of ultrasound in detecting moderate and severe steatosis is good, there is wide interobserver and intraobserver variability.[32] Other non-invasive techniques for steatosis assessment such as the Fatty Liver Index have moderate to good accuracy but are significantly confounded by fibrosis stage and are unable to monitor changes in steatosis.[33, 34] PDFF-Dixon has been shown in this study to have excellent accuracy in

differentiating patients with steatosis on liver biopsy from healthy volunteers with AUROC of 1.0. PDFF-Dixon also correlated strongly with PDFF-MRS (Rho=0.975, p<0.001), which is widely regarded at the most accurate method for non-invasive quantification of liver fat.[17, 35] Comparison of the accuracy of PDFF-Dixon and Controlled Attenuation Parameter for the detection of steatosis must be made with caution due to the small numbers of patients in this study with both a PDFF-Dixon and Controlled Attenuation Parameter reading. Both techniques had very high accuracy for the detection of any steatosis.

PDFF-Dixon demonstrated a clear, stepwise increase with advancing Brunt steatosis grade in patients with NAFLD suggesting that multiparametric MRI could be used as an accurate method of monitoring steatosis progression and regression and assessing the therapeutic response to lifestyle or drug interventions in the context of clinical trials. In this study, Controlled Attenuation Parameter did not demonstrate the same stepwise increase with Brunt grade observed with PDFF-Dixon measurement.

A single test to reliably exclude NAFLD would be of considerable value in clinical practice. In this study multiparametric MRI showed a high degree of accuracy for differentiating between healthy volunteers and those with NAFLD with AUROC (95% CI) of 0.93 (0.86-1.00). It should be recognised however, that the healthy volunteers and patients enrolled in this study were not well matched in terms of age, waist to hip ratio or BMI. Accepting this limitation, using a cT1 cut-off value of 875ms gave multiparametric MRI a sensitivity of 88.0% with specificity of 100% for the detection of any liver disease. This was superior to all

other non-invasive tests. In addition, the negative predictive value for excluding any liver disease was substantially higher than those for the other non-invasive techniques. This indicates an opportunity to consider further work to establish the ability of multiparametric MRI to be used as a single one-stop comprehensive MRI examination to rule out significant liver disease.

When used in a risk stratification pathway for people with NAFLD in whom biopsy is clinically indicated, these non-invasive tests could be cost-saving if applied as first line tests. Furthermore, the combination of transient elastography with multiparametric MRI would provide the lowest total cost and, because diagnostic accuracy is maintained, would result in a lower total cost per correct diagnosis.

The differentiation of those with NASH from those with simple steatosis is also an important distinction in clinical practice as our current understanding of NAFLD recognises NASH as the harbinger of progressive fibrosis and hepatocellular carcinoma.[2, 3, 36] Identifying individuals with NASH stratifies patients at risk of significant disease and may do so at an earlier stage than tests that reflect fibrosis alone.[37] Detection of NASH guides decision making about clinical management and follow-up intensity and identifies patients who may be eligible for recruitment to clinical studies.[38] To date, the differentiation of simple steatosis and NASH has been reliant on liver biopsy. Liver biopsy has low patient acceptability due to its invasiveness and associated risk. Liver biopsy is also prone to sampling error and interobserver variation of histological assessment. These factors reduce the suitability and

reliability of liver biopsy for disease stratification in NAFLD. Currently available methods to non-invasively differentiate NASH and simple steatosis are suboptimal. Conventional blood tests and imaging techniques have low accuracy for the differentiation of simple steatosis and NASH.[38]

To determine the severity of disease, cT1 showed a highly significant, positive correlation with NAFLD activity score. The correlation between NAFLD activity score and cT1 was stronger than between NAFLD activity score and any other evaluated test, althought the authors acknowledge that these tests are not designed to assess disease activity. This indicates the potential utility of multiparametric MRI as a sensitive diagnostic tool to monitor changes in disease activity. A NAFLD activity score of ≥5 is frequently used as a criterion to enrich clinical trials with patients with more significant liver disease. The performance of cT1 to make this distinction was comparable to the other non-invasive markers evaluated in this study.

Staging of fibrosis in NAFLD has been clearly shown to predict clinical outcomes [5-7] and thus is an important part of the assessment of patients with NAFLD in clinical practice and for inclusion in current late stage clinical trials. Additionally both cT1 and ELF have also been reported to have utility in predicting clinical outcomes.[39, 40] In this cohort, cT1 did not predict fibrosis defined by either Kleiner stage or collagen proportionate area whereas the other non-invasive markers assessed in this study performed in line with published work.[41, 42]

The lack of correlation between cT1 and fibrosis in this study was unexpected as previous work by Banerjee *et al*. in unselected patients[11] and by Pavlides *et al*. in patients with NAFLD[43] has shown a clear correlation between cT1 and fibrosis stage. It may be that our study is underpowered to detect this correlation, however, in our study it appears that the influence of disease activity on cT1 has hampered the ability of cT1 to detect stage differences in fibrosis; larger studies are needed to explore this further. **Figure 6** shows the heavy confounding of disease activity on fibrosis assessment in this cohort. When grouped by fibrosis stage, the only statistically significant difference in cT1 is between low and high NAFLD activity score in patients with early stage fibrosis. This identification of a group of patient with early stage fibrosis and less severe disease as graded by NAFLD activity score characterises a group of patients at low risk of progressive liver disease. As shown in **Table 8**, cT1 had comparable sensitivity, specificity and positive predictive value to ELF and liver stiffness for the exclusion of significant disease (NASH or fibrosis ≥F1) but notably greater negative predictive value suggesting that multiparametric MRI can identify patients without significant liver disease with confidence. These patients would potentially not need further investigations.

Our study was prospective and enrolled unselected consecutive patients across two sites. The statistical power is limited by recruitment volume, which may have resulted in some of the more subtle associations between variables being missed. Despite this, our independently collected data confirms the opportunities for new non-invasive biomarkers in liver disease severity assessment. Although liver biopsy remains the gold standard, as a comparator it has known limitations of sampling error and interobserver variation.[25, 44] However, in our study

we demonstrated that semi-quantitative histology scores correlated strongly with collagen proportionate area and other biomarkers performed as per previous publications.

In conclusion, multiparametric MRI with Liver*Multiscan*^{TM} has the ability to identify patients with NAFLD and to quantify steatosis and overall disease activity (by NAFLD activity score). Additionally, it stratified between healthy volunteers and patients with NAFLD as well as between patients at low risk and those at high risk of progressive liver disease. When evaluated alongside existing biomarkers, we conclude that different non-invasive tests may provide complementary diagnostic information and prove cost-effective when incorporated into clinical pathways by avoiding unnecessary liver biopsies. Multiparametric MRI was superior for the grading of steatosis, grading of NASH severity, and for excluding disease, but in this cohort was inferior for the staging of fibrosis. The potential application of multiparametric MRI in clinical and research settings thus remains of great interest and further studies in appropriate clinical populations will facilitate improved understanding of the opportunities for applying MR imaging in evaluation of liver disease.

References

1. Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73-84.
2. Dam-Larsen S, Franzmann M, Andersen IB, et al. Long term prognosis of fatty liver: risk of chronic liver disease and death. Gut 2004;53:750-5.
3. Adams LA, Lymp JF, St Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. Gastroenterology 2005;129:113-21.
4. McCullough AJ. The clinical features, diagnosis and natural history of nonalcoholic fatty liver disease. Clin Liver Dis 2004;8:521-33, viii.
5. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2015;149:389-97.e10.
6. Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. Hepatology 2015;61:1547-54.
7. Hagstrom H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017.
8. Nalbantoglu IL, Brunt EM. Role of liver biopsy in nonalcoholic fatty liver disease. World J Gastroenterol 2014;20:9026-37.
9. Rosenberg WM, Voelker M, Thiel R, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004;127:1704-13.

10. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology 2008;134:960-74.
11. Banerjee R, Pavlides M, Tunnicliffe EM, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol 2014;60:69-77.
12. National Institute for Health and Care Excellence. Nonalcoholic fatty liver disease (NAFLD): assessment and management. NICE guidelines (NG49) 2016.
13. National Institute for Health and Care Excellence. Cirrhosis in over 16s: assessment and management. NICE guidelines (NG50) 2016.
14. Banerjee R, Piechnik S, Robson M, et al. Multi-Parametric Magnetic Resonance Diagnosis & Staging of Liver Disease (US2014330106), 2014.
15. Noureddin M, Lam J, Peterson MR, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology 2013;58:1930-40.
16. Tang A, Tan J, Sun M, et al. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology 2013;267:422-31.
17. Permutt Z, Le TA, Peterson MR, et al. Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease - MRI accurately quantifies hepatic steatosis in NAFLD. Aliment Pharmacol Ther 2012;36:22-9.
18. St Pierre TG, Clark PW, Chua-anusorn W, et al. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. Blood 2005;105:855.

19. Armstrong MJ, Corbett C, Hodson J, et al. Operator training requirements and diagnostic accuracy of Fibroscan in routine clinical practice. Postgrad Med J 2013;89:685-92.

20. Boursier J, Zarski J-P, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013;57:1182-1191.

21. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317-25.

22. Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007;45:846-54.

23. Wilson M, Reynolds G, Kauppinen RA, et al. A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. Magnetic Resonance in Medicine 2011;65:1-12.

24. Wyatt J, Hubscher S, Bellamy C. Tissue pathways for liver biopsies for the investigation of medical disease and for focal lesions. Royal College of Pathologists guidelines 2014:1-29.

25. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. Hepatology 2009;49:1017-44.

26. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313-21.

27. Brunt EM, Janney CG, Di Bisceglie AM, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94:2467-74.

28. Scheuer PJ, Williams R, Muir AR. Hepatic pathology in relatives of patients with haemochromatosis. J Pathol Bacteriol 1962;84:53-64.

29. Younossi ZM, Stepanova M, Rafiq N, et al. Pathologic criteria for nonalcoholic steatohepatitis: Interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011;53:1874-1882.

30. Calvaruso V, Burroughs AK, Standish R, et al. Computer-assisted image analysis of liver collagen: relationship to Ishak scoring and hepatic venous pressure gradient. Hepatology 2009;49:1236-44.

31. 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.

32. Castera L. Non-invasive diagnosis of steatosis and fibrosis. Diabetes Metab 2008;34:674-9.

33. Fedchuk L, Nascimbeni F, Pais R, et al. Performance and limitations of steatosis biomarkers in patients with nonalcoholic fatty liver disease. Aliment Pharmacol Ther 2014;40:1209-22.

34. Festi D, Schiumerini R, Marzi L, et al. Review article: the diagnosis of non-alcoholic fatty liver disease -- availability and accuracy of non-invasive methods. Aliment Pharmacol Ther 2013;37:392-400.

35. Hamilton G, Middleton MS, Bydder M, et al. Effect of PRESS and STEAM sequences on magnetic resonance spectroscopic liver fat quantification. J Magn Reson Imaging 2009;30:145-52.

36. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012;55:2005-23.
37. Byrne CD, Targher G. Time to Replace Assessment of Liver Histology With MR-Based Imaging Tests to Assess Efficacy of Interventions for Nonalcoholic Fatty Liver Disease. Gastroenterology 2016;150:7-10.
38. Machado MV, Cortez-Pinto H. Non-invasive diagnosis of non-alcoholic fatty liver disease. A critical appraisal. J Hepatol 2013;58:1007-1019.
39. Pavlides M, Banerjee R, Sellwood J, et al. Multi-parametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. J Hepatol 2015.
40. Parkes J, Roderick P, Harris S, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. Gut 2010;59:1245-51.
41. Petta S, Wong VW, Camma C, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. Aliment Pharmacol Ther 2017.
42. Cui J, Ang B, Haufe W, et al. Comparative diagnostic accuracy of magnetic resonance elastography vs. eight clinical prediction rules for non-invasive diagnosis of advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease: a prospective study. Aliment Pharmacol Ther 2015;41:1271-80.
43. Pavlides M, Banerjee R, Tunnicliffe EM, et al. Multi-parametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int 2016.

44. Germani G, Hytiroglou P, Fotiadu A, et al. Assessment of fibrosis and cirrhosis in liver biopsies: an update. Semin Liver Dis 2011;31:82-90.

45. Wong VW-S, Vergniol J, Wong GL-H, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454-462.

Figure Legends

Figure 1: Box plots showing the relationships between non-invasive and histological assessment of steatosis. A) PDFF-Dixon and Brunt steatosis grade (n=38), B) PDFF-MRS and Brunt steatosis grade (n=43), C) Controlled Attenuation Parameter and Brunt steatosis grade (n=24). Inter-group differences were assessed with post-hoc tests with values shown in the supplementary material. CAP: controlled attenuation parameter, HV: healthy volunteer, PDFF: proton density fat fraction.

Figure 2: Receiver operating characteristic curve for the differentiation of patients with simple steatosis from those with NASH (n=47) for cT1, liver stiffness and enhanced liver fibrosis test. ELF: enhanced liver fibrosis test.

Figure 3: Box plots showing the relationships between the individual components of the NAFLD activity score and non-invasive markers of liver disease. p-values are derived from the Jonckheere–Terpstra test. NAS: NAFLD activity score, ELF: enhanced liver fibrosis test. A. cT1 showed a significant association with hepatocyte ballooning and total NAFLD activity score but not with lobular inflammation (n=50). B. Liver stiffness showed a significant association with hepatocyte ballooning, lobular inflammation and total NAFLD activity score (n=47). C. ELF score showed a significant association with hepatocyte ballooning and lobular inflammation but not total NAFLD activity score (n=50).

Figure 4: Box plots showing the relationships between non-invasive markers of liver disease and Kleiner fibrosis stage in the study cohort. cT1 did not have a statistically significant association with Kleiner fibrosis stage for patients with NAFLD. Liver stiffness (n=47), ELF, NAFLD fibrosis score, AST:ALT ratio and FIB-4 had a statistically significant association with Kleiner fibrosis stage. p-values shown were determined by the Jonckheere–Terpstra test across patients with NAFLD only and exclude healthy volunteers. HV: healthy volunteer, ELF: enhanced liver fibrosis test, NFS: NAFLD fibrosis score.

Figure 5: Scatter plots showing the relationships between non-invasive markers of liver disease and collagen proportionate area in the study cohort. cT1 did not have a statistically significant correlation with collagen proportionate area, assessed by Spearman's rank correlation. All other biomarkers had statistically significant correlations with collagen proportionate area as shown. Liver stiffness (n=47), enhanced liver fibrosis test, NAFLD fibrosis score, AST:ALT ratio, FIB-4. ELF: enhanced liver fibrosis test, NFS: NAFLD fibrosis score, CPA: collagen proportionate area.

Figure 6: Box plot showing the influence of overall disease activity (as assessed by NAFLD activity score) on cT1. In patients with low stage fibrosis there was significant difference in cT1 between those with NAFLD activity score <5 compared to those with a NAFLD activity score ≥5. This distinction had a trend towards significance in higher stage fibrosis. NAS: NAFLD activity score.

Table 1: Baseline characteristics of patients with NAFLD and healthy volunteers.

	Patients n=50	**Healthy Volunteers n=6**	**p-Value**
Age (years)	54 (18-73)	32 (23-55)	**0.011**
Male	28 (56%)	3 (50%)	1.000
Caucasian	43 (86%)	6 (100%)	1.000
BMI (kg/m^2)	33.6 ±5.1	24.0 ±2.5	**0.001**
waist to hip ratio			
Male	0.98 ±0.07	0.81 ±0.05	**0.001**
Female	0.90 ±0.06	0.72 ±0.03	**0.001**
Post-transplant	5 (10%)	n/a	-
Type 2 diabetes	26 (52%)	n/a	-
Hypertension	25 (50%)	n/a	-
Dyslipidaemia	26 (52%)	n/a	-
Smoking Status			1.000
Non-smoker	26 (58%)	4 (67%)	-
Ex-smoker	15 (30%)	2 (33%)	-
Current smoker	6 (12%)	0 (0%)	-
Consume alcohol	13 (26%)	6 (100%)	**0.001**
*UK units/week**	*8 (1-20)*	*13 (1-15)*	0.701

Data reported as mean ±SD, with p-values from t-tests; median (range), with p-values from Mann-Whitney tests; or n (%), with p-values from Fisher's exact tests, as applicable.
Bold p-values are significant at $p<0.05$
**In patients that consume alcohol*

Table 2: Liver histology characteristics of study participants

Characteristic	n	%
Kleiner Fibrosis Stage		
0	6	12%
1	10	20%
2	9	18%
3	20	40%
4	5	10%
Diagnosis		
Simple steatosis	12	24%
NASH	38	76%
Brunt Steatosis Grade		
0	0	0%
1	23	46%
2	17	34%
3	10	20%
Lobular Inflammation (NAFLD activity score)		
0	11	22%
1	23	46%
2	15	30%
3	1	2%
Hepatocyte Ballooning (NAFLD activity score)		
0	10	20%
1	15	30%
2	25	50%
Total NAFLD activity score		
0	0	0%
1-2	9	18%
3-4	16	32%
5-6	22	44%
7-8	3	6%

Table 3: Demographic, clinical and laboratory parameters showed no difference between patients with simple steatosis and NASH.

	NASH (n=38)	Simple Steatosis (n=12)	p-Value
Age (years)	54 (18-73)	46 (23-69)	0.216
Male	19 (50%)	9 (75%)	0.186
Caucasian	34 (90%)	9 (75%)	0.337
BMI (kg/m^2)	34.2 ±4.8	31.6 ±5.7	0.125
Type 2 diabetes	22 (58%)	4 (33%)	0.190
Hypertension	20 (53%)	5 (42%)	0.742
Hyperlipidaemia	21 (55%)	5 (42%)	0.514
Consume alcohol	8 (21%)	5 (42%)	0.256
*Alcohol intake (UK units/week)**	*7 (1-20)*	*12 (2-16)*	*0.831*
Bilirubin (µmol/L)	11 (4-45)	15 (5-50)	0.318
Aspartate transaminase (AST) (U/L)	38 (19-119)	41 (16-112)	0.526
Alanine transaminase (ALT) (U/L)	53 (18-153)	74 (15-176)	0.707
Alkaline phosphatase (ALP) (U/L)	94 (45-251)	76 (50-149)	0.114
Gamma-glutamyl transferase (gGT) (U/L)	78 (22-381)	58 (21-547)	0.071
Albumin (g/L)	45 ±4	46 ±4	0.477
Fasting glucose (mmol/L)	6.3 (2.8-17.3)	5.6 (4.6-11.4)	0.111
Cholesterol (mmol/L)	4.8 ±1.5	5.2 ±1.2	0.470
Triglycerides (mmol/L)	2.0 (0.7-5.8)	1.4 (0.9-2.7)	0.099
Ferritin (µg/L)	115 (10-689)	177 (45-346)	0.159
Transferrin saturation (%)	24.3 (7.5-49.6)	30.5 (14.7-43.7)	0.080
Creatinine (µmol/L)	73 (46-143)	77 (52-97)	0.225
Platelet count ($x10^9/L$)	219 ±71	197 ±37	0.172

Data reported as mean ±SD, with p-values from t-tests; median (range), with p-values from Mann-Whitney tests; or n (%), with p-values from Fisher's exact tests, as applicable.
**In patients that consume alcohol*

Table 4: cT1, liver stiffness and ELF showed significant differences between those with simple steatosis and those with NASH.

	NASH (n=38)	**simple steatosis (n=12)**	**p-Value**
cT1 (ms)	1007 ±94	907 ±120	**0.004**
Liver Stiffness (kPa)*	10.2 (4.9-27.7)	6.1 (3.6-9.1)	**<0.001**
ELF	9.3 ±1.0	7.8 ±0.8	**<0.001**
AST:ALT ratio	0.76 (0.27-1.56)	0.62 (0.34-1.07)	0.077
NAFLD fibrosis score	-0.95 ±1.64	-1.78 ±1.99	0.150
FIB-4	1.20 (0.40-5.80)	1.12 (0.38-4.61)	0.351

Data reported as mean ±SD, with p-values from t-tests or median (range), with p-values from Mann-Whitney tests, as applicable.
Bold p-values are significant at p<0.05
**Based on reliable scans only (n=47)*

Table 5: Correlations between non-invasive biomarkers and the components of the NAFLD activity score

	cT1	Liver stiffness*	ELF
Hepatocyte ballooning	Rho=0.308 (**p=0.030**)	Rho=0.450 (**p=0.002**)	Rho=0.366 (**p=0.009**)
Lobular inflammation	Rho=0.069 (p=0.635)	Rho=0.433 (**p=0.002**)	Rho=0.440 (**p=0.001**)
Total NAFLD activity score	Rho=0.514 (**p<0.001**)	Rho=0.419 (**p=0.003**)	Rho=0.460 (**p=0.001**)

Correlations assessed with Spearman's Rho
Bold p-values are significant at p<0.05
**Based on reliable scans only (n=47)*

Table 6: cT1, liver stiffness and ELF show significant differences between high risk patients, low risk patients and healthy volunteers.

	High risk patients* (n=40)	Low risk Patients** (n=10)	Healthy volunteers (n=6)	p-Value
cT1 (ms)	1007 ±93	890 ±122	790 ±42	**<0.001**
Liver Stiffness (kPa)†	9.9 (4.9-27.7)	6.1 (3.6-9.1)	4.5 (3.6-6.8)	**<0.001**
ELF	9.2 ±1.0	7.7 ±0.8	7.9 ±0.3	**<0.001**
AST:ALT ratio	0.76 (0.27-1.56)	0.65 (0.34-1.07)	1.02 (0.86-1.67)	**0.018**
NAFLD fibrosis score	-1.05 ±1.66	-1.54 ±2.09	-2.90 ±0.62	**0.047**
FIB-4	1.15 (0.38-5.80)	1.23 (0.53-4.61)	0.87 (0.55-0.98)	0.158

Data reported as mean ±SD, with p-values from one-way ANOVA or median (range), with p-values from Kruskal-Wallis tests, as applicable.
Bold p-values are significant at p<0.05
** Patients with either NASH or >F1 fibrosis*
*** Patients with simple steatosis and ≤F1 fibrosis*
† Based on reliable scans only (n=53)

Table 7: cT1, liver stiffness and ELF and stratification of low and high risk patients

	Low risk patients* (n=10) ***vs*** **High risk patients** (n=37)**	**Healthy volunteers (n=6)** ***vs*** **All patients (n=47)**	**Healthy volunteers and low risk patients (n=16)** ***vs*** **High risk patients (n=37)**
cT1	**0.73 (0.53-0.93)**	**0.93 (0.86-1.00)**	**0.83 (0.69-0.96)**
liver stiffness†	**0.82 (0.69-0.94)**	**0.89 (0.77-1.00)**	**0.86 (0.76-0.96)**
ELF	**0.89 (0.80-0.99)**	**0.81 (0.69-0.92)**	**0.89 (0.81-0.98)**
AST:ALT ratio	0.64 (0.45-0.84)	**0.82 (0.67-0.97)**††	0.52 (0.34-0.70)††
NAFLD fibrosis score	0.55 (0.32-0.77)	**0.79 (0.66-0.91)**	0.64 (0.47-0.81)
FIB-4	0.51 (0.31-0.71)	0.72 (0.59-0.85)	0.59 (0.43-0.75)

Data reported as AUROC (95% CI)
Bold values were significant at p<0.05
** Patients with simple steatosis and ≤F1 fibrosis*
*** Patients with either NASH or >F1 fibrosis*
† Based on reliable scans only
††Inverse relationship, i.e. AST:ALT ratio was higher in the healthy volunteer group

Table 8: Sensitivity, specificity, positive predictive value and negative predictive value at commonly accepted cut-off values for the differentiation of low and high risk patients.

		AUROC (95% CI)	Cut-off	Sensitivity	Specificity	PPV	NPV
Differentiation of low* and high risk** patients	cT1	0.73 (0.53-0.93)	822 ms[11]	97.5%	40.0%	86.7%	80.0%
			875 ms[39]	97.5%	50.0%	88.6%	83.3%
	LS	0.82 (0.69-0.94)	5.8 kPa[45]	89.2%	30.0%	82.5%	42.9%
			7.0 kPa[45]	75.7%	60.0%	87.5%	40.0%
			7.9 kPa[45]	64.9%	70.0%	88.9%	35.0%
			9.0 kPa[45]	62.2%	90.0%	95.8%	39.1%
	ELF	0.89 (0.80-0.99)	7.7[9]	92.5%	40.0%	86.0%	57.1%
			9.8[9]	30.0%	100.0%	100.0%	26.3%
Differentiation healthy volunteers and patients	cT1	0.93 (0.86-1.00)	822 ms	90.0%	83.3%	97.8%	50.0%
			875 ms	88.0%	100.0%	100.0%	50.0%
	LS	0.89 (0.77-1.00)	5.8 kPa	85.1%	66.7%	95.2%	36.4%
			7.0 kPa	68.1%	100.0%	100.0%	28.6%
	ELF	0.81 (0.69-0.92)	7.7	86.0%	16.7%	89.6%	12.5%
			9.8	24.0%	100.0%	100.0%	13.6%

** Patients with simple steatosis and ≤F1 fibrosis*

*** Patients with either NASH or >F1 fibrosis*

LS: liver stiffnes, PPV: positive predictive value, NPV: negative predictive value

Table 9: Cost and effectiveness of diagnostic pathways for patients with suspected NAFLD

		Overall outcome			Per 1000 patients		Total cost
	Cut-off	**Correct**	**Incorrect**	**Fail**	**Biopsies avoided**	**Total cost (£)**	**Cost Saving vs Biopsy (£)**	**Per correct diagnosis (£)**
cT1	822 ms	82.9%	12.1%	5.0%	381.9	538,345	101,265	649.57
	875 ms	88.3%	6.7%	5.0%	458.4	489,392	150,218	554.26
liver stiffness	5.8 kPa	63.6%	18.4%	18.0%	297.2	517,530	122,080	814.16
	7.0 kPa	66.6%	15.4%	18.0%	491.6	393,146	246,464	590.14
ELF	7.7	57.4%	42.6%	0.0%	151.1	654,010	-14,400	1138.43
	9.8	55.3%	44.7%	0.0%	858.9	201,322	438,288	363.97
liver stiffness (5.8 kPa) + cT1	822 ms	81.4%	15.0%	3.5%	734.6	338,260	301,350	415.37
	875 ms	83.4%	13.1%	3.5%	722.7	345,851	293,759	414.60
liver stiffness (7.0 kPa) + cT1	822 ms	76.8%	20.7%	2.5%	841.1	242,309	397,301	315.60
	875 ms	77.1%	20.3%	2.5%	848.7	237,488	402,122	307.92

Assumed costs: Biopsy (£639.91), MRI (£143), liver stiffness (£68), ELF (£111.06). Overall outcome rates are derived from the data in **Table 8***, with the rate of correct and incorrect outcomes being the sum of the true positives and true negatives (correct), or the false positives and false negatives (incorrect), respectively, for the pathway. Costs for each diagnostic pathway include all predicted test and biopsy costs for a cohort of 1,000 patients.*

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	7
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	9
	4	Study objectives and hypotheses	10
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	12
Participants	**6**	Eligibility criteria	12
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	12
	8	Where and when potentially eligible participants were identified (setting, location and dates)	12
	9	Whether participants formed a consecutive, random or convenience series	12
Test methods	**10a**	Index test, in sufficient detail to allow replication	13
	10b	Reference standard, in sufficient detail to allow replication	15/16
	11	Rationale for choosing the reference standard (if alternatives exist)	No alternatives
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	**Table 8**
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	15/16

	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	15
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	15
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	16
	15	How indeterminate index test or reference standard results were handled	n/a
	16	How missing data on the index test and reference standard were handled	18
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	n/a
	18	Intended sample size and how it was determined	18
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Supplementary material
	20	Baseline demographic and clinical characteristics of participants	18 / **Table 1**
	21a	Distribution of severity of disease in those with the target condition	**Table 2**
	21b	Distribution of alternative diagnoses in those without the target condition	n/a
	22	Time interval and any clinical interventions between index test and reference standard	13
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	**Table 4**

Table 4

	cT1	Liver stiffness*	ELF
Hepatocyte ballooning	Rho=0.308 (**p=0.030**)	Rho=0.450 (**p=0.002**)	Rho=0.366 (**p=0.009**)
Lobular inflammation	Rho=0.069 (p=0.635)	Rho=0.433 (**p=0.002**)	Rho=0.440 (**p=0.001**)
Total NAFLD activity score	Rho=0.514 (**p<0.001**)	Rho=0.419 (**p=0.003**)	Rho=0.460 (**p=0.001**)

Correlations assessed with Spearman's Rho
Bold p-values are significant at p<0.05
**Based on reliable scans only (n=47)*

Table 6 **Table 7**

	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	**Table 8**
	25	Any adverse events from performing the index test or the reference standard	n/a
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	25-30
	27	Implications for practice, including the intended use and clinical role of the index test	25-30
OTHER INFORMATION			
	28	Registration number and name of registry	12
	29	Where the full study protocol can be accessed	Supplementary material
	30	Sources of funding and other support; role of funders	3

1. Authors' declaration of personal interests:

(i) Peter J Eddowes, Phillip N Newsome and Gideon M Hirschfield are supported by the National Institute of Health Research Birmingham Biomedical Research Centre. Jonathan A Fallowfield is supported by a NHS Research Scotland/Universities Senior

(ii) Catherine J Kelly, Amy H Herlihy and Matthew D Kelly are employees of Perspectum Diagnostics ltd.

2. Declaration of funding interests:

(i) This study was funded in full by Innovate-UK, grant number 101679.

Supplementary Material to: Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease

MRI Sequence Parameters

T1 Mapping

A T1 relaxation time map was acquired using the Shortened Modified Look Locker Inversion recovery (shMOLLI) sequence. Shimming was performed across the volume of the liver based on a GRE field map acquired during a single expiration breath hold. The ShMOLLI sequence samples the T1 recovery curve using single-shot steady state free precession (SSFP) acquisitions using the following parameters. TR 2.14ms, TE 1.07ms, flip angle of 35°, field-of-view optimised per patient in the range 400-450mm. Acquisition matrix 192x134-160, depending on patient, with GRAPPA acceleration of 2 with 24 reference lines, yielding a typical interpolated voxel size 0.9 x 0.9 x 8mm. Images were acquired 340ms after the ECG R-wave with delay shortened in those with a heart rate >75beats per minute.

Proton magnetic resonance spectroscopy

^{1}H Magnetic Resonance Spectroscopy was performed in a 30x30x30 mm voxel in the right lobe of the liver using a stimulated echo sequence (STEAM), TR/TE=3000/20 ms. 5 measurements each of water-suppressed and non-water-suppressed data were acquired across 2 expiration breath holds. Measurements were automatically phase and frequency corrected individually using TARQUIN.[1] The 5 FID's were then averaged before automated fitting in

TARQUIN using customised liver-specific metabolite basis sets for water suppressed and non-suppressed data.

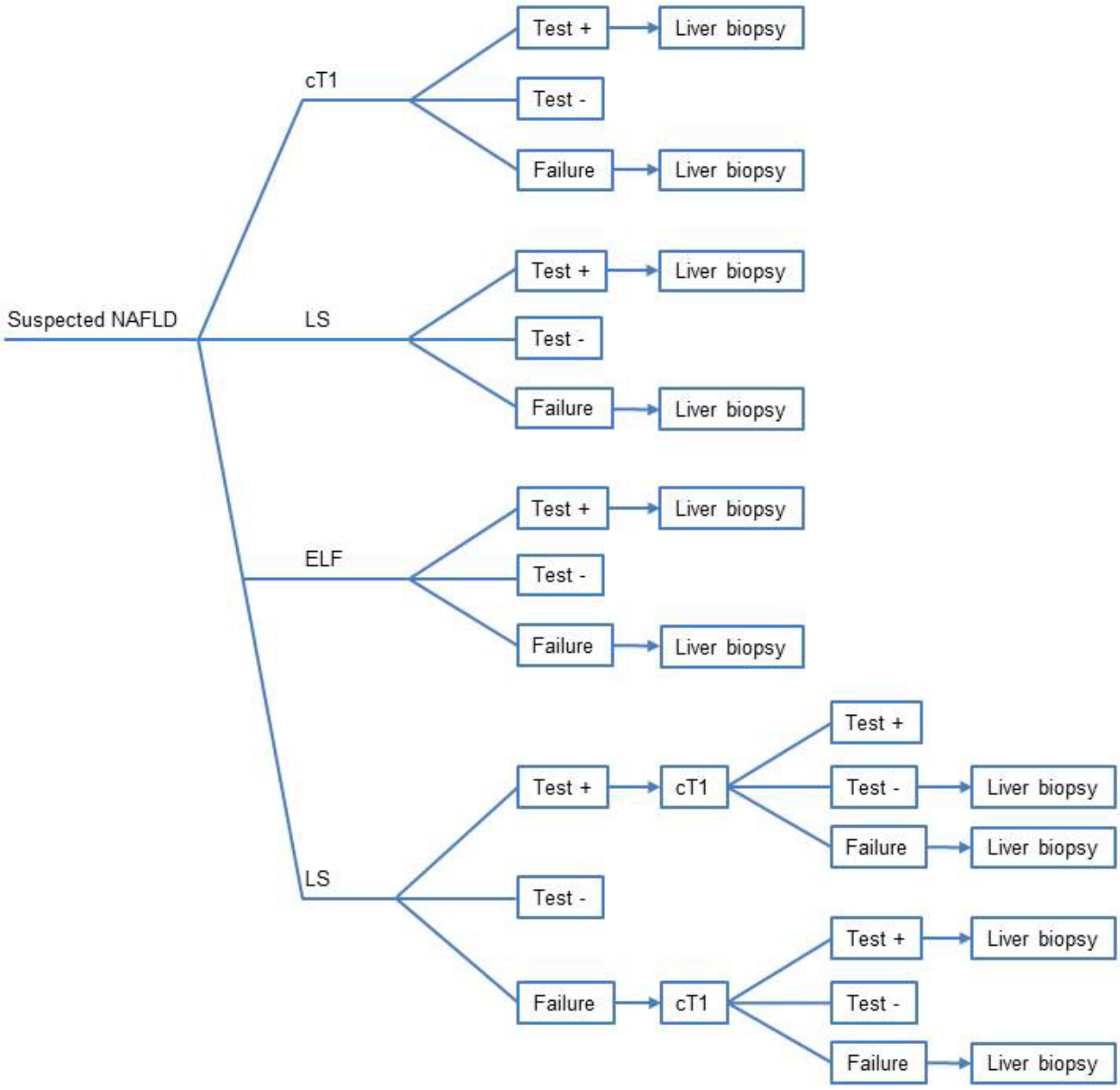
T2 mapping*

A multi-gradient-echo acquisition with RF spoiling is used to calculate a T2* map of the liver. The same field-of-view as in the T1 mapping sequence is used, with a matrix size of 192x128-160, depending on patient, slice thickness of 6mm and 2x GRAPPA acceleration, with the same delay after the R-wave before acquisition. The image is acquired in nine 4 segments with a TR of 26.5ms and flip angle of 20°. Echo times are selected as far as possible such that the signals from fat and water are in phase (TE = 2.46, 7.38, 12.30, 17.22 and 22.14 ms).

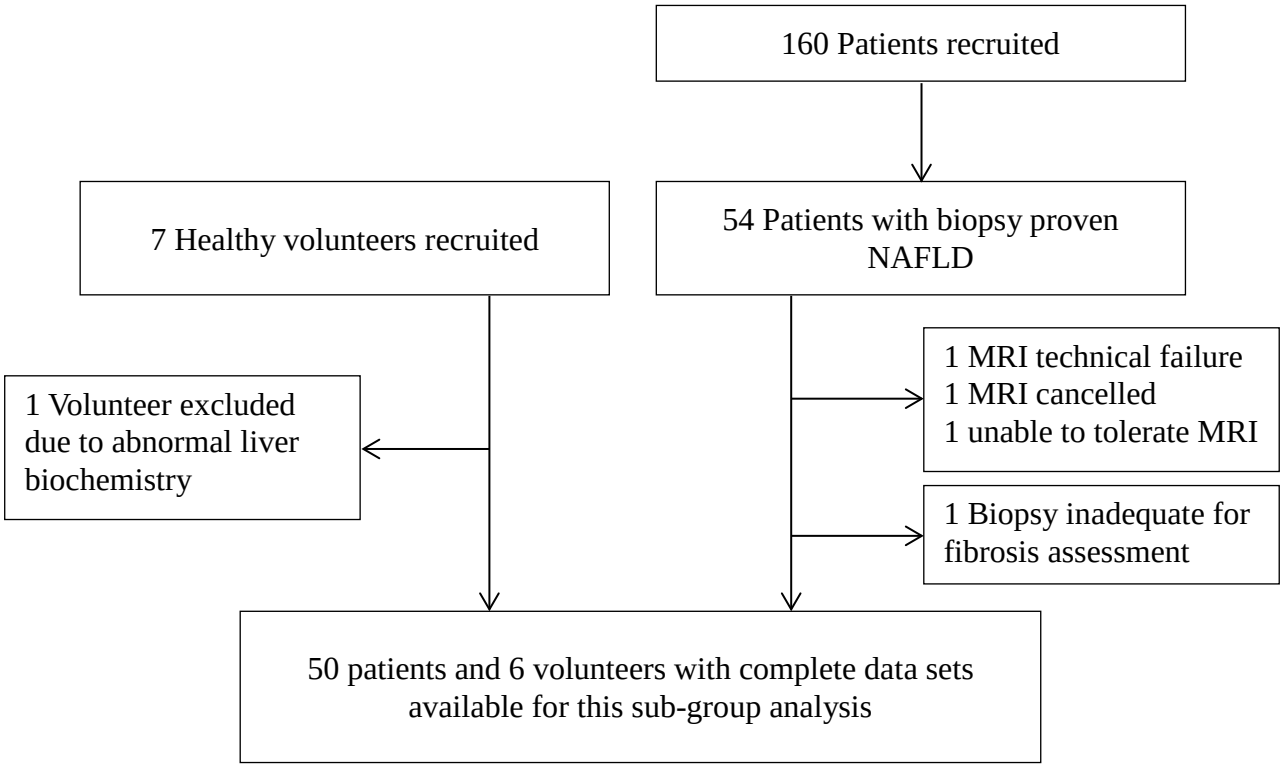
Modified Dixon Sequence

A further RF-spoiled multi-gradient-echo sequence was used to calculate the percentage fat content of liver tissue. Voxel size and acquisition parameters matched those of the T2* mapping sequence, but with echo times chosen to give alternating in- and out-of-phase fat and water signals. Calculation of fat fraction using this sequence is well described in the literature.[2]

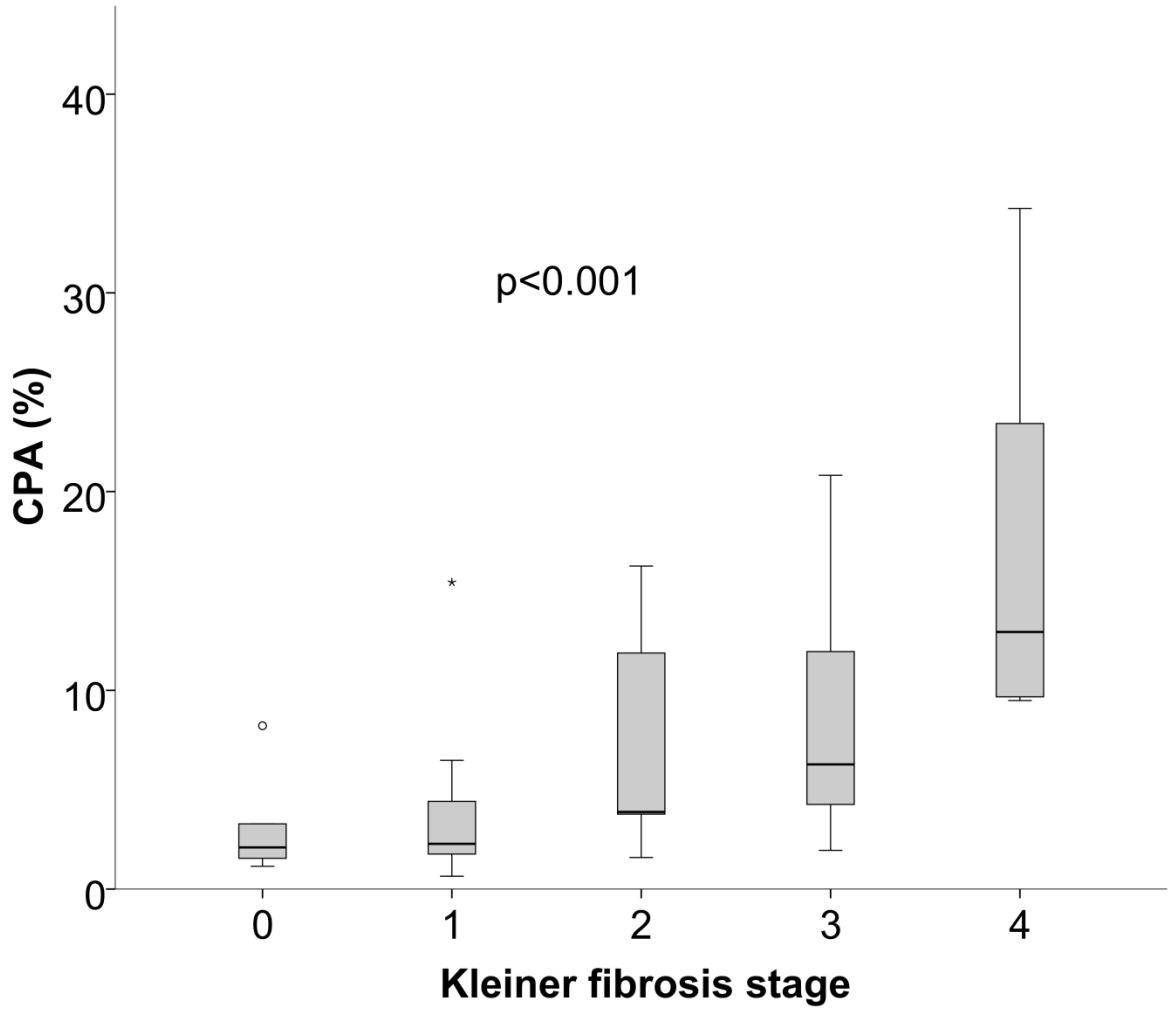
Supplementary Figures



Supplementary figure 1: Risk stratification pathway of patients with suspected NAFLD

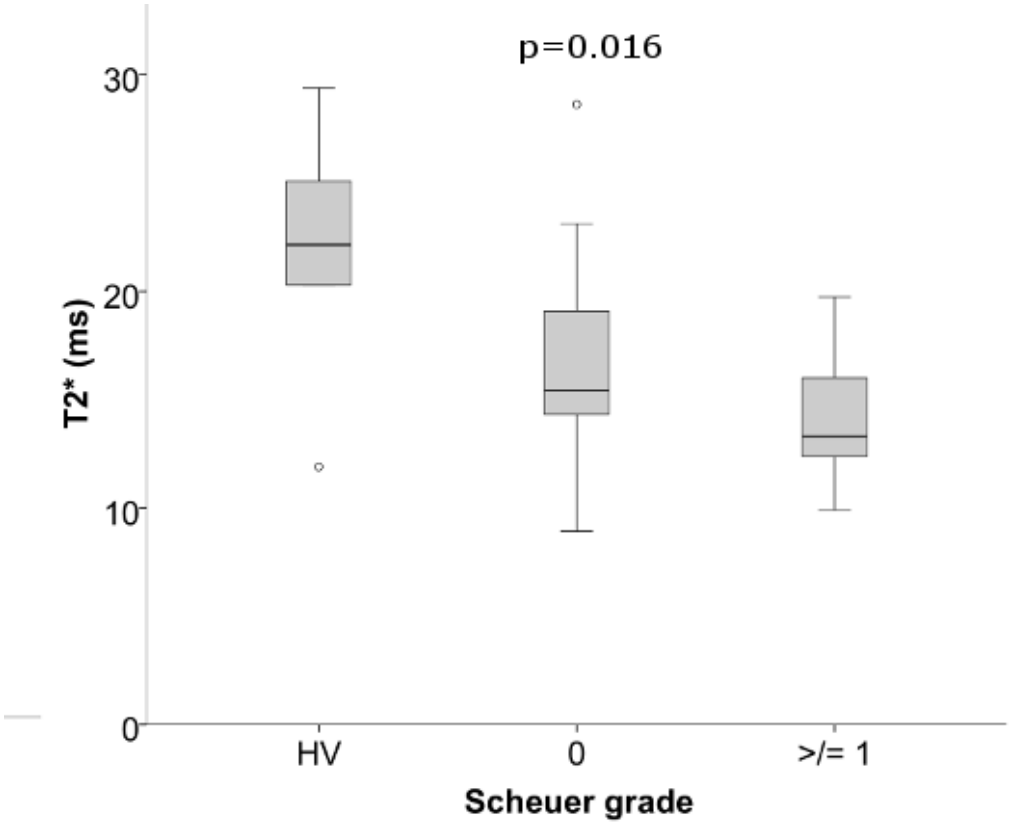


Supplementary figure 2: Study participant flowchart - 4 patients and 1 healthy volunteer were excluded from the final analysis.



Supplementary figure 3: The association of CPA and Kleiner stage in the study cohort.

P<0.001 by the Jonckheere-Terpstra test.



Supplementary figure 4: Relationship between Scheuer siderosis grade and T2*.

Supplementary data

Indication for liver biopsy

32 patients were diagnosed with NAFLD based on clinical assessment of risk factors, exclusion of other aetiologies and ultrasound findings consistent with hepatic steatosis. These patients had inconclusive non-invasive assessment of fibrosis and underwent liver biopsy for staging of fibrosis.

The remaining 18 patients were biopsied to make a diagnosis. These patients had undergone a range of non-invasive tests without a firm diagnosis being made.

Steatosis assessment

Comparison	Brunt vs. PDFF-Dixon	Brunt vs. PDFF-MRS	Brunt vs. CAP
Overall	**<0.001**	**<0.001**	**0.002**
HV vs. 1	0.211	0.312	0.090
HV vs. 2	**<0.001**	**<0.001**	**<0.001**
HV vs. 3	**<0.001**	**<0.001**	0.277
1 vs. 2	**0.011**	**0.004**	0.376
1 vs. 3	**0.002**	**<0.001**	1.000
2 vs. 3	1.000	1.000	1.000

Differences between groups in assessment of liver fat to accompany Figure 1. Overall significance by the Kruskal-Wallis test. Inter-group differences were assessed with post-hoc tests.

Grading of siderosis by multiparametric MRI

Seven out of 50 (14%) patients had grade 1 siderosis on biopsy and only 1/50 (2%) patients had grade 2 siderosis. Mean T2* in healthy volunteers, patients without siderosis on biopsy and patients with siderosis on biopsy (Scheuer grade ≥1) was 21.8 (±5.8), 16.7 (±3.7) and 14.1 (±3.1) milliseconds (ms) respectively (p=0.016) (Supplementary figure 4). AUROC for the differentiation of patients without and patients with siderosis on biopsy was 0.705 (0.498-0.912).

Scheuer Siderosis Grade		
0	42	84%
1	7	14%
2	1	2%
3	0	0%

Distribution of Scheuer siderosis grade on biopsy.

Comparison	**Scheuer vs. T2**
Overall	**0.016**
1 vs. 0	0.267
1 vs. HV	**0.012**
0 vs. HV	0.116

Differences between groups in assessment of liver iron to accompany Supplementary figure 4. Overall significance by the Kruskal-Wallis test. Inter-group differences were assessed with post-hoc tests.

Supplementary References

1. Wilson M, Reynolds G, Kauppinen RA, et al. A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. Magnetic Resonance in Medicine 2011;65:1-12.

2. Reeder SB, Cruite I, Hamilton G, et al. Quantitative Assessment of Liver Fat with Magnetic Resonance Imaging and Spectroscopy. J Magn Reson Imaging 2011;34:spcone.