External validation of the fatty liver index and lipid accumulation product indices, using 1H-magnetic resonance spectroscopy, to identify hepatic steatosis in healthy controls and obese, insulin-resistant individuals

Daniel J Cuthbertson1,2, Martin O Weickert3,4, Daniel Lythgoe5, Victoria S Sprung1,2, Rebecca Dobson1,2, Fariba Shoajee-Moradie6, Margot Umpbley6, Andreas F H Pfeiffer7,8, E Louise Thomas9, Jimmy D Bell9, Helen Jones10 and Graham J Kemp11

1Department of Obesity and Endocrinology, Clinical Sciences Centre, University Hospital Aintree, Liverpool, UK, 2Department of Obesity and Endocrinology, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK, 3Warwickshire Institute for the Study of Diabetes, Endocrinology and Metabolism, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK, 4Division of Metabolic and Vascular Health, Warwick Medical School, University of Warwick, Warwick, UK, 5Cancer Research UK Liverpool Cancer Trials Unit, University of Liverpool, Liverpool, UK, 6Department of Diabetes and Metabolic Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK, 7Department of Clinical Nutrition, German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany, 8The Department of Endocrinology, Diabetes and Nutrition, Charité-University-Medicine Berlin, Berlin, Germany, 9Metabolic and Molecular Imaging Group, Imperial College London, MRC Clinical Sciences Centre, London, UK, 10Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK and 11Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, Magnetic Resonance and Image Analysis Research Centre, University of Liverpool, Liverpool, UK

Abstract

Background and aims: Simple clinical algorithms including the fatty liver index (FLI) and lipid accumulation product (LAP) have been developed as surrogate markers for non-alcoholic fatty liver disease (NAFLD), constructed using (semi-quantitative) ultrasonography. This study aimed to validate FLI and LAP as measures of hepatic steatosis, as determined quantitatively by proton magnetic resonance spectroscopy (1H-MRS).

Methods: Data were collected from 168 patients with NAFLD and 168 controls who had undergone clinical, biochemical and anthropometric assessment. Values of FLI and LAP were determined and assessed both as predictors of the presence of hepatic steatosis (liver fat > 5.5%) and of actual liver fat content, as measured by 1H-MRS. The discriminative ability of FLI and LAP was estimated using the area under the receiver operator characteristic curve (AUROC). As FLI can also be interpreted as a predictive probability of hepatic steatosis, we assessed how well calibrated it was in our cohort. Linear regression with prediction intervals was used to assess the ability of FLI and LAP to predict liver fat content. Further validation was provided in 54 patients with type 2 diabetes mellitus.

Results: FLI, LAP and alanine transferase discriminated between patients with and without steatosis with an AUROC of 0.79 (IQR = 0.74, 0.84), 0.78 (IQR = 0.72, 0.83) and 0.83 (IQR = 0.79, 0.88) respectively although could not quantitatively predict liver fat. Additionally, the algorithms accurately matched the observed percentages of patients with hepatic steatosis in our cohort.
Conclusions: FLI and LAP may be used to identify patients with hepatic steatosis clinically or for research purposes but could not predict liver fat content.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is increasingly recognised as a major public health concern, being highly prevalent in the general population, particularly in individuals with features of the metabolic syndrome (1). NAFLD encompasses a disease spectrum, ranging from simple steatosis, through to an inflammatory state (non-alcoholic steatohepatitis, NASH) and culminating in fibrosis and liver cirrhosis (2). In addition to the known association with liver-related morbidity and mortality, NAFLD patients have an increased risk of type 2 diabetes mellitus (T2DM) and cardiovascular disease (3, 4).

Liver transaminases and abdominal ultrasonography are insensitive in detecting NAFLD, with liver function tests normal in up to 79% of patients and ultrasonography requiring a moderately high liver fat for NAFLD to be recognised (5). Only histological examination or proton magnetic resonance spectroscopy (1H-MRS) can quantitatively assess liver fat (more exactly, hepatocellular lipid content) (3, 4).

Non-invasive algorithms based on metabolic and anthropometric variables, such as the fatty liver index (FLI), lipid accumulation product (LAP), the hepatic steatosis index (HIS) and the Finnish Diabetes Risk Score (FINDRISC) (6, 7, 8, 9, 10, 11), have been used as a screening test for hepatic steatosis and identify potential patients for further clinical investigation or for epidemiologic studies. They have been applied in various clinical populations to assess prevalence of NAFLD (12) and to provide prognostic information about incident risk of metabolic syndrome and type 2 diabetes, cardiovascular disease and risk of mortality in various sub-groups (13, 14, 15, 16, 17, 18).

The original studies to propose FLI and LAP were validated using ultrasonography (2, 7, 19), as well as the SteatoTest, an alternative biochemical surrogate marker of liver steatosis (20). Liver biopsy is an invasive method. The only study to validate FLI using 1H-MRS, which is supposed to be the next best method as compared with liver biopsy, involved only 25 subjects, with the results suggesting a nonlinear relationship between FLI and hepatocellular lipid content (21).

This study aimed to evaluate the ability of FLI and the LAP to discriminate between patients with and without hepatic steatosis, based on simple clinical and biochemical variables; and to evaluate their ability to predict liver fat content based on our non-invasive measurement of liver fat by 1H-MRS. Here, we combined data from several large cohorts of participants with detailed characterisation of clinical, metabolic and anthropometric parameters, in whom 1H-MRS measurement of liver fat had been performed for several different research projects.

Methods

Study participants

We analysed data from participants recruited into human metabolic studies from four research centres (University of Liverpool, University of Surrey, Charite University Berlin and German Institute of Human Nutrition, Potsdam-Rehbruecke) from primary and secondary care. We recruited healthy controls and individuals with components of the metabolic syndrome including being overweight/obese (BMI > 25 kg/m²), with a waist circumference > 80 cm in females and > 94 cm in males, and with at least one additional feature of the metabolic syndrome according to International Diabetes Federation criteria (22). Clinical characteristics of the cohorts have been presented previously in detail (23, 24). All participants gave written informed consent, and ethical approval was obtained from the respective local ethics committee.

Exclusion criteria

We excluded participants with a history of type 1 diabetes mellitus, pregnancy, any significant history of endocrine, cardiovascular, renal or hepatic disease and standard MR contraindications. Other causes of chronic liver disease
were excluded by taking a careful alcohol and drug history
and performing an autoimmune liver screen and hepatitis
serology. We excluded males with alcohol intake > 21
units per week and females with > 14 units per week
and any individuals with a history of alcohol excess.

Anthropometric assessments
Trained physicians performed all anthropometry measure-
ments. BMI was calculated in kg/m²; waist circumference
was measured midway between the lower rib margin and
the iliac crest.

Biochemical assessment
All participants underwent a biochemical assessment, and
fasting triglycerides, alanine transferase (ALT) and
gamma-glutamyl transferase (GGT) were measured. Rout-
tine laboratory markers were measured from venous blood
samples using standard methods in the research labora-
tories of respective centres (University Hospital Aintree,
Liverpool, Royal Surrey County Hospital and in the
German Institute of Human Nutrition).

Proton magnetic resonance spectroscopy
All ¹H-MRS measurements of liver fat were performed in
the fasting state. Lipid content in the liver was measured by
localised ¹H-MRS, using an Intera 1.5T Achieva (Philips
Medical Systems, Best, The Netherlands) for the Surrey
participants (25), a Magnetom 1.5T Symphony MR (Sie-
mens Healthcare, Erlangen, Germany) for the Liverpool
participants (26), and a Magnetom 1.5T Avanto (Siemens
Healthcare) for the German participants. Single voxel
spectroscopy was used: for the German studies, STEAM
with VOI 30 x 30 x 30 mm, 32 acquisitions, TR = 4000 ms,
TM = 15 ms, TE = 10 ms; for the UK studies, PRESS with
VOI 20 x 20 x 20 mm (three voxels, results averaged), 64
acquisitions, TR = 1500 ms, TE 135 ms (25). Spectra were
quantified using the AMARES algorithm included in the
jMRUI software package, incorporating standard prior
knowledge. For the German studies, signal integrals of
water (H2O at 4.8 ppm) and lipids (CH2 and CH3 at 1.25
and 0.95 ppm) were quantified manually in fixed
frequency intervals (water: 3.1–6.2 ppm, lipids:
0.5–1.8 ppm); for the UK studies, these signal amplitudes
were obtained directly from the AMARES fit. Liver fat is
expressed as percentage of CH2 lipid signal amplitude
relative to water signal amplitude after correcting for T1
and T2 (25). Liver fat content (%) was quantified but also
coded ordinally as none (≤5.5%) or present (> 5.5%).

Calculations
FLI was calculated using BMI (kg/m²), serum triglyceride
(mg/dl) and GGT (U/l) concentrations and waist circum-
ference (cm) according to Bedogni et al. (7) to obtain
a score between 0 and 100:

\[
FLI = \frac{\exp(\eta)}{1 + \exp(\eta)} \times 100,
\]

where,

\[
\eta = 0.953 \times \ln(\text{triglycerides}) + 0.139 \times \text{BMI} + 0.718 \\
\times \ln(\text{ggt}) + 0.053 \times \text{waist circumference} - 15.754.
\]

LAP was calculated using serum triglycerides (mmol/l)
and waist circumference (cm) using sex-specific calcu-
lations (9):

\[
\begin{align*}
\text{LAP}_{\text{men}} &= (\text{waist circumference} - 65) \times \text{triglycerides} \\
\text{LAP}_{\text{women}} &= (\text{waist circumference} - 58) \times \text{triglycerides}.
\end{align*}
\]

Statistical analysis
Baseline demographic variables are reported as means and
S.D. or median and interquartile range depending on their
distribution. Distributional assumptions were assessed
using Q-Q plots. Categorical variables are reported as
frequencies and percentages. Statistical comparisons of
patients with and without hepatic steatosis were under-
taken for all demographic variables; \( \chi^2 \) tests were used for
categorical variables and unpaired \( t \)-tests or Mann–
Whitney \( U \) tests for continuous variables, depending on
whether relevant distributional assumptions were met.

To assess the ability of a variable to discriminate
between patients with and without hepatic steatosis,
receiver operator characteristic (ROC) curves were con-
structed for FLI, LAP, BMI, waist circumference and ALT.
In addition, for FLI and LAP, we measured a number of
other diagnostic statistics: the sensitivity, specificity,
positive likelihood ratio and negative likelihood ratio at
various cut-points.

As FLI can also be interpreted as a predictive
probability for hepatic steatosis, we assessed how well-
calibrated FLI was in our patient cohort using a calibration
plot (27). To do so, we grouped patients into deciles based
on measured FLI and within each decile calculated the
proportion of patients with hepatic steatosis. If a variable
is well calibrated, the observed percentages should be close
to the ‘line of equality’, which represents perfect cali-
bration: roughly 50% of patients should have hepatic
steatosis at an FLI of around 50 etc.
Linear regression with 95% prediction intervals (PI) was used to determine whether liver fat content could be predicted using FLI or LAP alone. PIs are to be interpreted as the range of values for liver fat we would assign to a new patient with a given FLI/LAP with 95% ‘confidence’. Linear regression assumptions were assessed using plots of residuals vs fitted values and Q-Q plots of the residuals. It was necessary to logarithmically transform both liver fat and LAP for these purposes; the linear regression line and PI, while linear on the log scale, are nonlinear, non-constant and asymmetric on the original scale. For simplicity, we evaluated their predictive performance at the mean FLI/LAP.

**Results**

**Characteristics of the participants without T2DM**

Table 1 reports the clinical, biochemical and anthropometric characteristics of the participants (178 males (53%)

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>Hepatic steatosis</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>168 (50%)</td>
<td>168 (50%)</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>Male</td>
<td>61 (34%)</td>
<td>116 (65%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>107 (67%)</td>
<td>52 (33%)</td>
<td></td>
</tr>
<tr>
<td>Liver fat (%)</td>
<td>1.8 (1.0, 3.1)</td>
<td>16.6 (9.1, 30.7)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Age (years)</td>
<td>48.6 (11.9)</td>
<td>50.3 (10.9)</td>
<td>0.19</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>83.3 (72.5, 93.2)</td>
<td>93.4 (84.6, 104.6)</td>
<td>0.29</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>29.2 (26.7, 32.3)</td>
<td>31.2 (28.9, 33.9)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Waist (cm)</td>
<td>98.4 (89.9, 104.8)</td>
<td>105.8 (100.5, 113.0)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Fasting glucose (mmol/l)</td>
<td>4.80 (4.50, 5.10)</td>
<td>5.00 (4.70, 5.40)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cholesterol (mmol/l)</td>
<td>5.00 (4.55, 5.80)</td>
<td>5.50 (4.87, 6.02)</td>
<td>0.0135</td>
</tr>
<tr>
<td>Triglycerides (mmol/l)</td>
<td>1.01 (0.80, 1.30)</td>
<td>1.60 (1.11, 2.32)</td>
<td>0.0001</td>
</tr>
<tr>
<td>HDL (mmol/l)</td>
<td>1.40 (1.20, 1.68)</td>
<td>1.20 (1.05, 1.42)</td>
<td>0.0001</td>
</tr>
<tr>
<td>LDL (mmol/l)</td>
<td>3.12 (2.64, 3.80)</td>
<td>3.30 (2.80, 3.92)</td>
<td>0.14</td>
</tr>
<tr>
<td>Chol:LDL ratio</td>
<td>3.67 (3.00, 4.24)</td>
<td>4.29 (3.81, 5.26)</td>
<td>0.0001</td>
</tr>
<tr>
<td>AST (U/l)</td>
<td>21 (18, 24)</td>
<td>28 (23, 35)</td>
<td>0.0001</td>
</tr>
<tr>
<td>ALT (U/l)</td>
<td>19 (16, 26)</td>
<td>37 (26, 56)</td>
<td>0.0001</td>
</tr>
<tr>
<td>GGT (U/l)</td>
<td>20 (13, 29)</td>
<td>37 (21, 58)</td>
<td>0.0001</td>
</tr>
<tr>
<td>FLI</td>
<td>18.8 (8.0, 37.6)</td>
<td>56.2 (31.3, 72.8)</td>
<td>0.0001</td>
</tr>
<tr>
<td>LAP</td>
<td>40.0 (25.1, 53.6)</td>
<td>72.8 (47.5, 99.2)</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Categorical variables reported as frequency (%) and continuous variables as mean (s.d.) or median (IQR).

**ROC curves**

ROC curves were constructed and area under the ROC curve (AUROC) with NAFLD defined as liver fat >5.5% on MRS (Fig. 2). Both LAP and FLI were able to discriminate between patients with and without hepatic steatosis. The AUROC for LAP was 0.78 (IQR=0.72, 0.83) and for FLI was 0.79 (IQR=0.74, 0.84). There was no evidence that AUROC for LAP and FLI differed (P=0.49). There was no significant difference in AUROC for either FLI or LAP between males and females (Table 2).

We also considered the AUROC for BMI, waist circumference, ALT, triglycerides and GGT, which were 0.64 (IQR=0.58, 0.70), 0.73 (IQR=0.67, 0.79), 0.83 (IQR=0.79, 0.88), 0.74 (IQR=0.69, 0.79) and 0.73 (IQR=0.67, 0.78), respectively. We conducted exploratory pairwise comparisons between all seven of these variables and found evidence that both FLI and LAP were superior to waist circumferences and BMI. FLI was also superior to GGT (P=0.03) but not to triglycerides (P=0.10).
The reverse was true for LAP, which was superior to triglycerides \((P=0.01)\) but not GGT \((P=0.12)\). Interestingly, ALT had a similar AUROC to both FLI and LAP.

### Sensitivity, specificity, positive likelihood ratios and negative likelihood ratio of FLI and LAP

Table 3 gives the sensitivity, specificity, positive likelihood ratios and negative likelihood ratios for a range of ten-unit intervals for FLI. The intervals chosen were those used by Bedogni et al. (7). A cut-off of FLI \(\geq 10\) gives a sensitivity of 95% and a LR− of 0.15, i.e. an individual without hepatic steatosis is around seven times more likely to have an FLI < 10. A cut-off of FLI \(\geq 60\) gives a specificity of 91% and a LR+ of 5.10, i.e. an individual with hepatic steatosis is around five times more likely to have an FLI \(\geq 60\). Using these cut-offs, 42% of patients would have avoided the need for a more definitive test (with 77% being classified correctly); 58% would have had an indeterminate classification (an FLI value within the cut-offs).

Table 3 gives the sensitivity, specificity, positive likelihood ratios and negative likelihood ratios for a range of ten-unit intervals for LAP. A cut-off of LAP \(\geq 20\) gives a sensitivity of 99% and LR− of 0.08, i.e. an individual without hepatic steatosis is around ten times more likely to have a LAP < 20. A cut-off of LAP \(\geq 80\) has a specificity of 94% and LR+ of 4.93, i.e. an individual with hepatic steatosis is around five times more likely to have a LAP \(\geq 80\). Using these cut-offs, 35% of patients would have avoided the need for a more definitive test (with 86% being classified correctly); 65% would have had an indeterminate classification.

We report these cut-offs in particular because they yield a LR+ > 5 and a LR− < 0.2 and therefore might be used as reasonable ‘rule-in’ and ‘rule-out’ criteria respectively.

### Calibration of FLI

FLI provides a ‘predicted probability’ of a patient having hepatic steatosis. The calibration plot in Fig. 3 addresses how satisfactory these predicted probabilities were in our cohort. We have established that FLI discriminates between patients with and without hepatic steatosis (using AUROC analysis), and this is evidenced by the horizontal separation of the two clouds of points, clustered at low FLI values for those without steatosis, and at higher FLI values for those with hepatic steatosis.

We further determined how closely an individual’s actual FLI corresponds to their probability of having hepatic steatosis and this relates to calibration. We assessed calibration of FLI by grouping patients according to FLI and calculating the proportion of individuals in each group with hepatic steatosis.

If these proportions closely match the (dashed) line of equality, then FLI is well calibrated and FLI can reliably be used as a predictive probability of hepatic steatosis. The CIs generally include the line of equality and therefore our results indicate that the predicted probabilities of hepatic steatosis given by FLI are consistent with the observed percentages in our cohort, i.e. that FLI is reasonably well calibrated. The point estimates being above the line of equality indicate that FLI may underestimate the probability of a patient having hepatic steatosis and perhaps might be considered a pragmatic lower limit.

![Figure 1](distribution_of_liver_fat.png)

**Figure 1**

Distribution of liver fat in those with hepatic steatosis.

![Figure 2](roc_analysis.png)

**Figure 2**

Receiver operating characteristics (ROC) analysis of fatty liver index (FLI) to predict the presence or absence of non-alcoholic fatty liver disease (NAFLD).
Predicting hepatocellular lipid content using FLI and LAP

Both FLI and log-transformed LAP were linearly related to log-transformed liver fat (Fig. 4). However, the values of log-transformed liver fat varied considerably about the regression line. This variability is the primary contributor to the width of the 95% PIs and hence determines the estimated predictive ability of each algorithm. If we consider the width of the PIs at FLI = 30, a patient’s liver fat for this score could plausibly be between around 0 and 40%. For LAP = 30, the predicted liver fat could plausibly be between around 0 and 45%. Varying the predictive values of FLI and LAP at other values remained uninformative, suggesting FLI and LAP cannot be used to quantitatively determine liver fat content.

Validation of the models in patients with T2DM

Fifty-four patients with T2DM were recruited to validate the models in patients with T2DM, of whom 45 had hepatic steatosis and nine had normal liver fat. We performed the ROC analysis using only those patients with T2DM then compared it with the initial ROC analysis (Table 2). The non-significant P values suggest that the ability of each parameter to correctly classify a subject does not differ for the T2DM patients.

Discussion

Here, we provide external validation for the use of two previously reported indices, the FLI and the LAP, to determine any given individual’s probability of having hepatic steatosis, based on simple clinical parameters.
The FLI and LAP values are most useful to determine the probability of an individual having hepatic steatosis, but the strength of their relationship with liver fat content is insufficient for accurate prediction. A previous, elegant study by Kotronen et al. (29) in a large cohort of Finnish adults, using 1H-MRS measurements of liver fat, developed a NAFLD liver fat score and an equation distinct from FLI, which has been applied to predict NAFLD and liver fat content. However, in contrast to the indices discussed here, that score required measurement of fasting serum insulin concentrations. Significantly, serum ALT has a similar discriminatory ability to both of these algorithms.

To date, FLI has only been validated in a small group of females (n=25) using 1H-MRS measures of liver fat demonstrating a nonlinear relationship between FLI and liver fat content, limiting its predictive ability (21). Determining the severity of steatosis has arguably only limited clinical utility, mere identification of an individual as having hepatic steatosis, as part of the NAFLD spectrum, being sufficient to trigger prompt assessment and treatment of associated cardio-metabolic complications, and determination of the presence of NASH or fibrosis either non-invasively or by liver biopsy, to reduce the long-term risk of cardiovascular and hepatic complications.

FLI, adopted as a surrogate marker of hepatic steatosis, has been applied in numerous prospective, epidemiological studies and can predict the risk of incident T2DM (30), the incidence of atherosclerosis and cardiovascular disease (18) and of hepatic-related mortality after 15 years (16). Thus, FLI values have both diagnostic and prognostic significance.

Strengths of this study include the large number of well-characterised individuals with liver fat measured by the gold standard, non-invasive method. Furthermore, the study comprised a cross-section of individuals with normal liver fat and with hepatic steatosis (of a mild, moderate or severe degree). We acknowledge a limitation of the study was that individuals were recruited from three different research sites, thus there was a lack of standardisation of the analytical techniques (i.e. biochemical assays and MRS) between the three centres. However, analysis of the individual data sets from each of the three centres demonstrated similar results. A further limitation, inherent to these algorithms, is that...
although these values can predict the probability of hepatic steatosis, they have no predictive ability in identifying individuals who may have progressed along the NAFLD spectrum, with NASH or fibrosis.

In summary, we provide an external validation for the use of FLI and LAP using MRS. These results provide reassurance about its legitimacy as a surrogate marker for hepatic steatosis. Their application may be useful clinically, for metabolic research or epidemiological studies.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
Dr D J Cuthbertson and Prof. M Umpleby were supported by the European Foundation for the Study of Diabetes (EFSD).

Author contribution statement
D J Cuthbertson and G J Kemp contributed to the conception and design of the study, drafting and redrafting of the final manuscript. D Lythgoe contributed to the assembly, analysis and interpretation of data. M O Weickert, V S Sprung, R Dobson, F Shojaee-Moradie, M Umpleby, A F H Pfeiffer, E L Thomas, J D Bell and H Jones participated in the generation, collection and assembly of data. V S Sprung, M Umpleby and H Jones also participated in the drafting and redrafting of the manuscript. All authors approved the final version of the manuscript.

References


15 Kahn HS. The lipid accumulation product is better than BMI for identifying diabetics: a population-based comparison. Diabetes Care 2006 29 151–153. (doi:10.2337/diacare.29.01.06.dc05-1805)


