Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease

Eddowes, Peter J.; Sasso, Magali; Allison, Michael; Tsochatzis, Emmanuel; Anstee, Quentin M.; Sheridan, David; Guha, Indra N.; Cobbold, Jeremy F.; Deeks, Jonathan J.; Paradis, Valérie; Bedossa, Pierre; Newsome, Philip N.

DOI:
10.1053/j.gastro.2019.01.042

Document Version
Peer reviewed version

Citation for published version (Harvard):
Eddowes, PJ, Sasso, M, Allison, M, Tsochatzis, E, Anstee, QM, Sheridan, D, Guha, IN, Cobbold, JF, Deeks, JJ, Paradis, V, Bedossa, P & Newsome, PN 2019, 'Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease', *Gastroenterology*, vol. 156, no. 6, pp. 1717-1730. https://doi.org/10.1053/j.gastro.2019.01.042

Accepted Manuscript

Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Non-alcoholic Fatty Liver Disease

Peter J. Eddowes, Magali Sasso, Michael Allison, Emmanuel Tsochatzis, Quentin M. Anstee, David Sheridan, Indra N. Guha, Jeremy F. Cobbold, Jonathan J. Deeks, Valérie Paradis, Pierre Bedossa, Philip N. Newsome

PII: S0016-5085(19)30105-2
DOI: https://doi.org/10.1053/j.gastro.2019.01.042
Reference: YGAST 62409

To appear in: *Gastroenterology*
Accepted Date: 15 January 2019

> 450 patients with suspicion of NAFLD prospectively recruited

> Underwent liver biopsy within 2 weeks of FibroScan

(M or XL probe according to the automatic probe recommendation tool)

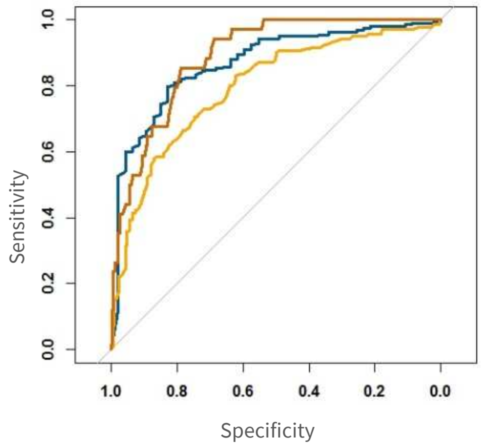
→ Steatosis

→ Fibrosis

→ CAP (dB/m)

→ LSM (kPa)

> Results and conclusions



CAP for steatosis (S≥1):
> AUC = 0.87 (0.82-0.92)

LSM for advanced fibrosis (F≥3):
> AUC = 0.80 (0.75-0.84)

LSM for cirrhosis (F=4):
> AUC = 0.89 (0.84-0.93)

> Steatosis or probe type had no impact on LSM (multivariable analysis)

>>> CAP and LSM by FibroScan are reliable biomarkers to non-invasively assess liver steatosis and fibrosis respectively in NAFLD

Gastroenterology

Title: Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Non-alcoholic Fatty Liver Disease

Short title: Diagnostic accuracy of CAP and LSM in NAFLD patients

Peter J Eddowes[1,2,3,4], Magali Sasso[5], Michael Allison[6], Emmanuel Tsochatzis[7], Quentin M Anstee[8], David Sheridan[9], Indra N Guha[4], Jeremy F Cobbold[10], Jonathan J Deeks[11], Valérie Paradis[12], Pierre Bedossa[12], Philip N Newsome*[1,2,3].

[1]National Institute for Health Research Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham

[2]Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham

[3]Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham

[4]National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, United Kingdom

[5]Echosens, R&D department, Paris, France

[6]Liver Unit, Addenbrooke's Hospital, Cambridge Biomedical Research Centre, Cambridge, United Kingdom

[7]University College London Institute for Liver and Digestive Health, Royal Free Hospital, London, United Kingdom

[8]Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[9]Institute of Translational and Stratified Medicine, Faculty of Medicine and Dentistry, University of Plymouth, United Kingdom

[10]Department of Gastroenterology and Hepatology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, United Kingdom

[11]National Institute for Health Research Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the Institute of Applied Health Research, University of Birmingham

[12]Department of Pathology, Physiology and Imaging, Beaujon Hospital Paris Diderot University, Paris, France

Grant support: This work was funded by Echosens who were sponsors of this study. PJE and PNN were supported by the National Institute of Health Research (NIHR) Birmingham Biomedical Research Centre (BRC). JFC was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). ING was supported by the National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Abbreviations:

A2M: alpha-2 macroglobulin

ALT: alanine transaminase

AST: aspartate aminotransferase

AUROC: area under the receiver operating characteristic curve

BIC: Bayesian information criteria

CAP: controlled attenuation parameter

CI: confidence interval

CK18-M30: cytokeratin 18 neo-epitope M30

CRP: C-reactive protein

FLIP: fatty liver: inhibition of progression

FN: false negative

FP: false positive

GGT: gamma-glutamyl transferase

HDL: high-density lipoprotein

HSI: hepatic steatosis index

IQR: interquartile range

INR: international normalized ratio

LDL: low-density lipoprotein

LB: liver biopsy

LR+: positive likelihood ratio

LR-: negative likelihood ratio

LSM: liver stiffness measurement

NAFL: non-alcoholic fatty liver

NAFLD: non-alcoholic fatty liver disease

NAS: non-alcoholic fatty liver disease activity score

NASH: non-alcoholic steato-hepatitis

NFS: NAFLD fibrosis score

NPV: negative predictive value

PPV: positive predictive value

ROC: receiver operating characteristic

SAF: steatosis activity fibrosis

Se: sensitivity

Sp: specificity

STARD: standards for reporting of diagnostic accuracy studies

TN: true negative

TP: true positive

VCTE: vibration-controlled transient elastography

* Corresponding Author

Professor Philip Newsome

NIHR Birmingham Biomedical Research Centre and Centre for Liver and Gastrointestinal Research

5th Floor Institute of Biomedical Research

University of Birmingham

Birmingham

B15 2TT

UK

Telephone: +44-121-415-8700

Fax: +44-121-415-8701

Email: P.N.Newsome@bham.ac.uk

Disclosures: No relevant disclosures.

Author contributions: PNN had the original concept and contributed to the design of the study protocol. PJE, with the assistance of the other recruiting sites performed the study and generated all of the data for the manuscript. MS, on behalf of the sponsor, performed the statistical analysis. PNN and MS wrote the first draft of the manuscript, and all authors reviewed the final version. PNN is guarantor.

Abstract

Background & Aims: We estimated the accuracy of FibroScan vibration-controlled transient elastography controlled attenuation parameter (CAP) and liver stiffness measurements (LSMs) in assessing steatosis and fibrosis in patients with suspected NAFLD.

Methods: We collected data from 450 consecutive adults who underwent liver biopsy analysis for suspected NAFLD at 7 centers in the United Kingdom from March 2014 through January 2017. FibroScan examinations with M or XL probe were completed within the 2 weeks of the biopsy analysis (404 had a valid examination). The biopsies were scored by 2 blinded expert pathologists according to non-alcoholic steatohepatitis clinical research network criteria. Diagnostic accuracy was estimated using the area under the receiver operating characteristic curves (AUROC) for the categories of steatosis and fibrosis. We assessed effects of disease prevalence on positive and negative predictive values. For LSMs, the effects of histological parameters and probe type were appraised using multivariable analysis.

Results: Using biopsy analysis as the reference standard, we found that CAP identified patients with steatosis with an AUROCs of 0.87 (95% CI, 0.82–0.92) for S≥S1, 0.77 (95% CI, 0.71–0.82) for S≥S2, and 0.70 (95% CI, 0.64–0.75) for S=S3. Youden cut-off values for S≥S1, S≥S2 and S≥S3 were 302 dB/m, 331 dB/m, and 337 dB/m respectively. LSM identified patients with fibrosis with AUROCs of 0.77 (95% CI, 0.72–0.82) for F≥F2, 0.80 (95% CI, 0.75–0.84) for F≥F3, and 0.89 (95% CI, 0.84–0.93) for F=F4. Youden cut-off values for F≥F2, F≥F3 and F=F4 were 8.2 kPa, 9.7 kPa, and 13.6 kPa respectively. Applying the optimal cut-off values, determined from this cohort, to populations of lower fibrosis prevalence increased negative predictive values and reduced positive predictive values. Multivariable analysis found that the only parameter that significantly affect LSMs was fibrosis stage (*P*<10-16); we found no association with steatosis or probe type.

Conclusions: In a prospective analysis of patients with NAFLD, we found CAP and LSMs by FibroScan to assess liver steatosis and fibrosis, respectively, with AUROC values ranging from 0.7 to 0.89. Probe type and steatosis did not affect LSMs. **Study registration:** ClinicalTrials.gov Identifier: NCT01985009.

KEY WORDS: VCTE, NASH, non-invasive, biomarker

Background & Aims:

Non-alcoholic fatty liver disease is an increasingly common cause of chronic liver disease, and is expected to soon become the commonest indication for liver transplantation[1, 2]. Estimates of its prevalence vary from 20-40% in the general population, although only 1-3% have evidence of significant inflammation and fibrosis[3]. The presence of liver fibrosis in particular is an important predictor of clinical events, both in terms of overall mortality and also liver-related morbidities and mortality[4, 5]. The challenge therefore remains how to identify those individuals with NAFLD that have more significant pathology in a manner which is non-invasive and affordable by healthcare systems.

Vibration-controlled transient elastography (VTCE) is one such approach which is in widespread clinical usage and for which there is an increasing understanding of clinically relevant cut-off values. By the use of a pulse-echo ultrasonic acquisition, vibration-controlled transient elastography (VCTE) can quantify the speed of a mechanically induced shear wave in liver tissue and hence generate an estimate of the degree of liver fibrosis with a liver stiffness measurement (LSM)[6, 7]. More recently this has been supplemented by the ability to quantify hepatic steatosis by measuring ultrasonic attenuation of the echo wave, termed the controlled attenuation parameter (CAP)[8, 9], which has been compared to liver biopsy in prospective studies with the M probe[10-12].

Previous studies have demonstrated the limitations of the M probe in patients with an increased skin to liver capsular distance as can occur commonly in NAFLD and overweight/obese patients[13, 14]; there is a much higher failure rate which led to the development of the XL probe. However, much of the published literature with the XL probe and CAP consists of either retrospective[15] or small/medium prospective cohort studies[16-19], with the exception of the recent NASH CRN studies[20, 21]. However, none have been the subject of large prospective powered diagnostic studies adhering to standards for reporting of diagnostic accuracy studies (STARD) guidelines[22].

Importantly, there are still uncertainties about the impact of other histological features on LSM readings with reports suggesting that steatosis may be a contributor[23, 24], although these studies were limited in that only the M probe was used. Similarly, whilst the advent of the XL probe has markedly reduced the failure rate in overweight/obese individuals[25], there are reports suggesting that cut-off ranges differ according to probe choice[26].

We designed a large prospective diagnostic study across 7 centres in the United Kingdom to evaluate the diagnostic accuracy of CAP measured either with the M or XL probe (depending on the FibroScan device automatic probe recommendation tool) in patients being investigated for potential NAFLD compared to a reference standard of histological evaluation of steatosis. The secondary objectives were to evaluate the diagnostic accuracy of LSM (with either M or XL probe) compared to a reference standard based on histological evaluation of fibrosis, and study of impact of histological parameters and probe type on LSM reading. In addition we aimed to identify cutoffs for use in clinical practice with both CAP and LSM.

Methods

Study participant and design

The study was a cross-sectional prospective multi-centre study, with the primary and secondary outcomes being to assess the diagnostic accuracy of CAP and LSM against liver histology which is the gold standard to evaluate the liver steatosis and fibrosis. NAFLD was suspected on the basis of the presence of abnormal liver enzymes in the presence of an ultrasound scan showing and echobright liver was the principle reason, usually in the presence of metabolic syndrome components. The STARD guidelines were followed to report the methods and results of this study[22] (see Supplementary Table 1 for further details). Consecutive patients were prospectively recruited between March 2014 and January 2017 in 7 liver centres across the United Kingdom (University Hospitals Birmingham NHS Foundation Trust, Birmingham; Addenbrooke's Hospital, Cambridge; Royal Free Hospital, London; Freeman Hospital, Newcastle upon Tyne; University Hospitals Plymouth NHS Trust, Plymouth; Queen's Medical Centre, Nottingham and John Radcliffe Hospital, Oxford).

The study (NCT01985009) was approved by the North Wales Research Ethics Committee (13/WA/0385) and by the Local Research Ethics Committee at each centre. All patients gave written informed consent to participate in the study. The study was conducted in accordance with the declaration of Helsinki and in agreement with the International Conference on Harmonisation (ICH) guidelines on Good Clinical Practice (GCP). All authors had access to the study data and reviewed and approved the final manuscript.

Main analyses: The primary outcome of the protocol was to evaluate the diagnostic accuracy of CAP measured either with the M or XL probe (depending on the FibroScan device automatic probe recommendation tool) against histological evaluation of steatosis. A secondary outcome of the protocol was to evaluate the diagnostic accuracy of liver stiffness measured either with M or XL probe (depending on the FibroScan device automatic probe recommendation tool) against histological evaluation of fibrosis.

Inclusion and exclusion criteria

Inclusion criteria were as follows: patients were ≥18 years of age, able to give written informed consent and were scheduled, independently from this study, to have a liver biopsy (LB) for investigation of assumed NAFLD within 2 weeks of Fibroscan examination (before or after). Patients were also negative for HBsAg, anti-HCV, HCV-RNA and HBVDNA. Exclusion criteria were as follows: patients with ascites, pregnant women, patient with any active implantable medical device (such as pacemaker or defibrillator), patients who had undergone liver transplantation, patients with cardiac failure and/or significant valvular disease, patients with haemochromatosis, patients that refused to undergo liver biopsy or blood tests, patients with an alcohol consumption above recommended limits (>14 units/week for women and >21 units/week for men; 1 unit = 8 g of ethanol), patients with a confirmed diagnosis of active malignancy, or other terminal disease, patient participating in another clinical trial within the preceding 30 days.

Patient Characteristics

The following characteristics were recorded for each patient: age, gender, BMI, presence of diabetes, hypertension, and hypercholesterolemia. For each patient, a 12 hour fasting blood collection was performed locally on the same day of the FibroScan procedure and was then shipped to a central laboratory for assessment of the following laboratory parameters: platelets count, international normalized ratio (INR), aspartate transaminase (AST), alanine transaminase (ALT), gamma-glutamyl-transferase (GGT), alkaline phosphatase, albumin, bilirubin, fasting glucose, total cholesterol, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol, triglyceride, ferritin, urea, creatinine, alpha-2-macroglobulin (A2M), hyaluronic acid, C-reactive protein (CRP) and cytokeratin 18 neo-epitope M30 (CK18-M30).

Histopathologic evaluation

Percutaneous LB was performed on all patients according to local standard procedure LB specimens were fixed in formalin, embedded in paraffin and stained with Hematoxylin and Eosin and Sirius Red for fibrosis evaluation. Slides were analysed independently by two experienced pathologists (PB and VP) who were blinded to each other's reading and also to the patient's clinical and Fibroscan data if available. In case of disagreement, they reviewed the slides together to reach consensus.

Steatosis (from 0 to 3), ballooning (from 0 to 2), lobular inflammation (from 0 to 3), fibrosis (from 0 to 4) and NAFLD activity score (NAS) were scored using the NASH clinical research network (NASH CRN) scoring system [27]. NASH was diagnosed using the "fatty liver: inhibition of progression" (FLIP) definition (presence of steatosis, hepatocyte ballooning and lobular inflammation with at least 1 point for each category). In addition, steatosis was semi-quantitatively assessed in percentage and the activity score (Ballooning (0-2) plus lobular inflammation (0-2)) according to the Steatosis Activity Fibrosis (SAF) was also assessed [28]. The presence of portal inflammation was also recorded. Biopsies were categorised by the pathologists as normal liver (no liver pathology), NAFL (steatosis but no NASH), NASH or other diagnosis when no NAFLD but other histological features suggestive of another diagnostic were observed (*e.g.* granulomatous hepatitis, biliary disease, autoimmune hepatitis). Interpretability for liver biopsy was based on the standard criteria of length, width and lack of major fragmentation. These criteria were occasionally over-looked by the pathologist when the biopsy showed obvious histological criteria of NASH, septal fibrosis or cirrhosis even if the biopsy was small or fragmented.

FibroScan liver stiffness measurement and controlled attenuation parameter

FibroScan (Echosens, Paris, France) examination was performed in each centre by nurses or physicians trained and certified by the manufacturer and blinded to the patient's histological evaluation. The FibroScan used in each center was a FibroScan 502 Touch model, equipped with both M and XL probes. An automatic probe selection tool was embedded in the device software which recommends the appropriate probe for each patient according to the real time assessment of the skin to liver capsule distance. The FibroScan examination procedure has been detailed previously[6, 29]. Briefly, all patients were asked to fast at least 3 hours prior to the examination, and then placed in the supine position with their right arm fully abducted. Measurements were performed by scanning the right liver lobe through an intercostal space.

The FibroScan device simultaneously measures LSM and CAP using VCTE technology. CAP has been designed to measure liver ultrasonic attenuation (go and return path) at 3.5 MHz on both M and XL probes[8], on signals acquired by the Fibroscan. The principle of CAP measurement has been described elsewhere[8, 9], and CAP was computed only when the associated LSM was valid and using the same signals as the one used to measure liver stiffness. At the beginning of the study, CAP was not available on the XL probe, therefore, the raw ultrasonic radio-frequency signals were stored in the Fibroscan examination file to enable computation of CAP off-line. CAP computation was performed blinded to all patients' clinical and histological data using the exact same configuration and algorithm to the one embedded in the commercial device for N=116 patients. When CAP was commercially available for the XL probe, all software were updated and the CAP value was displayed on the device screen for both probes during the procedure. The final CAP and LSM results were expressed in dB/m and kPa respectively. Only examinations with at least 10 valid individual measurements were deemed valid.

Statistical Analysis

Sample size estimation: Since no study had been performed previously using the probe recommendation on the FibroScan device, the sample size was calculated for patient measured with the XL probe only. It was hypothesized that approximately 1/3 of the total patients would be measured with M probe. Given the expected performance of CAP to detect steatosis (S≥S1) with an AUROC≥0.80[9, 30, 31], a projected sample size of 212 patients was deemed necessary to estimate an AUROC of 0.80 with the XL probe with an (1-α) confidence interval, α being set to 5%, at a 5% standard error level, for the XL probe only. The total number of patients measured using both probes was set to 312 patients and the final number of patients was set at 450 assuming a 30% drop-out rate

For descriptive statistics, continuous variables were expressed as medians [interquartile range (IQR)] and categorical variables as absolute figures with percentages. Confidence intervals were reported at the 95% level. Evidence for differences between CAP and LSM between steatosis grades and fibrosis stages was assessed using Kruskal-Wallis test followed by Dunn's tests with *post hoc* comparison. P values of < 0.05 were considered statistically significant.

Overall diagnostic accuracy of CAP and LSM was estimated as the area under the ROC curve (AUROC) together with its 95% confidence interval (CI). Data are reported for thresholds of steatosis and fibrosis. Cut-off values for CAP and LSM were identified that (a) maximise the Youden index, and also (b) at fixed values of sensitivity and specificity of 90%. For each cut-off value, we reported sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-) together with 95% confidence intervals. In additional analyses we investigated the performance of the tests in settings with different prevalence using Bayes equation to estimate post-test probabilities from the estimated likelihood ratios. For these computations we focused on fibrosis thresholds of F≥F2 and F=4 which are of particular importance as they correspond with stages which result in changes in patient management. We also identified cutoffs which minimized the consequences of test errors across different relative weightings of false positives and false negatives (see Supplementary Methods).

Factors influencing LSM: To evaluate the impact of histological parameters that possibly influenced LSM, a multivariable linear regression model was constructed with fibrosis stage, steatosis grade, ballooning grade, lobular inflammation and portal inflammation as candidate covariates and LSM as the outcome variable. In addition, the probe type used (M or XL) was also entered as a candidate covariate to evaluate if it had an impact on LSM when adjusted on histological parameters. All first order interactions were entered into the model. LSM was Box-Cox transformed to approximate a normal distribution. Final model selection was performed with a backward elimination procedure based on Bayesian information criteria (BIC). Multi-collinearity of independent variables was checked using the variance inflation factor. In addition to this multivariable analysis, LSM versus fibrosis stage stratified by probe type and by semi-quantitative steatosis percentage quartile was represented using a boxplot. Univariate analysis was performed using Kendall rank correlation coefficient between each histological parameter and LSM and was performed using the Mann-Whitney U test between the probe type and LSM.

The sensitivity analyses on CAP and LSM diagnostic accuracy and the analyses relative to the influence of disease prevalence on PPV and NPV, the cutoffs which minimized the consequences of test errors across different relative weightings of false positives and false negatives and factors influencing LSM were exploratory analyses which were not pre-specified.

For all analyses, only patients with histological results and median LSM or CAP values available with at least ten valid measurements were analyzed. In addition, no replacement of missing data has been performed. All analyses were performed using the software R, version 3.3.0[32].

Results

Patient Characteristics

The study flow chart is represented in Figure 1. Table 1 details the clinical, serological, histological characteristics and Fibroscan data of 383 patients with a valid FibroScan reading and an interpretable liver biopsy.

FibroScan applicability

Of 415 patients evaluated using the FibroScan (Figure 1), 138 (33%) were with the M probe and 277 (67%) with the XL probe. FibroScan readings were valid (with at least 10 valid individual measurements as per the manufacturer's recommendations) in 404 patients leading to an applicability value of 97%. For the 11 patients for whom a valid FibroScan was not achieved; 2 were with the M probe and 9 with the XL probe. Of note 4 of these 11 patients had 9 valid measurements (rather than the 10 required). Patients with less than 9 valid measurements (n=7) had a significantly higher BMI than others (46.5 [13.6] $kg.m^{-2}$ versus 36.4 [9.2] $kg.m^{-2}$; P = 0.003). Within the 404 patients with valid FibroScan, patients assessed with the XL probe (N=268) had a significantly higher BMI than patients measured by the M probe (36.3 [7.8] $kg.m^{-2}$ versus 29.3 [4.7] $kg.m^{-2}$; $P < 10^{-16}$). No adverse event has been reported related to the use of the FibroScan device.

Liver biopsies

A total of 412 patients underwent LB (see Figure 1: 433 eligible patients minus 16 patients who did not have LB, 4 patients who had LB cancelled by the investigator and 1 patient who withdrew consent before LB). The LB slides of 3 patients were lost during shipment and a further 15 LB were judged as non-interpretable by the pathologist leaving 394 (96%) as having an interpretable LB. A further ten patients had a LB that although interpretable by the pathologist could not be staged according to the NASH CRN scoring system. A description of those LB is provided in Supplementary Table 2 (2 patients being NAFLD with associated lesions and 8 being not NAFLD but not normal liver). Of note, 33 patients (8% of the patients with interpretable LB) had a histological diagnosis other than NAFLD or normal liver. A description of those LB is provided in Supplementary Table 2. After LB, 3 adverse events were reported: 1 patient had a syncopal episode following LB and pain at LB site requiring oral analgesia, 1 patient had hemorrhage following LB requiring hospitalization and 1 patient was admitted with pain and fever.

Assessment of steatosis using controlled attenuation parameter

Of 415 patients, 380 patients had an interpretable liver biopsy and valid CAP values (Figure 1). According to histological assessment, steatosis grade distribution was as follows: S0 = 47 (12%), S1 = 89 (23%), S2 = 107 (28%), S3 = 137 (36%) and the boxplot of CAP versus steatosis grade is shown in Figure 2a. CAP was significantly different between S0, S1 and S2 but not S2 and S3 (Kruskal-Wallis H = 97.70, $P < 10^{-16}$; Dunn's post hoc tests, P = 0.19 between CAP in S2 and CAP in S3, $P < 10^{-3}$ otherwise). Areas under the ROC curve (AUROC) as well as diagnostic performance of CAP cut-off values optimized using Youden’s index, a sensitivity of 90% or a specificity of 90% are detailed in Table 2 for S0 versus S1 and above, S0-S1 versus S2-S3 and S0-S2 versus S3. Accuracy was highest at the S≥S1 threshold, with an AUROC of 0.87 (95% CI: 0.82-0.92) and sensitivity of 0.80 (0.75-0.84) and specificity of 0.83 (0.69-0.92) at a threshold of 302 dB/m selected by maximizing Youden’s Index. Accuracy dropped to an AUC of 0.77 (0.71-0.82) for the S≥S2 threshold, with the corresponding sensitivity of 0.70 (0.63-0.75) and specificity of 0.76 (0.68-0.83) at the threshold of 331 dB/m maximizing Youden’s index and to an AUROC of 0.70 (0.64-0.75) for the S=S3 threshold with the corresponding sensitivity of 0.72 (0.63-0.79) and a specificity of 0.63 (0.56-0.69) at the threshold of 337 dB/m maximizing Youden's index. The ROC plots for S≥S1, S≥S2 and S=S3 are given in Supplementary Figure 1. Performance of CAP to diagnose NASH was also assessed. Corresponding AUC was 0.71 (0.65-0.76).

The use of quality criteria based on the IQR of CAP as proposed by Caussy *et al* [33] and Wong *et al* [34] which recommend excluding patients with IQR of CAP greater or equal to 30 dB/m or 40 dB/m, respectively was tested in our cohort. A large proportion of patients had an IQR of CAP ≥30 or 40 dB/m (57% and 39%, respectively), and performance was no better in patients with an IQR of CAP <30 or <40 dB/m (Supplementary Table 3). Indeed for the diagnosis of higher stages of steatosis performance was even lower in patient with an IQR of CAP <30 or <40 dB/m. To determine the influence of serum ALT on CAP diagnostic performance patients were stratified by ALT values (≤ULN, between ULN and 2xULN and >2xULN), but this did not influence CAP AUROCs (Supplementary Table 4). Performance of CAP was compared to the hepatic steatosis index (HSI) [35] in a subset of patients (N=375, due to 5 missing biological data). CAP significantly outperformed HSI for each steatosis grade S≥S1, S≥S2 and S=S3 (Supplementary Table 5).

Assessment of fibrosis using liver stiffness measurement

Of the 384 patients with valid LSM and interpretable LB, only 373 had fibrosis interpretable according to the NASH CRN scoring system (Figure 1). Differences in characteristics between the 373 patients used for fibrosis staging analysis and the 10 patients with fibrosis not staged are given in Supplementary Table 6.

Fibrosis stage distribution was as follows: F0: 62 (17%), F1: 86 (23%), F2: 85 (23%), F3: 106 (28%), F4: 34 (9%). LSM versus fibrosis stage is presented as a boxplot in Figure 2b.

LSM was significantly different between all fibrosis stages with the exception of F0 and F1 (Kruskal-Wallis H = 119.8, $P < 10^{-16}$; Dunn's post hoc tests, P = 1 between LSM in F0 and LSM in F1, P < 0.05 otherwise). AUC as well as diagnostic performance of LSM cut-off values optimized using Youden's index, a sensitivity of 90% or a specificity of 90% are detailed in Table 3 for F0-F1 versus F2 and above, F0-F2 versus F3-F4 and F0-F3 versus F4. Accuracy was highest at the F=F4 threshold, with an AUC of 0.89 (95% CI: 0.84-0.93) and sensitivity of 0.85 (0.69-0.95) and specificity of 0.79 (0.74-0.83) at a threshold of 13.6 kPa selected by maximizing Youden's Index. Accuracy was lower at lower fibrosis thresholds dropping to an AUROC of 0.80 (0.75-0.84) for F≥F3 with the corresponding sensitivity of 0.71 (0.62-0.78) and a specificity of 0.75 (0.69-0.80) at a threshold of 9.7 kPa maximizing the Youden's index and to an AUROC of 0.77 (0.72-0.82) for the F≥F2 threshold, with the corresponding sensitivity of 0.71 (0.64-0.77) and specificity of 0.70 (0.62-0.77) at the threshold of 8.2 kPa maximizing the Youden's index. The ROC plots for F≥F2, F≥F3 and F=F4 are given in Supplementary Figure 2. Performance of LSM to diagnose NASH was also assessed. Corresponding AUC was 0.68 (0.62-0.74).

The performance of the Boursier criteria[36] as a quality control for Fibroscan were evaluated in this cohort (IQR/median<30% in patient with LSM≥7.1 kPa). Whilst 43 (12%) patients did not reach the Boursier criteria, analysis in this cohort did not find evidence that these criteria improved performance of Fibroscan (Supplementary Table 7) where we have assessed AUROC for patients reliable according to Boursier's criteria only. The influence of ALT on LSM diagnostic performance was evaluated by stratifying patients on ALT values (≤ULN, between ULN and 2xULN and >2xULN). No significant influence of the effect of ALT on the LSM AUROC for each fibrosis stage was observed (Supplementary Table 8). The performance of the Baveno VI cut-offs[37], in relation to patients with compensated advanced chronic liver disease with advanced fibrosis (F≥F3) were tested in this cohort. The NPV associated with the ≤10 kPa cutoff was 0.80 and the PPV associated with the ≥15 kPa cutoff was 0.75.

Performance of LSM was also compared to Fib4[38] and the NAFLD fibrosis score (NFS[39]). Diagnostic performance in terms of AUROC for each fibrosis stage (≥F2, F≥F3 and F=F4) are provided in Supplementary Table 9. LSM outperformed Fib4 and NFS for the diagnosis of cirrhosis and NFS for the diagnosis of F≥2. For the diagnosis of advanced fibrosis, performance of LSM was compared using the dual cut-offs (cut-off for Se≥0.90 = 7.1 kPa and cut-off for Sp≥0.90 = 14.1 kPa determined in the present cohort) against the dual cut-offs for Fib4 (1.30 and 3.25)[38] and NFS (-1.455 and 0.676)[39]. LSM had a higher Se for the confirmation of advanced fibrosis (F≥3) with a PPV = 0.74 (Supplementary Table 10).

Further analysis was performed to identify cutoffs which minimized the consequences of test errors across different relative weightings of false positives and false negatives (see Supplementary Results and Supplementary Table 11). In these analyses the consequences of diagnostic error were explored in situations where the priority was to either avoid false positive diagnoses (for the diagnostic of F≥F2) or false negative diagnoses (for the diagnostic of F=F4). The analyses were performed under a range of scenarios with the cost of a false positive (FP) being set at 2 times, 5 times and 10 times worse than a false negative (FN) for the diagnostic of F≥F2. The effect on threshold is shown in Supplementary Table 11 along with the corollary analyses for the diagnostic of F=F4.

Impact of fibrosis prevalence on predictive value of liver stiffness measurement

We set out to determine the impact of fibrosis prevalence on PPV and NPV values by utilising a range of different pre-test probabilities values (prevalence). The prevalence figures used represent values from this cohort (60, 38% and 9% for F≥F2, F≥F3 and F=4 respectively) and also values seen in cohorts of patients with type 2 diabetes mellitus, patients at risk of liver disease and the general population[40-42]. For a diagnosis of F≥F2, F≥F3 and F=F4 there was a marked reduction in the PPV as the prevalence of fibrosis was lowered (Table 4). Rounding the proposed cut-offs did not affect the PPV and NPV, irrespective of prevalence (see Supplementary Table 12).

Influence of probe type and histological parameters on liver stiffness measurement

We next investigated the influence of probe type and histological parameters on LSM values. In univariate analysis, no significant difference was found between LSM and the probe type (P = 0.55); all histological parameters were significantly correlated to LSM: fibrosis stage ($\tau = 0.43$, $P < 10^{-16}$), ballooning grade ($\tau = 0.22$, $P < 10^{-7}$), lobular inflammation grade ($\tau = 0.21$, $P < 10^{-6}$), portal inflammation grade ($\tau = 0.17$, $P < 10^{-4}$) and steatosis grade ($\tau = 0.11$, $P = 0.004$). Then, a multivariable linear regression analysis was performed. Following a backward selection procedure based on BIC, the only covariate influencing LSM was fibrosis stage ($\beta = 0.18$, 95% CI = (0.15-0.21), $P < 10^{-16}$). When adjusted for fibrosis stage, there was no significant influence of probe type or steatosis grade on the LSM value. To further illustrate this, a boxplot of LSM versus fibrosis stage stratified by probe type is presented in Figure 3a and a boxplot of LSM stratified by semi-quantitative steatosis percentage quartile is presented in Figure 3b.

Conclusions

This prospective study examined the association of contemporaneous VTCE and liver histology in a cohort of patients undergoing liver biopsy for investigation for suspected NAFLD, and the results were reported according to the STARD guidelines. It demonstrates the high applicability rate of VTCE (97%) in a large UK NAFLD cohort with BMI up to 53.2 kg/m² and provides optimised cut-off values for staging steatosis and fibrosis depending on prevalence and clinical context (Youden criteria, 90% sensitivity or 90% specificity). This study also provides novel approaches to threshold setting taking into account the prevalence of fibrosis in the population to be tested and also basing thresholds around clinical priorities such as minimising false positive diagnoses of F≥F2 or false negative diagnoses of F=4. Critically this study demonstrates that only fibrosis stage, and not probe type or any other histological parameters, influence LSM values.

Whilst the cut-offs for steatosis grade increase progressively from S0 to S3 when set for high sensitivity or high specificity there is not much difference between S2 and S3 when using the Youden cut-off values which were 331 dB/m and 337 dB/m respectively. Nevertheless in clinical practice the identification of moderate steatosis is of greater utility than distinctions between S2 and S3, and thus the Youden cut-off for S ≥S2 of 331 dB/m is sufficient. The determination of steatosis by CAP is relevant for the confirmation of any degree of steatosis and also potentially as a serial measure in response to lifestyle or pharmacological/surgical intervention. The former is demonstrably feasible in this study whereas the latter will require examination in intervention studies.

With regards to the association between LSM values and histological evaluation of liver fibrosis there is a clear demarcation between the different degrees of fibrosis for Youden cut-off as well as for those with high sensitivity or specificity. As expected the cut-off for liver cirrhosis is markedly higher at 20.9 kPa when the specificity is set at 90%. The Youden cut-off values from this study for F≥F2, F≥F3 and F=F4 were 8.2 kPa, 9.7 kPa, and 13.6 kPa respectively, which demonstrate a clear upward increment with progressive liver fibrosis. These cut-off values have good sensitivity and specificity with a good PPV (0.78) for ≥F2 and an excellent NPV (0.98) for F4. Distinguishing F0-F2 versus F3-4 can be achieved despite a slightly lower PPV (0.63), although there is a higher NPV (0.81) with the cut-off for F≥F3.

The diagnostic performance of LSM and cutoffs for stages of fibrosis in this study are broadly in keeping with data from a US cohort[20] (Supplementary Table 13) and those recommended in a UK guideline[43]. The cutoffs from a range of other published studies are included in Supplementary Table 14 for comparison. Whilst reasonably similar there are some differences in the UK cohort such as gender (45% female vs 68% female in US cohort) and presence of diabetes mellitus (50% vs 44% in US cohort). For CAP however, diagnostic performance is higher in our cohort than in the US cohort (AUROC 0.87 (0.82-0.92) for the diagnostic of S≥1 in our cohort versus 0.76 (0.64-0.89) in the US cohort. This difference may be accounted to the prevalence of patients with S≥S1 steatosis which is 88% in our cohort versus 95% in the US cohort. Another possibility is that the delay between FibroScan and LB was up to 12 months in NASH CRN study whereas in this study it was only 2 weeks.

Reports have suggested that factors other than liver fibrosis, such as steatosis[23], may influence LSM readings. To evaluate this question we performed multivariable analysis including all potentially relevant factors and notably the only factor that predicted LSM was the degree of liver fibrosis. Explicitly, neither the degree of steatosis or inflammation was associated with differences in LSM. This is likely because prior studies had not included other factors such as degree of fibrosis in their analyses, which when taken into account reveal that other histological elements do not influence LSM readings[23]. Also these studies only used the M probe which is likely to give an incorrect reading in many patients with NAFLD. Similarly, groups have suggested that LSM cut-offs differ according to probe choice[20, 26], although in this study we did not find this to be the case.

The threshold values will also be significantly impacted by the prevalence of the underlying condition. In Table 4 the effect of changing prevalence is demonstrated again allowing for appropriate choice of cut-off values depending on the clinical setting. This modelling data demonstrates that as the prevalence of liver fibrosis (≥F2 or F4) decreases there is a commensurate reduction in PPV and increase in NPV. This is relevant as cut-offs generated in secondary care are often applied in primary care without taking into account the marked difference in prevalence. In this situation a negative test would be very reassuring although a positive test would have a low likelihood of capturing a true positive and raises the question of needing further confirmatory tests.

Conventional cut-off criteria for grades of steatosis and fibrosis whilst useful, do not capture the importance to clinical decision making and its dependence on the relevant clinical setting. To better model this we explored two settings; one in which the presence of ≥F2 or F4 was being tested (Supplementary Appendix). In the former setting (≥F2) the assumption was made that a false positive was two, five or ten times worse than a false negative, with concomitant increases in the threshold. In contrast for F4 the opposite view was taken, namely that it was more important to not miss a diagnosis (Supplementary Table 11). This allows for healthcare organisations to make decision depending on how they value the ratio of false positive to false negatives.

Our study has several strengths; it is a large prospective appropriately powered study, and captures real world clinical practice of clinicians evaluating patients with potential NAFLD. By incorporating the automatic probe recommendation tool we also ensured that the correct probe was used to generate LSM and CAP values. It defines a number of cut-offs which can be used according to the clinical setting and also provides modelling data on the impact of prevalence on performance.

A potential weakness of our study is that a number of biopsies were not interpretable as they did not show NAFLD but there again this is representative of real-world examination of this technology. In addition, we did not establish whether repeat VTCE examination would have generated consistent readings as demonstrated recently[20].

In summary, this study confirms the high applicability/low failure rate of VTCE in a cohort of patients with potential NAFLD, and demonstrate that LSM readings are not influenced by other histological components or choice of probe. Finally, our study provides a comprehensive range of cut-offs for LSM and CAP depending on the value a clinician places on false positive/false negatives as well as taking into account the prevalence of the degree of fibrosis. This will be critical for the roll-out of VTCE in a range of clinical settings.

Figure legends

Figure 1. Study flow chart.

Of 450 patients enrolled, 433 were eligible, 415 had the FibroScan examination performed and 404 had a valid FibroScan examination. Eventually 383 had a valid controlled attenuation parameter (CAP) measurements and steatosis grade assessed on liver biopsy (LB) and 373 had a valid liver stiffness measurement (LSM) and fibrosis stage assessed on LB.

Figure 2. Boxplot of (a) controlled attenuation parameter (CAP) versus steatosis grade, (b) liver stiffness measurement (LSM) versus fibrosis stage.

(a) CAP values increase with increasing steatosis grade (Kruskal–Wallis test $p < 10^{-16}$, Dunn's *post hoc* tests, $p = 0.19$ between CAP in S2 and CAP in S3, $p < 10^{-3}$ otherwise); (b) LSM values increase significantly with increasing fibrosis stage (Kruskal-Wallis $p < 10^{-16}$; Dunn's *post hoc* tests, $p = 1$ between LSM in F0 and LSM in F1, $p < 0.05$ otherwise).

Figure 3. Boxplot of LSM versus fibrosis stage stratified by (a) probe type, (b) quartile of semi-quantitative steatosis percentage.

The boxplot represent the LSM distribution for each fibrosis stage (a) according to the probe used. Patients were scanned either with the M or XL probe as proposed by the automatic probe recommendation tool. (b) stratified by steatosis amount: for each fibrosis stage, patients are stratified by steatosis quartile in the fibrosis stage.

Table legends

Table 1. Patient characteristics

Table 2. Diagnostic performance of controlled attenuation parameter (CAP) for steatosis grade greater or equal than 1, greater or equal than 2 and equal to 3.

Table 3. Diagnostic performance of liver stiffness measurement (LSM) for each fibrosis stage greater or equal than 2, greater or equal than 3 and equal to 4.

Table 4. Impact of prevalence of F≥F2 and F=4 on positive predictive value (PPV) and negative predictive value (NPV) for cut-offs.

References

1. Younossi ZM, Blissett D, Blissett R, et al. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. Hepatology 2016;64:1577-1586.
2. Charlton MR, Burns JM, Pedersen RA, et al. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. Gastroenterology 2011;141:1249-53.
3. Younossi ZM, Koenig AB, Abdelatif D, et al. Global Epidemiology of Non-Alcoholic Fatty Liver Disease-Meta-Analytic Assessment of Prevalence, Incidence and Outcomes. Hepatology 2015.
4. 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.
5. Dulai PS, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology 2017;65:1557-1565.
6. Sandrin L, Fourquet B, Hasquenoph JM, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705-13.
7. Tsochatzis EA, Gurusamy KS, Ntaoula S, et al. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. J Hepatol 2011;54:650-9.
8. Sasso M, Audiere S, Kemgang A, et al. Liver Steatosis Assessed by Controlled Attenuation Parameter (CAP) Measured with the XL Probe of the FibroScan: A Pilot Study Assessing Diagnostic Accuracy. Ultrasound Med Biol 2016;42:92-103.
9. Sasso M, Beaugrand M, de Ledinghen V, et al. Controlled attenuation parameter (CAP): a novel VCTE guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. Ultrasound Med Biol 2010;36:1825-35.
10. de Ledinghen V, Vergniol J, Capdepont M, et al. Controlled attenuation parameter (CAP) for the diagnosis of steatosis: a prospective study of 5323 examinations. J Hepatol 2014;60:1026-31.
11. Jun BG, Park WY, Park EJ, et al. A prospective comparative assessment of the accuracy of the FibroScan in evaluating liver steatosis. PLoS One 2017;12:e0182784.
12. Runge JH, Smits LP, Verheij J, et al. MR Spectroscopy-derived Proton Density Fat Fraction Is Superior to Controlled Attenuation Parameter for Detecting and Grading Hepatic Steatosis. Radiology 2018;286:547-556.
13. Myers RP, Pomier-Layrargues G, Kirsch R, et al. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. Hepatology 2012;55:199-208.
14. Tapper EB, Challies T, Nasser I, et al. The Performance of Vibration Controlled Transient Elastography in a US Cohort of Patients With Nonalcoholic Fatty Liver Disease. Am J Gastroenterol 2016;111:677-84.
15. Naveau S, Voican CS, Lebrun A, et al. Controlled attenuation parameter for diagnosing steatosis in bariatric surgery candidates with suspected nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol 2017;29:1022-1030.

16. Chan WK, Nik Mustapha NR, Wong GL, et al. Controlled attenuation parameter using the FibroScan(R) XL probe for quantification of hepatic steatosis for non-alcoholic fatty liver disease in an Asian population. United European Gastroenterol J 2017;5:76-85.
17. Park CC, Nguyen P, Hernandez C, et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. Gastroenterology 2017;152:598-607 e2.
18. Garg H, Aggarwal S, Shalimar, et al. Utility of transient elastography (fibroscan) and impact of bariatric surgery on nonalcoholic fatty liver disease (NAFLD) in morbidly obese patients. Surg Obes Relat Dis 2018;14:81-91.
19. de Ledinghen V, Hiriart JB, Vergniol J, et al. Controlled Attenuation Parameter (CAP) with the XL Probe of the Fibroscan((R)): A Comparative Study with the M Probe and Liver Biopsy. Dig Dis Sci 2017;62:2569-2577.
20. Vuppalanchi R, Siddiqui MS, Van Natta ML, et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. Hepatology 2018;67:134-144.
21. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al. Vibration-controlled Transient Elastography to Assess Fibrosis and Steatosis in Patients With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol 2018.
22. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
23. Petta S, Maida M, Macaluso FS, et al. The severity of steatosis influences liver stiffness measurement in patients with nonalcoholic fatty liver disease. Hepatology 2015;62:1101-10.
24. Petta S, Wong VW, Camma C, et al. Improved noninvasive prediction of liver fibrosis by liver stiffness measurement in patients with nonalcoholic fatty liver disease accounting for controlled attenuation parameter values. Hepatology 2017;65:1145-1155.
25. Wong VW, Vergniol J, Wong GL, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862-71.
26. de Ledinghen V, Wong VW, Vergniol J, et al. Diagnosis of liver fibrosis and cirrhosis using liver stiffness measurement: comparison between M and XL probe of FibroScan(R). J Hepatol 2012;56:833-9.
27. 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.
28. Bedossa P, Poitou C, Veyrie N, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology 2012;56:1751-9.
29. de Ledinghen V, Vergniol J. Transient elastography (FibroScan). Gastroenterol Clin Biol 2008;32:58-67.
30. de Ledinghen V, Vergniol J, Foucher J, et al. Non-invasive diagnosis of liver steatosis using controlled attenuation parameter (CAP) and transient elastography. Liver Int 2012;32:911-8.
31. Myers RP, Pollett A, Kirsch R, et al. Controlled Attenuation Parameter (CAP): a noninvasive method for the detection of hepatic steatosis based on transient elastography. Liver Int 2012;32:902-10.

32. R Core Team. R: A Language and Environment for Statistical Computing. In: https://www.R-project.org/, ed. Vienna, Austria: R Foundation for Statistical Computing, 2016.
33. Caussy C, Alquiraish MH, Nguyen P, et al. Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. Hepatology 2017.
34. Wong VW, Petta S, Hiriart JB, et al. Validity criteria for the diagnosis of fatty liver by M probe-based controlled attenuation parameter. J Hepatol 2017.
35. Lee JH, Kim D, Kim HJ, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis 2010;42:503-8.
36. Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013;57:1182-91.
37. de Franchis R, Baveno VIF. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. J Hepatol 2015;63:743-52.
38. 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.
39. 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.
40. Roulot D, Costes JL, Buyck JF, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut 2011;60:977-84.
41. Harris R, Harman DJ, Card TR, et al. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. Lancet Gastroenterol Hepatol 2017;2:288-297.
42. Kwok R, Choi KC, Wong GL, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut 2016;65:1359-68.
43. Newsome PN, Cramb R, Davison SM, et al. Guidelines on the management of abnormal liver blood tests. Gut 2018;67:6-19.

Table 1. Patient characteristics

Characteristic	*N*	*Distribution*	*Range*
Centre	383	Birmingham: 102 (27%) Newcastle: 51 (13%) London: 52 (14%) Nottingham: 40 (10%) Plymouth: 48 (13%) Cambridge: 60 (16%) Oxford: 30 (8%)	–
Age (years)	383	54 [18]	[19-77]
BMI (kg.m^{-2})	383	33.8 [9.2],	[19.5-53.2]
Female gender	383	171 (45%)	–
Diabetes mellitus	383	193 (50%)	–

Hypertension	383	207 (54%)	–
Hypercholesterolemia	383	199 (52%)	–
Platelets count (x10^9/L)	373	236 [84]	[57-446]
INR	361	1.08 [0.09]	[0.81-2.54]
AST (IU/L)	378	36 [25]	[9-203]
ALT (IU/L)	378	50 [40]	[7-298]
GGT (IU/L)	378	59 [88]	[9-1718]
Alkaline phosphatase (IU/L)	377	82 [40]	[4-738]
Albumin (g/dL)	379	4.5 [0.4]	[3.6-5.5]
Bilirubin (mg/dL)	378	0.50 [0.35]	[0.12-3.96]
Fasting glucose (mg/dL)	376	106 [51]	[50-312]
Total cholesterol (mg/dL)	363	179 [64]	[80-274]

HDL cholesterol (mg/dL)	351	43 [17]	[15-101]
LDL cholesterol (mg/dL)	350	102 [51]	[3-189]
Triglyceride (mg/dL)	362	161 [92]	[51-501]
Ferritin (ng/mL)	378	134 [214]	[7-4320]
Urea (mg/dL)	378	29 [11]	[12-84]
Creatinine (mg/dL)	379	0.85 [0.22]	[0.36-1.94]
A2M (mg/dL)	376	205 [121]	[91-523]
Hyaluronic acid (ug/L)	379	40 [55]	[19-1850]
CRP (mg/dL)	378	0.31 [0.47]	[0.02-7.53]
CK18-M30 (IU/L)	369	415 [395]	[74-1825]
Time between FibroScan and liver biopsy (day)	383	0 [7]	[0-14]

XL probe	383	255 (67%)	–
LSM (kPa), range 1.5-75 kPa	383	8.8 [7.8]	[1.7-75.0]
CAP (dB/m), range 100-400 dB/m	380	336 [74]	[100-400]
Length of liver biopsy specimen (mm)	383	23 [10]	[5-60]
Fibrosis stage	373	F0: 62 (17%) F1: 86 (23%) F2: 85 (23%) F3: 106 (28%) F4: 34 (9%)	–
Steatosis grade	383	S0: 47 (12%) S1: 89 (23%) S2: 109 (28%)	–

		S3: 138 (36%)	
Ballooning grade	383	B0: 106 (28%) B1: 147 (38%) B2: 130 (34%)	–
Lobular inflammation grade	383	I0: 90 (23%) I1: 235 (61%) I2: 51 (13%) I3: 7 (2%)	–
NAS score	383	0-2: 90 (23%) 3-4: 122 (32%) 5-8: 171 (45%)	–
Activity grade (according to SAF)	383	A0: 55 (14%) A1: 80 (21%) A2: 102 (27%)	–

		A3: 110 (29%) A4: 36 (9%)	
Portal inflammation present	382	172 (45%)	–
Pathologists diagnosis	383	Normal liver: 17 (4%) NAFL: 91 (24%) NASH: 242 (63%) Other: 33 (9%)	–

Distribution is expressed as median [interquartile range] or figure (percentage).

A2M: alpha-2 macroglobulin, ALT: alanine transaminase, AST: aspartate aminotransferase, BMI: body mass index, CK18-M30: cytokeratin 18 neoepitope M30, CAP: controlled attenuation parameter, CRP: C-reactive protein, GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, INR: international normalized ratio, LDL: low-density lipoprotein, LSM: liver stiffness measurement, NAFL: non-alcoholic fatty liver, NAFLD: NAFL disease, NASH: non-alcoholic steato-hepatitis, NAS: NAFLD activity score.

Table 2. Diagnostic performance of controlled attenuation parameter (CAP) for steatosis grade greater or equal than 1, greater or equal than 2 and equal to 3.

		S≥S1 (≥5% steatosis)	S≥S2 (≥34% steatosis)	S=S3 (≥67% steatosis)
AUROC (95%CI)		0.87 (0.82-0.92)	0.77 (0.71-0.82)	0.70 (0.64-0.75)
Prevalence (N)		0.88 (N=303)	0.64 (N=244)	0.36 (N=137)
Youden Index	Cut-off (dB/m)	302	331	337
	Se (95%CI)	0.80 (0.75-0.84)	0.70 (0.63-0.75)	0.72 (0.63-0.79)
	TP/(TP+FN)	*(266/333)*	*(170/244)*	*(98/137)*
	Sp (95%CI)	0.83 (0.69-0.92)	0.76 (0.68-0.83)	0.63 (0.56-0.69)
	TN/(TN+FP)	*(39/47)*	*(104/136)*	*(152/243)*
	PPV (95% CI)	0.97 (0.94-0.98)	0.84 (0.78-0.88)	0.52 (0.45-0.62)
	NPV (95% CI)	0.37 (0.31-0.59)	0.58 (0.52-0.68)	0.80 (0.73-0.84)
	LR+ (95% CI)	4.69 (2.49-8.84)	2.96 (2.16-4.05)	1.91 (1.57-2.32)

	LR- (95% CI)	0.24 (0.19-0.31)	0.40 (0.32-0.49)	0.46 (0.34-0.60)
Se=0.90	Cut-off (dB/m)	274	290	302
	Se (95%CI) *TP/(TP+FN)* Sp (95%CI) *TN/(TN+FP)*	Se = 0.90 (0.87-0.93) *(301/333)* Sp = 0.60 (0.44-0.74) *(28/47)*	Se = 0.90 (0.86-0.94) *(220/244)* Sp = 0.44 (0.36-0.53) *(60/136)*	Se = 0.90 (0.83-0.94) *(123/137)* Sp = 0.38 (0.32-0.44) *(92/243)*
	PPV (95% CI) NPV (95% CI)	PPV = 0.94 (0.90-0.96) NPV = 0.47 (0.38-0.62)	PPV = 0.74 (0.67-0.82) NPV = 0.71 (0.62-0.78)	PPV = 0.45 (0.38-0.61) NPV = 0.87 (0.79-0.90)
	LR+ (95% CI) LR- (95% CI)	LR+ = 2.24 (1.58-3.17) LR- = 0.16 (0.11-0.24)	LR+ = 1.61 (1.38-1.88) LR- = 0.22 (0.15-0.34)	LR+ = 1.44 (1.29-1.62) LR- = 0.27 (0.16-0.45)
Sp=0.90	Cut-off (dB/m)	325	370	398
	Se (95%CI) *TP/(TP+FN)*	Se = 0.66 (0.61-0.71]) *(220/333)*	Se = 0.34 (0.28-0.40) *(83/244)*	Se = 0.14 (0.09-0.21) *(19/137)*

	Sp (95%CI) *TN/(TN+FP)*	Sp = 0.90 (0.77-0.96) *(42/47)*	Sp = 0.90 (0.83-0.94) *(122/136)*	Sp = 0.90 (0.86-0.94) *(219/243)*
	PPV (95% CI) NPV (95% CI)	PPV = 0.98 (0.95-0.98) NPV = 0.27 (0.23-0.55)	PPV = 0.86 (0.77-0.89) NPV = 0.43 (0.36-0.59)	PPV = 0.44 (0.34-0.56) NPV = 0.65 (0.52-0.75)
	LR+ (95% CI) LR- (95% CI)	LR+ = 6.21 (2.70-14.27 LR- = 0.38 (0.32-0.45)	LR+ = 3.30 (1.95-5.59) LR- = 0.74 (0.66-0.82)	LR+ = 1.40 (0.80-2.47) LR- = 0.96 (0.88-1.03)

AUROC: area under the receiver operating curve, CI: confidence interval, FN: number of false negative, FP: number of false positive, LR-: negative likelihood ratio, LP+: positive likelihood ratio, NPV: negative predictive value, PPV: positive predictive value, S: steatosis, Se: sensitivity, Sp: specificity, TN: true negative, TP: true positive.

Table 3. Diagnostic performance of liver stiffness measurement (LSM) for each fibrosis stage greater or equal than 2, greater or equal than 3 and equal to 4.

		F≥F2	**F≥F3**	**F=F4**
AUROC (95%CI)		**HIS**	**0.80 (0.75-0.84)**	**0.89 (0.84-0.93)**
Prevalence (N)		0.60 (N=225)	0.38 (N=140)	0.09 (N=34)
Youden Index	Cut-off (kPa)	8.2	9.7	13.6
	Se (95%CI) *TP/(TP+FN)* Sp (95%CI) *TN/(TN+FP)*	Se = 0.71 (0.64-0.77) *(159/225)* Sp = 0.70 (0.62-0.77) *(103/148)*	Se = 0.71 (0.62-0.78) *(99/140)* Sp = 0.75 (0.69-0.80) *(174/233)*	Se = 0.85 (0.69-0.95) *(29/34)* Sp = 0.79 (0.74-0.83) *(267/339)*
	PPV (95% CI) NPV (95% CI)	PPV = 0.78 (0.71-0.83) NPV = 0.61 (0.54-0.69)	PPV = 0.63 (0.55-0.71) NPV = 0.81 (0.74-0.85)	PPV = 0.29 (0.24-0.57) NPV = 0.98 (0.95-0.99)
	LR+ (95% CI)	LR+ = 2.32 (1.80-3.01)	LR+ = 2.79 (2.19-3.57)	LR+ = 4.02 (3.13-5.15)

	LR- (95% CI)	LR- = 0.42 (0.34-0.53)	LR- = 0.39 (0.30-0.51)	LR- = 0.19 (0.08-0.42)
Se=0.90	Cut-off (kPa)	6.1	7.1	10.9
	Se (95%CI) *TP/(TP+FN)* Sp (95%CI) *TN/(TN+FP)*	Se = 0.90 (0.86-0.94) *(203/225)* Sp = 0.38 (0.30-0.46) *(56/148)*	Se = 0.90 (0.84-0.94) *(126/140)* Sp = 0.50 (0.43-0.56) *(116/233)*	Se = 0.91 (0.76-0.98) *(31/34)* Sp = 0.70 (0.64-0.74) *(236/339)*
	PPV (95% CI) NPV (95% CI)	PPV = 0.69 (0.61-0.78) NPV = 0.72 (0.62-0.78)	PPV = 0.52 (0.45-0.67) NPV = 0.89 (0.83-0.92)	PPV = 0.23 (0.19-0.61) NPV = 0.99 (0.96-0.99)
	LR+ (95% CI) LR- (95% CI)	LR+ = 1.45 (1.27-1.66) LR- = 0.26 (0.17-0.40)	LR+ = 1.79 (1.56-2.06) LR- = 0.20 (0.12-0.34)	LR+ = 3.00 (2.48-3.64) LR- = 0.13 (0.04-0.37)
Sp=0.90	Cut-off (kPa)	12.1	14.1	20.9
	Se (95%CI) *TP/(TP+FN)*	Se = 0.44 (0.38-0.51) *(100/225)*	Se = 0.48 (0.39-0.56) *(67/140)*	Se = 0.59 (0.41-0.75) *(20/34)*

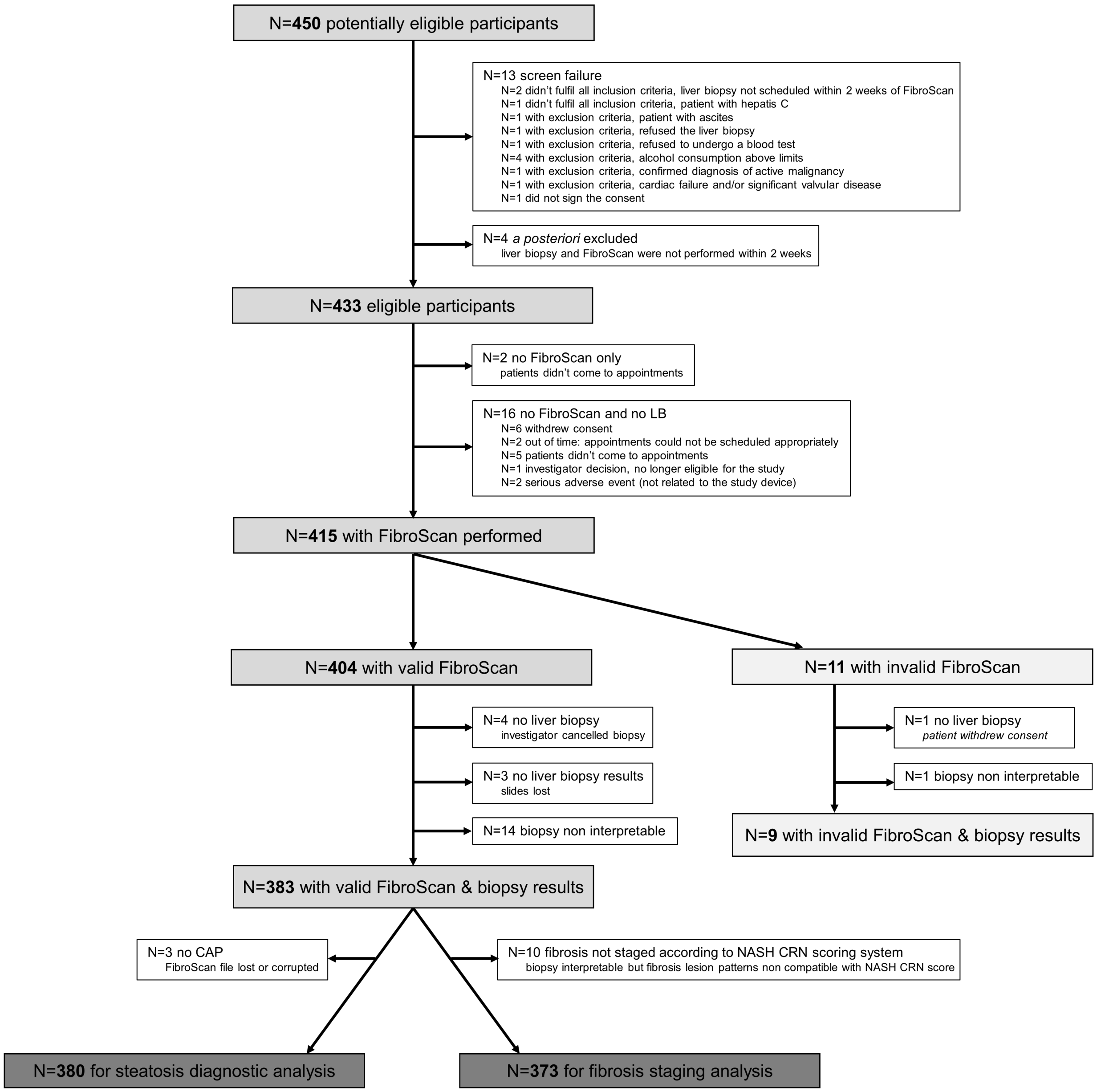
	Sp (95%CI) *TN/(TN+FP)*	Sp = 0.91 (0.85-0.95) *(134/148)*	Sp = 0.90 (0.86-0.94) *(210/233)*	Sp = 0.90 (0.86-0.93) *(305/339)*
	PPV (95% CI) NPV (95% CI)	PPV = 0.88 (0.80-0.90) NPV = 0.52 (0.45-0.67)	PPV = 0.74 (0.65-0.80) NPV = 0.74 (0.67-0.82)	PPV = 0.37 (0.29-0.56) NPV = 0.96 (0.91-0.97)
	LR+ (95% CI) LR- (95% CI)	LR+ = 4.70 (2.79-7.90) LR- = 0.61 (0.54-0.70)	LR+ = 4.85 (3.17-7.41) LR- = 0.58 (0.49-0.68)	LR+ = 5.87 (3.83-8.97) LR- = 0.46 (0.31-0.69)

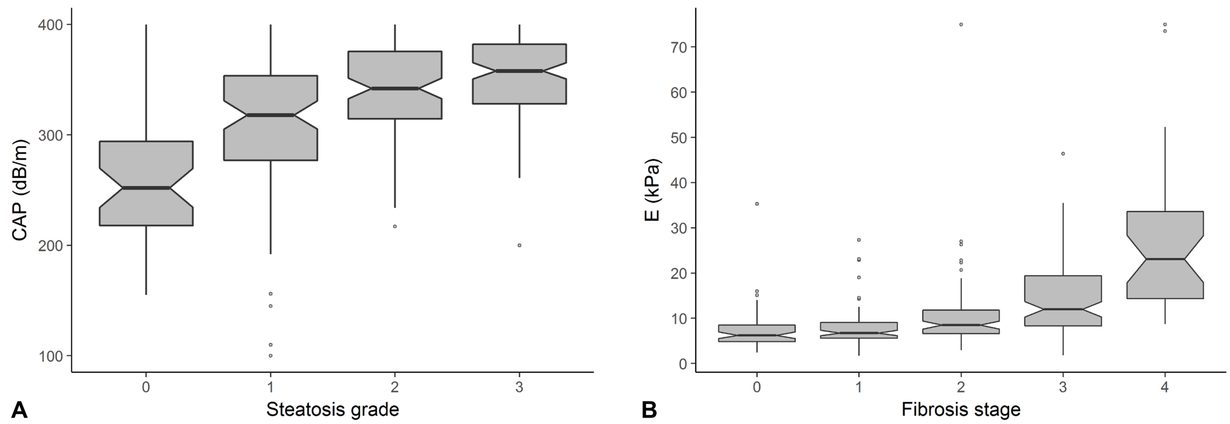
AUROC: area under the receiver operating curve, CI: confidence interval, FN: number of false negative, FP: number of false positive, LR-: negative likelihood ratio, LP+: positive likelihood ratio, NPV: negative predictive value, PPV: positive predictive value, Se: sensitivity, Sp: specificity, TN: true negative, TP: true positive.

Table 4. Impact of prevalence of F≥F2, F≥F3 and F=4 on positive predictive value (PPV) and negative predictive value (NPV) together with their (95% confidence interval) of LSM for the cutoff for Se=0.90, for the Youden index cutoff and for the cutoff for Sp=0.90.

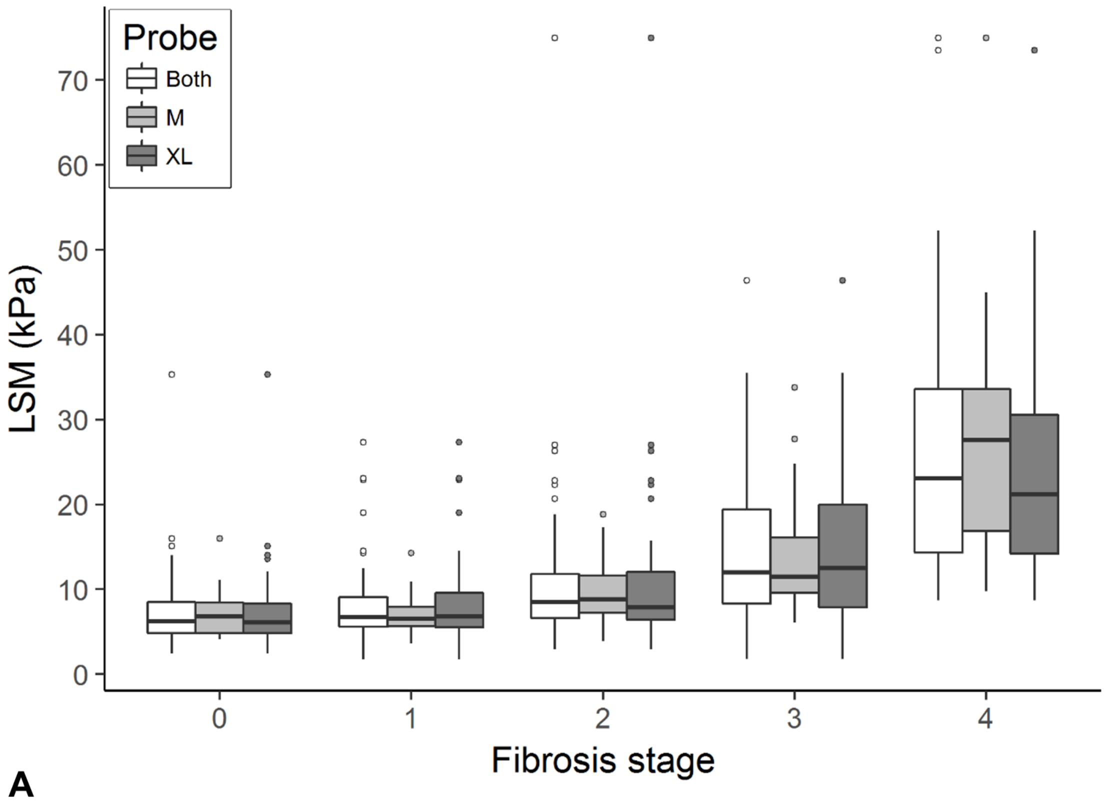
	Prevalence	Justification	Cutoff for Se=0.90	Youden index cutoff	Cutoff for Se=0.90
Diagnostic of F≥F2	-	-	*Cutoff = 6.1 kPa*	*Cutoff = 8.2 kPa*	*Cutoff = 12.1 kPa*
	60%	Actual prevalence in our population	PPV=69% (66%-71%) NPV=72% (62%-80%)	PPV=78% (73%-82%) NPV=61% (56%-67%)	PPV=88% (81%-92%) NPV=52% (49%-55%)
	40%	Estimated prevalence in diabetic clinic [42]	PPV=49% (46%-53%) NPV=85% (79%-90%)	PPV=61% (54%-67%) NPV=78% (74%-82%)	PPV=76% (65%-84%) NPV=71% (68%-74%)
	7%	Estimated prevalence in general population [40]	PPV=10% (9%-11%) NPV=98% (97%-99%)	PPV=15% (12%-18%) NPV=97% (96%-98%)	PPV=26% (17%-37%) NPV=96% (95%-96%)
Diagnostic of F≥F3	-	-	*Cutoff = 7.1 kPa*	*Cutoff = 9.7 kPa*	*Cutoff = 14.1 kPa*
	38%	Actual prevalence in our population	PPV = 52% (45%-67%) NPV = 89% (83%-92%)	PPV = 63% (55%-71%) NPV = 81% (74%-85%)	PPV = 74% (65%-80%) NPV = 74% (67%-82%)
	18%	Estimated prevalence in diabetic clinic [42]	PPV=28% (24%-32%) NPV=96% (92%-98%)	PPV=38% (30%-46%) NPV=92% (89%-94%)	PPV=52% (37%-66%) NPV=89% (87%-91%)
	2%	Estimated prevalence in general population [41]	PPV=4% (3%-4%) NPV=99.6% (99.2%-99.8%)	PPV=5% (4%-7%) NPV=99.2% (98.9%-99.4%)	PPV=9% (5%-15%) NPV=98.8% (98.6%-99.1%)

Diagnostic of F=F4	-	-	*Cutoff = 10.9 kPa*	*Cutoff = 13.6 kPa*	*Cutoff = 20.9 kPa*
	9%	Actual prevalence in our population	PPV=23% (20%-26%) NPV=98.7% (96.5%-99.6%)	PPV=28% (24%-34%) NPV=98.2% (96.0%-99.2)	PPV=37% (27%-47%) NPV=95.7% (93.7%-97.1%)
	3%	Estimated prevalence in population at risk of liver disease [41]	PPV=8% (7%-10%) NPV=99.6% (98.9%-99.9%)	PPV=11% (9%-14%) NPV=99.4% (98.7%-99.8%)	PPV=15% (11%-22%) NPV=98.6% (97.9%-99.1%)
	1%	Estimated prevalence in general population [41]	PPV=3% (2%-4%) NPV=99.9% (99.6%-100%)	PPV=4% (3%-5%) NPV=99.8% (99.6%-99.9%)	PPV=6% (4%-8%) NPV=99.5% (99.3%-99.7%)

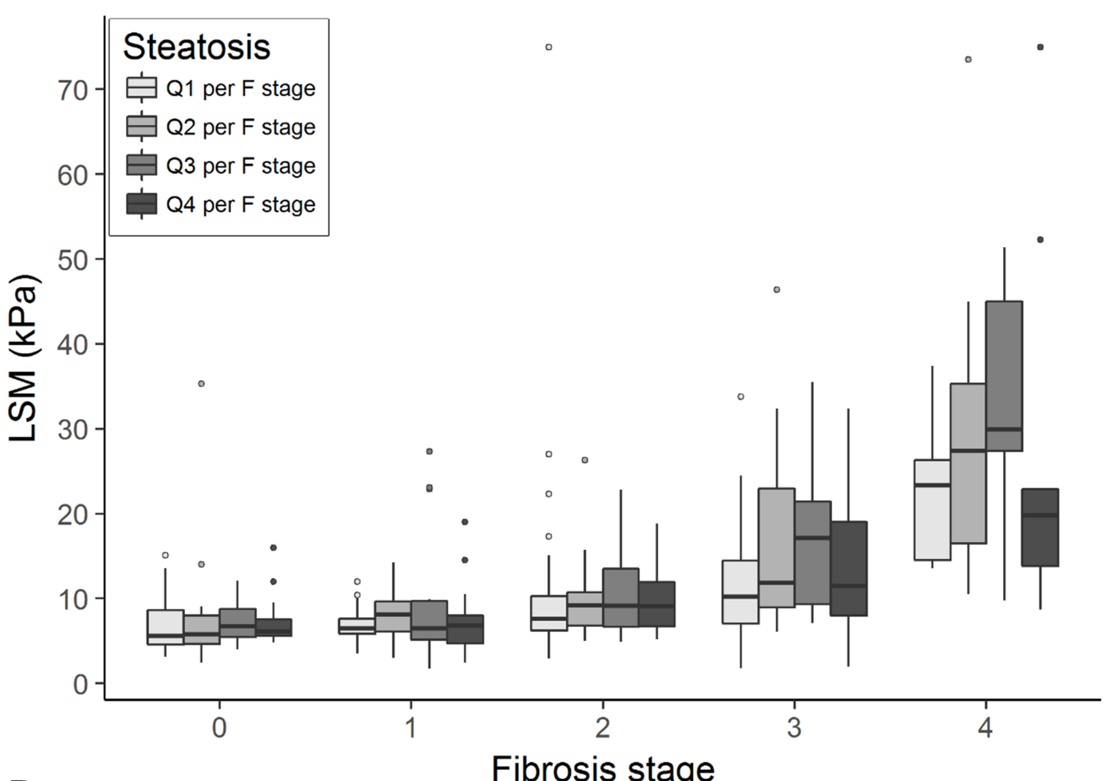

N=450 potentially eligible participants
N=13 screen failure
N=2 didn't fulfil all inclusion criteria, liver biopsy not scheduled within 2 weeks of FibroScan
N=1 didn't fulfil all inclusion criteria, patient with hepatis C
N=1 with exclusion criteria, patient with ascites
N=1 with exclusion criteria, refused the liver biopsy
N=1 with exclusion criteria, refused to undergo a blood test
N=4 with exclusion criteria, alcohol consumption above limits
N=1 with exclusion criteria, confirmed diagnosis of active malignancy
N=1 with exclusion criteria, cardiac failure and/or significant valvular disease
N=1 did not sign the consent
N=4 a posteriori excluded
liver biopsy and FibroScan were not performed within 2 weeks
N=433 eligible participants
N=2 no FibroScan only
patients didn't come to appointments
N=16 no FibroScan and no LB
N=6 withdrew consent
N=2 out of time: appointments could not be scheduled appropriately
N=5 patients didn't come to appointments
N=1 investigator decision, no longer eligible for the study
N=2 serious adverse event (not related to the study device)
N=415 with FibroScan performed
N=404 with valid FibroScan
N=11 with invalid FibroScan
N=4 no liver biopsy
investigator cancelled biopsy
N=3 no liver biopsy results
slides lost
N=14 biopsy non interpretable
N=1 no liver biopsy
patient withdrew consent
N=1 biopsy non interpretable
N=9 with invalid FibroScan & biopsy results
N=383 with valid FibroScan & biopsy results
N=3 no CAP
FibroScan file lost or corrupted
N=10 fibrosis not staged according to NASH CRN scoring system
biopsy interpretable but fibrosis lesion patterns non compatible with NASH CRN score
N=380 for steatosis diagnostic analysis
N=373 for fibrosis staging analysis



CAP (dB/m)
400
300
200
100
0
1
2
3
Steatosis grade
A
E (kPa)
70
60
50
40
30
20
10
0
0
1
2
3
4
Fibrosis stage
B


Probe
Both
M
XL
LSM (kPa)
0
10
20
30
40
50
60
70
0
1
2
3
4
Fibrosis stage

A


Steatosis
Q1 per F stage
Q2 per F stage
Q3 per F stage
Q4 per F stage
LSM (kPa)
0
10
20
30
40
50
60
70
0
1
2
3
4
Fibrosis stage

B

Supplementary material

Supplementary Methods

Influence of the consequences of diagnostic error and of disease prevalence on LSM cut-offs: further analysis on cut-offs was performed for the diagnostic of F≥F2 and F=F4 to take into account the consequences of incorrect classifications on the diagnosis and the disease prevalence. This can be achieved by finding the cut-off value (C) that minimizes the misclassification-cost term[1, 2]:

$$MCT(C) = \frac{C_{FN}}{C_{FP}} P\big(1 - Se(C)\big) + (1 - P)(1 - Sp(C)) \qquad \text{(Eq. 3)}$$

where: C_{FN} is the cost associated with a false negative (FN), C_{FP} is the cost associated with a false positive (FP), *P* is the prevalence. Of note $P(1 - Se)$ is the probability of false-negative Prob(FN) and $(1 - P)(1 - Sp)$ is the probability of false-positive Prob(FP). For the diagnostic of F≥F2, a FP is worse than a FN, therefore we computed the cut-off value for a cost of an FP 2 times, 5 times and 10 times the cost of an FN[1, 2]. For the diagnostic of F=4, a FN is worse than a FP, therefore we computed the cut-off value for a cost of an FN 2 times, 5 times and 10 times the cost of an FP. Finally, we assessed the impact of disease prevalence on the computed cut-offs by varying the prevalence in (Eq. 1) from 5% to 70% for F≥F2 and from 0% to 10% for F=F4.

Supplementary Results

Using clinical consequences to determine optimal cut-offs

Understanding the consequences of diagnostic error, which will vary depending on the clinical setting, can make a major impact on the choice of cut-offs. In Supplementary Table 11 we modelled several scenarios for a diagnosis of F≥F2 and then for F=F4. In a low prevalence setting there may be a greater priority on reducing false positive rate and thus we examined scenarios where the cost of a false positive (FP) was 2 times, 5 times and 10 times worse than a false negative (FN). In another setting there may be prioritisation on not missing a patient with cirrhosis, and here the cost of a FN 2 times, 5 times and 10 times worse than a FP. The impact of the prevalence on those computed cut-offs is given in Supplementary Figure 3 by varying the prevalence from 5% to 70% for F≥F2 and from 0% to 10% for F=F4.

Supplementary Table 1. Standards for Reporting of Diagnostic Accuracy (STARD) check-list

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)	6
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	8-9
	4	Study objectives and hypotheses	9
METHODS			
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	10
Participants	6	Eligibility criteria	11
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	10
	8	Where and when potentially eligible participants were identified (setting, location and dates)	10

	9	Whether participants formed a consecutive, random or convenience series	10
Test methods	10a	Index test, in sufficient detail to allow replication	13-14
	10b	Reference standard, in sufficient detail to allow replication	12-13
	11	Rationale for choosing the reference standard (if alternatives exist)	10
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	14-16
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	12-13
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	13
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	12
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	14-15
	15	How indeterminate index test or reference standard results were handled	16
	16	How missing data on the index test and reference standard were handled	16
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	14-16
	18	Intended sample size and how it was determined	14

RESULTS			
Participants	19	Flow of participants, using a diagram	Figure 1
	20	Baseline demographic and clinical characteristics of participants	Table 1
	21a	Distribution of severity of disease in those with the target condition	18-19 & Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	Table 1
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Table 2 & Table 3
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Table 2 & Table 3
	25	Any adverse events from performing the index test or the reference standard	17-18
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	23-26
	27	Implications for practice, including the intended use and clinical role of the index test	23-26
OTHER INFORMATION			
	28	Registration number and name of registry	7 & 10
	29	Where the full study protocol can be accessed	Available upon request to the

			corresponding author
	30	Sources of funding and other support; role of funders	2

Supplementary Table 2. Histological description of patients with histological diagnoses other than NAFLD or normal liver (including those for whom it was not possible to stage fibrosis according to the NASH CRN scoring system (F=NA)).

Number of cases	Pathology diagnosis	Pathology comment	SAF score*
4	Cryptogenic cirrhosis	Burnt out NASH or other aetiology	N=3: S0A0F4 N=1: S0A1F4
2	Inflammatory cirrhosis	Other disease	N=2: S0A2F4
7	Fibrosis without any sign of NAFLD	Burnt out NASH or other aetiology	N=3: S0A0F2 N=3: S0A0F3 N=1: S0A1F3
3	NAFLD and associated lesions	Granuloma or lesions suggesting active chronic hepatitis	N=1: S2A2F3 N=1: S1A1F$_{=NA}$ N=1: S1A2F$_{=NA}$
17	Not NAFLD but not normal liver	Inflammatory lesion or other cause. None have steatosis, all have portal inflammation	N=1: S0A0F0 N=2: S0A0F1 N=1: S0A1F0 N=4: S0A1F1 N=1: S0A1F2 N=1: S0A0F$_{=NA}$ N=1: S0A0F$_{=NA}$ N=2: S0A1F$_{=NA}$ N=2: S0A1F$_{=NA}$ N=1: S0A1F$_{=NA}$ N=1: S0A2F$_{=NA}$

*: SAF score is given in patients for whom fibrosis could be staged. For others, only steatosis and activity grade are given, fibrosis stage is mentioned as F = NA.

A: activity, F: fibrosis, NA: not applicable, NAFLD: non-alcoholic fatty liver disease, NASH: non-alcoholic steato-hepatitis, S: steatosis.

Supplementary Table 3: AUROC (95% CI) for the diagnostic of steatosis grade ≥ 1, ≥ 2 and = to 3 when dichotomizing patients by their IQR of CAP value (< and ≥ 30 dB/m and (< and ≥ 40 dB/m). P-value corresponds to the AUROC comparison using Delong test.

	N (proportion)	*AUROC (95%CI) for S≥S1*	*P value*	*AUROC (95%CI) for S≥S2*	*P value*	*AUROC (95%CI) for S=S3*	*P value*
IQR CAP<30	164 (43%)	0.88 (0.77-0.99)	0.60	0.69 (0.59-0.79)	0.09	0.64 (0.56-0.73)	0.01
IQR CAP≥30	216 (57%)	0.85 (0.78-0.91)		0.80 (0.74-0.85)		0.78 (0.72-0.84)	
IQR CAP<40	232 (61%)	0.87 (0.79-0.95)	0.91	0.74 (0.66-0.82)	0.51	0.68 (0.61-0.74)	0.07
IQR CAP≥40	148 (39%)	0.85 (0.78-0.91)		0.80 (0.74-0.85)		0.78 (0.72-0.84)	

Supplementary Table 4: Diagnostic performance of controlled attenuation parameter (CAP) for each steatosis grade ≥1, ≥2 and = to 3 stratified by ALT value.

	S≥1	**S≥2**	**S=3**
Stratum 1 **AUROC (95%CI) for ALT≤ULN**	0.86 (0.80-0.93)	0.74 (0.67-0.82)	0.73 (0.65-0.81)
	Pr=0.86	*Pr=0.56*	*Pr=0.22*
Stratum 2 **AUROC (95%CI) for ULN>ALAT≤2*ULN**	0.87 (0.73-1.00)	0.78 (0.68-0.88)	0.69 (0.60-0.77)
	Prevalence=0.92	*Prevalence=0.75*	*Prevalence=0.49*
Stratum 3 **AUROC (95%CI) for ALAT>2*ULN**	0.95 (0.88-1.00)	0.84 (0.68-1.00)	0.67 (0.48-0.85)
	Prevalence=0.95	*Prevalence=0.80*	*Prevalence=0.55*
AUROC comparison	Stratum 1/2: P=0.89 Stratum 1/3: P=0.08 Stratum 2/3: P=0.34	Stratum 1/2: P=0.53 Stratum 1/3: P=0.29 Stratum 2/3: P=0.55	Stratum 1/2: P=0.47 Stratum 1/3: P=0.54 Stratum 2/3: P=0.83

Supplementary Table 5: AUROC of CAP, HSI and FLI for the diagnosis of S≥1, S≥2 and S=3. P value corresponds to the AUROC comparison with CAP AUROC using Delong test.

		S≥1	**S≥2**	**S=3**
CAP	AUROC	0.87 (0.81-0.92)	0.77 (0.72-0.83)	0.70 (0.65-0.75)
HSI	AUROC	0.63 (0.55-0.71)	0.63 (0.58-0.69)	0.59 (0.53-0.65)
	P value	$<10^{-8}$	$<10^{-5}$	0.01

Supplementary Table 6. Comparison of patient characteristics between the 10 patients with fibrosis not staged according to the NASH CRN scoring system and the 373 patients used for fibrosis staging analysis (Figure 1).

All bio-clinical parameters from Table 1 were tested. Only those with a P-value < 0.20 for the comparison are represented in the table. Distribution is expressed as median [interquartile range] or figure (percentage). Comparison was performed using Mann-Whitney U test for continuous variables and using $\chi 2$ test or Fisher-exact test, as applicable for binary or categorical variables.

Characteristic	N=10 Patients with fibrosis not staged according to NASH CRN	N=373 Patients with fibrosis staged according to NASH CRN	P-value
Centre	Birmingham: 1 (10%) Newcastle: 1 (10%) London: 0 (0%) Nottingham: 0 (0%) Plymouth: 6 (60%) Cambridge: 0 (0%) Oxford: 2 (20%)	Birmingham: 101 (27%) Newcastle: 50 (13%) London: 52 (14%) Nottingham: 40 (11%) Plymouth: 42 (11%) Cambridge: 60 (16%) Oxford: 28 (8%)	$<10^{-3}$
Female gender	8 (80%)	163 (44%)	0.05
Alkaline phosphatase (IU/L)	161 [100]	81 [38]	0.006
Fasting glucose (mg/dL)	87 [15]	107 [52]	0.02
HDL cholesterol (mg/dL)	54 [14]	43 [17]	0.06

Ferritin (ng/mL)	111 [92]	135 [216]	0.15
Creatinine (mg/dL)	0.76 [0.14]	0.86 [0.22]	0.05
CK18-M30 (IU/L)	310 [210]	416 [402]	0.09
CAP (dB/m)	242 [63]	337 [73]	$<10^{-3}$
Steatosis grade	S0: 8 (80%) S1: 2 (20) S2: 0 (0%) S3: 0 (0%)	S0: 39 (10%) S1: 87 (23%) S2: 109 (29%) S3: 138 (37%)	$<10^{-6}$
Ballooning grade	B0: 9 (90%) B1: 1 (10%) B2: 0 (0%)	B0: 97 (26%) B1: 146 (39%) B2: 130 (35%)	$<10^{-4}$
NAS score	0-2: 8 (80%) 3-4: 2 (20%) 5-8: 0 (0%)	0-2: 82 (22%) 3-4: 120 (32%) 5-8: 171 (46%)	$<10^{-4}$
Activity grade	A0: 2 (20%) A1: 6 (60%) A2: 2 (20%) A3: 0 (0%) A4: 0 (0%)	A0: 53 (14%) A1: 74 (20%) A2: 100 (27%) A3: 110 (29%) A4: 36 (10%)	0.02
Portal inflammation present	10 (100%)	162 (44%)	$<10^{-3}$
Pathologists diagnostic	Normal liver: 0 (0%)	Normal liver: 17 (5%)	$<10^{-10}$

	NAFL: 0 (0%) NASH: 0 (0%) Other: 10 (100%)	NAFL: 91 (24%) NASH: 242 (65%) Other: 23 (6%)	

GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, NAFL: non-alcoholic fatty liver, NAFLD: NAFL disease, NASH: non-alcoholic steato-hepatitis, NAS: NAFLD Activity Score.

Supplementary Table 7: AUROC (95% CI) for the diagnosis of fibrosis stage ≥2, ≥3 and = to 4 in all patients and patients with Fibroscan fulfilling Boursier's criteria[3].

	N	F≥F2	F≥F3	F=F4
Patients with Fibroscans fulfilling Boursier's criteria	331	AUROC=0.78 (0.73-0.83) *Prevalence=0.70*	AUROC=0.80 (0.75-0.86) *Prevalence=0.53*	AUROC=0.90 (0.86-0.95) *Prevalence=0.07*

Supplementary Table 8: Diagnostic performance of liver stiffness measurement (LSM) for each fibrosis stage ≥2, ≥3 and = to 4 stratified by ALT value. The AUROC comparison was performed using Delong test.

	F≥2	**F≥3**	**F=4**
Stratum 1 **AUROC (95%CI) for ALT≤ULN**	0.80 (0.71-0.88)	0.81 (0.72-0.90)	0.87 (0.78-0.95)
	Prevalence=0.46	*Prevalence =0.34*	*Prevalence =0.10*
Stratum 2 **AUROC (95%CI) for ULN>ALAT≤2*ULN**	0.75 (0.67-0.83)	0.77 (0.70-0.85)	0.93 (0.89-0.98)
	Prevalence =0.64	*Prevalence =0.40*	*Prevalence =0.09*
Stratum 3 **AUROC (95%CI) for ALAT>2*ULN**	0.79 (0.69-0.89)	0.81 (0.71-0.90)	0.86 (0.72-0.99)
	Prevalence =0.68	*Prevalence =0.36*	*Prevalence =0.07*
AUROC comparison	Stratum 1/2: P=0.41 Stratum 1/3: P=0.93 Stratum 2/3: P=0.51	Stratum 1/2: P=0.58 Stratum 1/3: P=0.96 Stratum 2/3: P=0.62	Stratum 1/2: P=0.17 Stratum 1/3: P=0.90 Stratum 2/3: P=0.30

Supplementary Table 9: AUROC of LSM, Fib4 and NFS for the diagnosis of fibrosis stage ≥2, ≥3 and = to 4. P value corresponds to AUROC comparison with LSM AUROC using Delong test.

		F≥2	F≥3	F=4
LSM	**AUROC**	0.77 (0.72-0.82)	0.80 (0.75-0.85)	0.90 (0.85-0.94)
Fib4	**AUROC**	0.72 (0.67-0.78)	0.78 (0.73-0.83)	0.84 (0.78-0.91)
	P value	0.09	0.31	0.03
NFS	**AUROC**	0.69 (0.63-0.74)	0.75 (0.70-0.80)	0.81 (0.73-0.90)
	P value	0.006	0.07	0.01

Supplementary Table 10: Performance comparison of LSM, Fib4 and NFS using dual-cutoff approach for the diagnosis of advanced fibrosis (F≥3).

LSM			Fib4			NFS		
Lower cut-off (< 7.1 kPa)	Grey zone	Upper cut-off (≥14.1 kPa)	Lower cut-off (<1.30)	Grey zone	Upper cut-off (≥3.25)	Lower cut-off (<-1.455)	Grey zone	Upper cut-off (≥0.676)
N = 127 (35%)	N = 148 (41%)	N = 87 (24%)	N = 209 (58%)	N = 131 (36%)	N = 22 (6%)	N = 153 (42%)	N = 170 (47%)	N = 39 (11%)
Sp=0.50	–	Se=0.48	Sp=0.73	–	Se=0.14	Sp=0.56	–	Se=0.22
NPV=0.90		PPV=0.74	NPV=0.80		PPV=0.86	NPV=0.84		PPV=0.74
F<3: 114 (50%)	F<3: 93 (41%)	F<3: 22 (10%)	F<3: 168 (73%)	F<3: 58 (25%)	F<3: 3 (1%)	F<3: 128 (56%)	F<3: 91 (40%)	F<3: 10 (4%)
F≥3: 13 (10%)	F≥3: 55 (41%)	F≥3: 65 (49%)	F≥3: 41 (31%)	F≥3: 73 (55%)	F≥3: 19 (14%)	F≥3: 25 (19%)	F≥3: 79 (59%)	F≥3:29 (22%)

Supplementary Table 11: Diagnostic performance of liver stiffness measurement (LSM) taking into account the consequences of diagnostic error: for the diagnostic of F≥F2 with a cost false positive (FP) 2 times, 5 times and 10 times worse than a false negative (FN); for the diagnostic of F=F4 with a cost FN 2 times, 5 times and 10 times worse than a FP.

	F≥F2		F=F4	
Cut-off	FP 2 times worse than FN	10.3	FN 2 times worse than FP	27.4
Se / Sp		Se=0.55 / Sp=0.85		Se=0.41 / Sp=0.97
PPV / NPV		PPV=0.85 / NPV=0.55		PPV=0.61 / NPV=0.94
LR+ / LR-		LR+=3.68 / LR-=0.53		LR+=15.51 / LR-=0.60
CC		0.67		0.92
FP / FN		FP=22 / FN=102		FP=9 / FN=20
Cut-off	FP 5 times worse than FN	16.8	FN 5 times worse than FP	19.8
Se / Sp		Se=0.30 / Sp=0.96		Se=0.65 / Sp=0.89
PPV / NPV		PPV=0.92 / NPV=0.47		PPV=0.37 / NPV=0.96

LR+ / LR-		LR+=7.35 / LR-=0.73		LR+=5.93 / LR-=0.40
CC		0.56		0.87
FP / FN		FP=6 / FN=158		FP=37 / FN=12
Cut-off	FP 10 times worse than FN	23.3	FN 10 times worse than FP	C=13.6
Se		Se=0.15 / Sp=0.99		Se=0.85 / Sp=0.79
PPV / NPV		PPV=0.94 / NPV=0.43		PPV=0.29 / NPV=0.98
LR+ / LR-		LR+=11.18 / LR-=0.86		LR+=4.02 / LR-=0.19
CC		0.48		0.79
FP / FN		FP=2 / FN=191		FP=72 / FN=5

AUROC: area under the receiver operating curve, CC: proportion of correctly classified, F: fibrosis, FN: number of false negative, FP: number of false positive, LR-: negative likelihood ratio, LP+: positive likelihood ratio, NPV: negative predictive value, PPV: positive predictive value, Se: sensitivity, Sp: specificity.

Supplementary Table 12: Impact of rounding cut-offs from Table (Impact of prevalence of F≥F2, F≥F3 and F=4) on positive predictive value (PPV) and negative predictive value (NPV) of LSM for cut-offs for Se=0.90, Youden index cutoff and Sp=0.90).

	Prevalence	Justification	Cutoff for Se=0.90		Youden index cutoff		Cutoff for Se=0.90	
Diagnosis of F≥F2	-	-	*Cutoff = 6.1 kPa*	*Rounded cutoff = 6.0 kPa*	*Cutoff = 8.2 kPa*	*Rounded cutoff = 8.0 kPa*	*Cutoff = 12.1 kPa*	*Rounded cutoff = 12.0 kPa*
	60%	Actual prevalence in our population	PPV=69% / NPV=72%	PPV=69% / NPV=72%	PPV=78% / NPV=61%	PPV=77% / NPV=61%	PPV=88% / NPV=52%	PPV=88% / NPV=52%
	40%	Estimated prevalence in diabetic clinic [4]	PPV=49% / NPV=85%	PPV=49% / NPV=85%	PPV=61% / NPV=78%	PPV=59% / NPV=78%	PPV=76% / NPV=71%	PPV=76% / NPV=71%
	7%	Estimated prevalence in general population [5]	PPV=10% / NPV=98%	PPV=10% / NPV=98%	PPV=15% / NPV=97%	PPV=14% / NPV=97%	PPV=26% / NPV=96%	PPV=26% / NPV=96%
Diagnosis	-	-	*Cutoff = 7.1 kPa*	*Rounded cutoff = 7.0 kPa*	*Cutoff = 9.7 kPa*	*Rounded cutoff = 10.0 kPa*	*Cutoff = 14.1 kPa*	*Rounded cutoff = 14.0 kPa*

of F≥F3	38%	Actual prevalence in our population	PPV=52% / PPV= 89%	PPV=52% / PPV=89%	PPV=63% / NPV=81%	PPV=63% / NPV=79%	PPV = 74% / NPV = 74%	PPV=74% / NPV=74%
	18%	Estimated prevalence in diabetic clinic [4]	PPV=28% / NPV=96%	PPV=28% / PPV=96%	PPV=38% / NPV=92%	PPV=38% / NPV=91%	PPV=52% / NPV=89%	PPV=52% / NPV=89%
	2%	Estimated prevalence in general population [6]	PPV=4% / NPV=99.6%	PPV=4% / PPV=99.5%	PPV=5% / NPV=99.2%	PPV=5% / NPV=99.1%	PPV=9% / NPV=98.8%	PPV=9% / NPV=98.8%
Diagnosis of F=F4	-	-	*Cutoff = 10.9 kPa*	*Rounded cutoff = 11.0 kPa*	*Cutoff = 13.6 kPa*	*Rounded cutoff = 14.0 kPa*	*Cutoff = 20.9 kPa*	*Rounded cutoff = 21.0 kPa*
	9%	Actual prevalence in our population	PPV=23% / NPV=98.7%	PPV=23% / NPV =98.3%	PPV=28% / NPV=98.2%	PPV=29 / NPV=97.2%	PPV=37% / NPV=95.7%	PPV=38% / NPV=95.6%
	3%	Estimated prevalence in population at risk of liver disease [6]	PPV=8% / NPV=99.6%	PPV = 8% / NPV=99.5%	PPV=11% / NPV=99.4%	PPV=11% / NPV=99.1%	PPV=15% / NPV=98.6%	PPV=16% / NPV=98.6%
	1%	Estimated prevalence	PPV=3% /	PPV=3% /	PPV=4% /	PPV=4% /	PPV=6% /	PPV=6% /

		in general population [6]	NPV=99.9%	NPV=99.8%	NPV=99.8%	NPV=99.7%	NPV=99.5%	NPV=99.5%

Supplementary Table 13: Comparison of the main results from Siddiqui et al.[7] and from the present study.

			Present study	Siddiqui et al.[7] study
Patients main characteristics*		*N*	384	398
		Age (year)	54 [18]	51±11
		Female gender	45%	68%
		BMI (kg.m-2)	33.8 [9.2]	34.4±6.4
		AST (IU/L)	36 [25]	49±37
		ALT (IU/L)	50 [40]	64±44
Diagnostic performance of LSM	***F≥F2***	*Prevalence*	0.60	0.51
		AUROC (95% CI)	0.77 (0.72-0.82)	0.79 (0.74-0.83)
		Cut-off for Youden's index	8.2 kPa	8.6 kPa
		Cut-off for Se=0.90	6.1 kPa	5.6 kPa
		Cut-off for Sp=0.90	12.1 kPa	11.9 kPa
	F≥F3	*Prevalence*	0.38	0.32
		AUROC (95% CI)	0.80 (0.75-0.84)	0.83 (0.79-0.87)
		Cut-off for Youden's index	9.7 kPa	8.6 kPa
		Cut-off for Se=0.90	7.1 kPa	6.5 kPa
		Cut-off for Sp=0.90	14.1 kPa	12.1 kPa
	F=F4	*Prevalence*	0.09	0.09
		AUROC (95% CI)	0.89 (0.84-0.93)	0.93 (0.90-0.97)
		Cut-off for Youden's index	13.6 kPa	13.1 kPa
		Cut-off for Se=0.90	10.9 kPa	12.1 kPa
		Cut-off for Sp=0.90	20.9 kPa	14.9 kPa
Diagnostic performance of CAP	***S≥S1***	*Prevalence*	0.88	0.95
		AUROC (95% CI)	0.87 (0.82-0.92)	0.76 (0.64-0.89)
		Cut-off for Youden's index	302 dB/m	285 dB/m
		Cut-off for Se=0.90	274 dB/m	263 dB/m
		Cut-off for Sp=0.90	325 dB/m	353 dB/m
	S≥S2	*Prevalence*	0.64	0.58
		AUROC (95% CI)	0.77 (0.71-0.82)	0.70 (0.64-0.75)
		Cut-off for Youden's index	331 dB/m	311 dB/m

		Cut-off for Se=0.90	290 dB/m	280 dB/m
		Cut-off for Sp=0.90	370 dB/m	367 dB/m
	S=S3	*Prevalence*	0.36	0.27
		AUROC (95% CI)	0.70 (0.64-0.75)	0.58 (0.51-0.64)
		Cut-off for Youden's index	337 dB/m	306 dB/m
		Cut-off for Se=0.90	302 dB/m	274 dB/m
		Cut-off for Sp=0.90	398 dB/m	380 dB/m

*: results are given as median [inter-quartile range] for the present study and as mean±standard deviation for the Siddiqui et al.[7] study.

AUROC: area under the receiver operating curve, ALT: alanine transaminase, AST: aspartate aminotransferase, BMI: body mass index, CAP: controlled attenuation parameter, CI: confidence interval, F: fibrosis, LSM: liver stiffness measurement, S: steatosis, Se: sensitivity, Sp: specificity.

Supplementary Table 14: Published Youden cutoffs in NAFLD studies, except for Siddiqui et al[7]

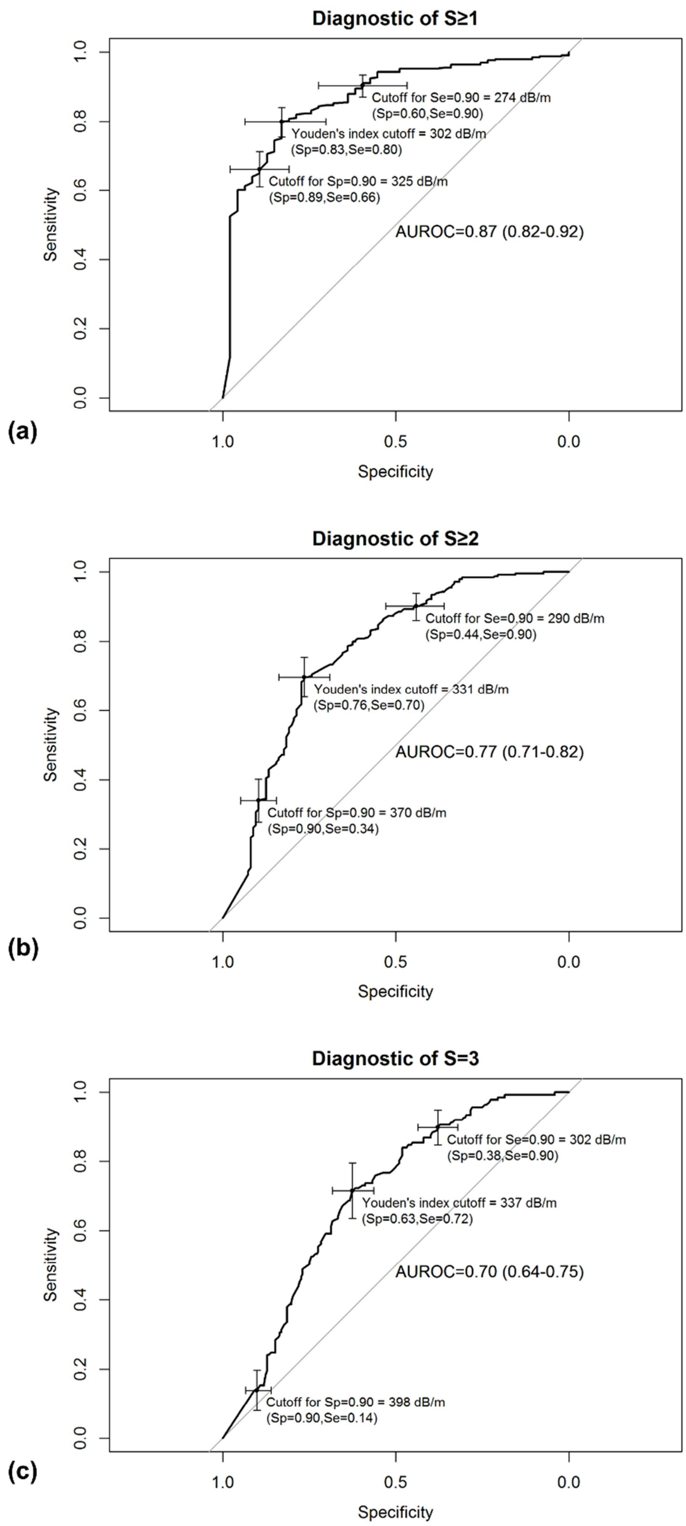
Reference	N	BMI (kg.m^{-2})	Probe usage	Diagnostic target	Prevalence	AUC	Youden cutoff (kPa)	Se/Sp
Chen et al.[8]	111	40.3	M or XL probe according to manufacturer's recommendations.	F≥F2	0.36	0.91	7.8	82/78
				F≥F3	0.20	0.87	7.6	84/64
Imajo et al.[9]	127	28.1	M only	F≥F2	0.54	0.82	11.0	65/89
				F≥F3	0.32	0.88	11.4	86/84
				F=F4	0.08	0.92	14.0	100/76
Petta et al.[10]	324	40% of patients >30	M only	F≥F2	0.58	0.81	8.5	74/74
				F≥F3	0.36	0.86	10.1	78/78
Kumar et al.[11]	120	26.1	M only	F≥F2	0.45	0.85	7.0	77/78
				F≥F3	0.23	0.94	9.0	85/88
				F=F4	0.08	0.96	11.8	90/88
Naveau et al.[12]	100	42.3	M or XL probe according to manufacturer's recommendations	F≥F2	0.22	0.81	7.6	73/78
				F≥F3	0.09	0.85	7.6	100/74
Mahadeva et al.[13]	120	33% of patients	M only	F≥F2	0.57	0.67	6.9	59/69

		>30		F≥F3	0.22	0.77	7.1	70/67
				F=F4	0.06	0.95	11.3	88/89
Tapper et al.[14]	120	31.3	M only	F≥F3	0.18	0.93	9.9	95/77
Wong et al.[15]	246	28.0	M only	F≥F2	0.41	0.84	7.0	79/76
				F≥F3	0.23	0.93	8.7	84/83
				F=4	0.10	0.95	10.3	92/88
Wong et al.[16]	193	28.9	Both probes used on each patient regardless of manufacturer's recommendations.	F≥F2	0.54	M: 0.83 XL: 0.80	M: 7.0 XL: 6.2	79/64 73/66
				F≥F3	0.33	M: 0.87 XL: 0.85	M: 8.7 XL: 7.2	83/78 78/78
				F=4	0.14	M: 0.89	M: 10.3	81/83

						XL: 0.91	XL: 7.2	92/70

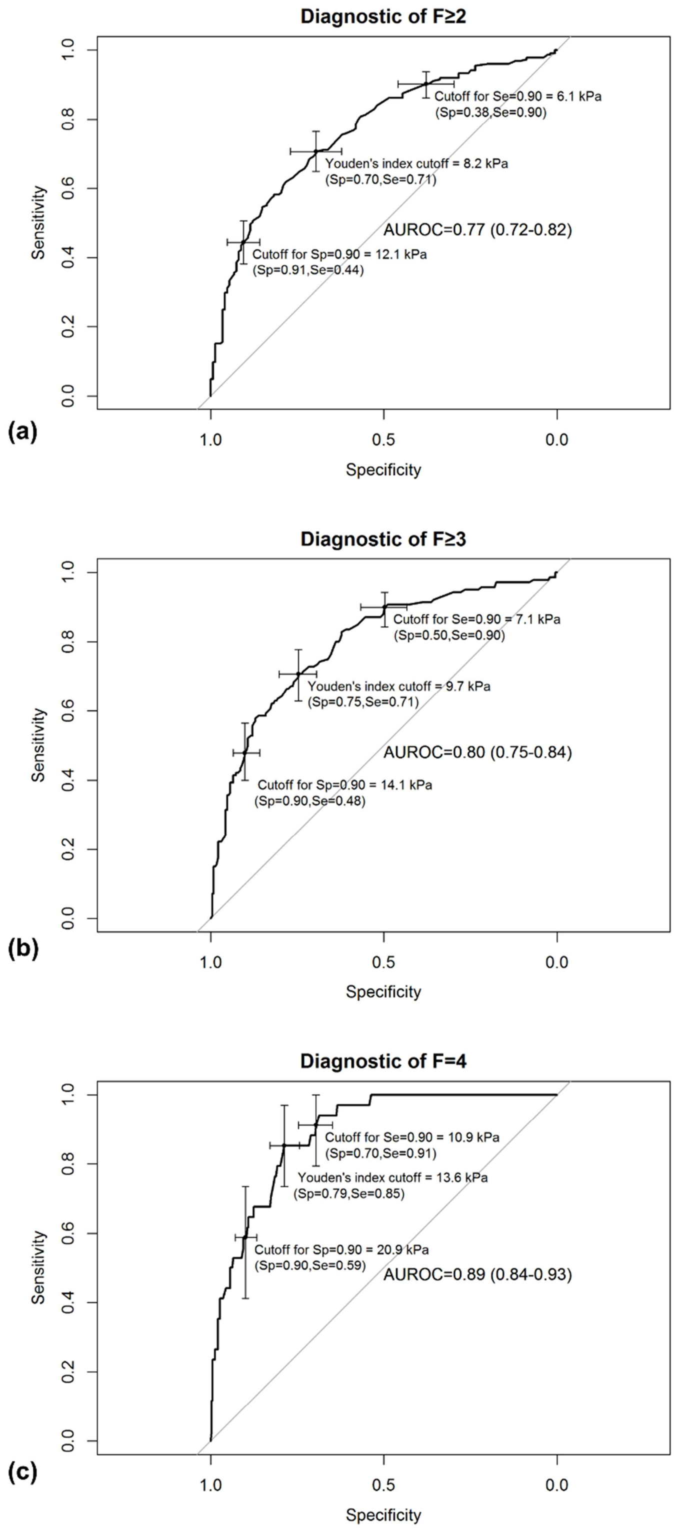
Supplementary Figure 1. Receiver operating characteristic (ROC) curve of controlled attenuation parameter (CAP) for identifying (a) S≥S1, (b) S≥S2 and (c) S=S3.

For each steatosis threshold, are overprinted: area under ROC curve (AUROC) with its 95% CI and the cut-off values maximizing Youden's index, for a fixed sensitivity (Se) and Specificity (Sp) of 0.90.



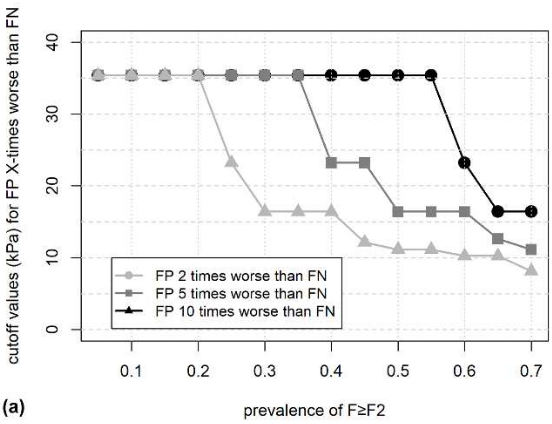
Supplementary Figure 2. Receiver operating characteristic (ROC) curve of liver stiffness measurement (LSM) for identifying (a) F≥F2, (b) F≥F3 and (c) F=F4.

For each fibrosis threshold, are overprinted: area under ROC curve (AUROC) with its 95% CI and the cut-off values maximizing Youden's index, for a fixed sensitivity (Se) and Specificity (Sp) of 0.90.

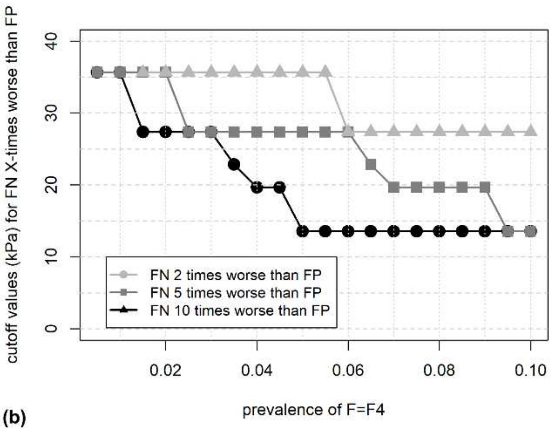


Supplementary Figure 3. Impact of the prevalence on the liver stiffness measurement (LSM) cut-offs computed taking into account the consequences of diagnostic error.

This was undertaken (a) for the diagnostic of F≥F2 with a cost false positive (FP) 2 times, 5 times and 10 times worse than a false negative (FN), (b) for the diagnostic of F=F4 with a cost FN 2 times, 5 times and 10 times worse than a FP. The range of prevalence is 5 to 70% for F≥F2 and 0% to 10% for F=F4, respectively. For F≥F2, for a prevalence up to 20% the cut-offs value is 35.4 kPa. The cut-off value decreases from a prevalence of 20% for a cost for a FP 2 times worse than a FN, from a prevalence of 35% for a cost for a FP 5 times worse than a FN and from 55% for a cost for a FP 10 times worse than a FN. For F=F4, for a prevalence up to 1% the cut-offs value is 35.7 kPa. The cut-off value decreases from a prevalence of 1% for a cost for a FN 10 times worse than a FP, from a prevalence of 5.5% for a cost for a FN 5 times worse than a FP and from 5.5% for a cost for a FN 10 times worse than a FP.



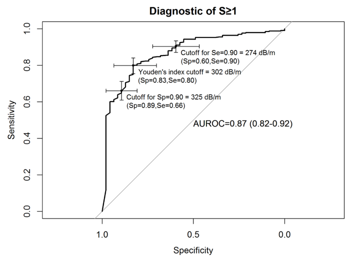
(a)



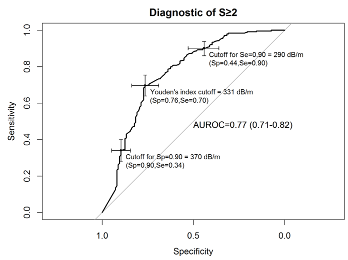
(b)

References

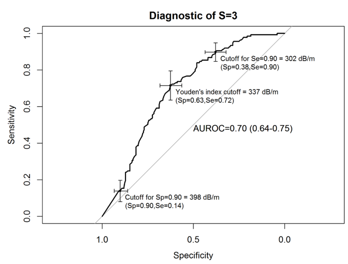
1. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 2000;45:23-41.
2. Bakir M, Dosanjh DP, Deeks JJ, et al. Use of T cell-based diagnosis of tuberculosis infection to optimize interpretation of tuberculin skin testing for child tuberculosis contacts. Clin Infect Dis 2009;48:302-12.
3. Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013;57:1182-91.
4. Kwok R, Choi KC, Wong GL, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut 2016;65:1359-68.
5. Roulot D, Costes JL, Buyck JF, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut 2011;60:977-84.
6. Harris R, Harman DJ, Card TR, et al. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. Lancet Gastroenterol Hepatol 2017;2:288-297.
7. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al. Vibration-controlled Transient Elastography to Assess Fibrosis and Steatosis in Patients With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol 2018.
8. Chen J, Yin M, Talwalkar JA, et al. Diagnostic Performance of MR Elastography and Vibration-controlled Transient Elastography in the Detection of Hepatic Fibrosis in Patients with Severe to Morbid Obesity. Radiology 2017;283:418-428.
9. Imajo K, Kessoku T, Honda Y, et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. Gastroenterology 2016;150:626-637 e7.
10. Petta S, Maida M, Macaluso FS, et al. The severity of steatosis influences liver stiffness measurement in patients with nonalcoholic fatty liver disease. Hepatology 2015;62:1101-10.
11. Kumar R, Rastogi A, Sharma MK, et al. Liver stiffness measurements in patients with different stages of nonalcoholic fatty liver disease: diagnostic performance and clinicopathological correlation. Dig Dis Sci 2013;58:265-74.
12. Naveau S, Lamouri K, Pourcher G, et al. The Diagnostic Accuracy of Transient Elastography for the Diagnosis of Liver Fibrosis in Bariatric Surgery Candidates with Suspected NAFLD. Obes Surg 2014.
13. Mahadeva S, Mahfudz AS, Vijayanathan A, et al. Performance of transient elastography (TE) and factors associated with discordance in non-alcoholic fatty liver disease. J Dig Dis 2013;14:604-10.
14. Tapper EB, Challies T, Nasser I, et al. The Performance of Vibration Controlled Transient Elastography in a US Cohort of Patients With Nonalcoholic Fatty Liver Disease. Am J Gastroenterol 2016;111:677-84.
15. Wong VW, Vergniol J, Wong GL, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454-62.
16. Wong VW, Vergniol J, Wong GL, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862-71.


Diagnostic of S≥1
Cutoff for Se=0.90 = 274 dB/m
(Sp=0.60,Se=0.90)
Youden's index cutoff = 302 dB/m
(Sp=0.83,Se=0.80)
Cutoff for Sp=0.90 = 325 dB/m
(Sp=0.89,Se=0.66)
AUROC=0.87 (0.82-0.92)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.0
0.5
0.0
Specificity

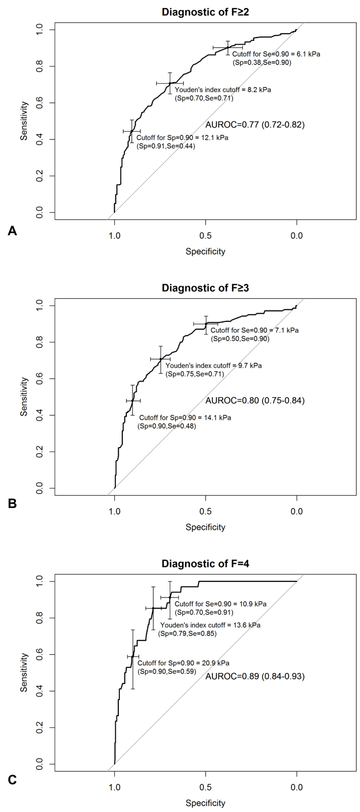
A

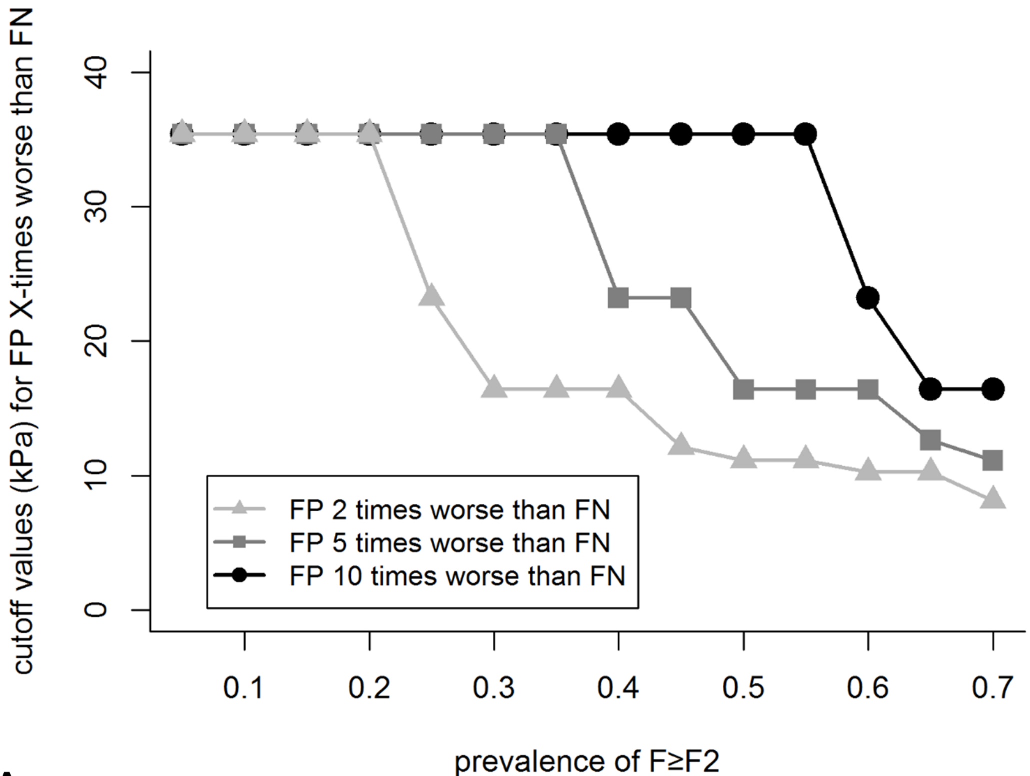

Diagnostic of S≥2
Cutoff for Se=0.90 = 290 dB/m
(Sp=0.44,Se=0.90)
Youden's index cutoff = 331 dB/m
(Sp=0.76,Se=0.70)
AUROC=0.77 (0.71-0.82)
Cutoff for Sp=0.90 = 370 dB/m
(Sp=0.90,Se=0.34)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.0
0.5
0.0
Specificity

B

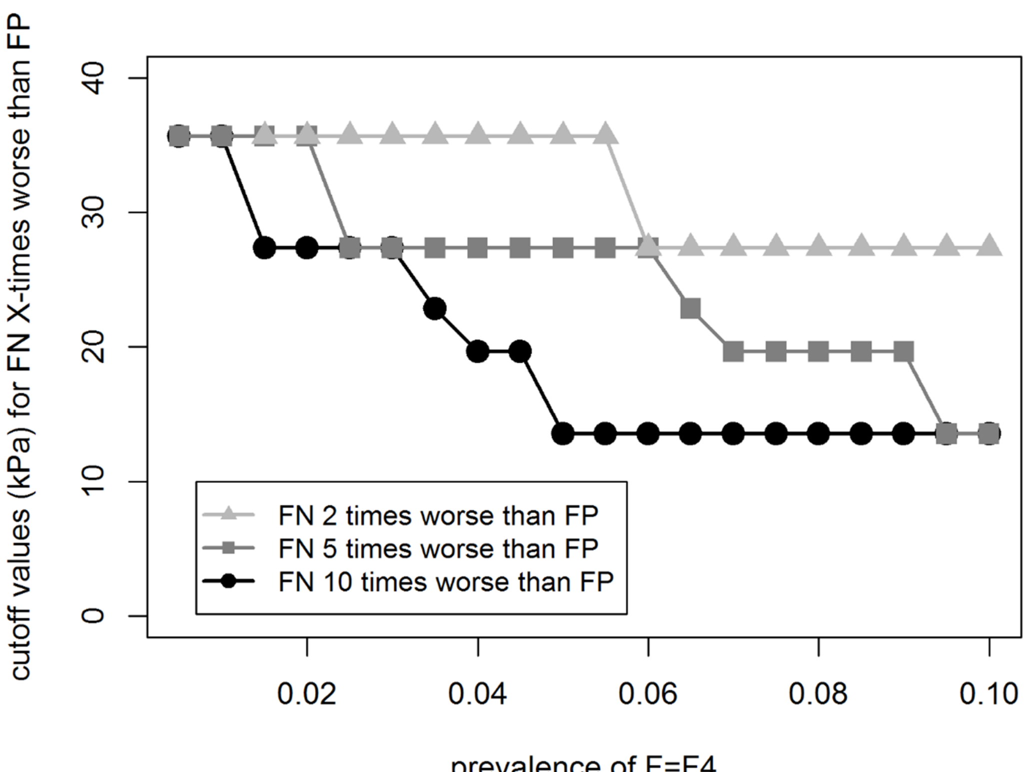

Diagnostic of S=3
Cutoff for Se=0.90 = 302 dB/m
(Sp=0.38,Se=0.90)
Youden's index cutoff = 337 dB/m
(Sp=0.63,Se=0.72)
AUROC=0.70 (0.64-0.75)
Cutoff for Sp=0.90 = 398 dB/m
(Sp=0.90,Se=0.14)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.0
0.5
0.0
Specificity

C


Diagnostic of F≥2
Cutoff for Se=0.90 = 6.1 kPa
(Sp=0.38,Se=0.90)
Youden's index cutoff = 8.2 kPa
(Sp=0.70,Se=0.71)
AUROC=0.77 (0.72-0.82)
Cutoff for Sp=0.90 = 12.1 kPa
(Sp=0.91,Se=0.44)
Sensitivity
Specificity
A
Diagnostic of F≥3
Cutoff for Se=0.90 = 7.1 kPa
(Sp=0.50,Se=0.90)
Youden's index cutoff = 9.7 kPa
(Sp=0.75,Se=0.71)
AUROC=0.80 (0.75-0.84)
Cutoff for Sp=0.90 = 14.1 kPa
(Sp=0.90,Se=0.48)
Sensitivity
Specificity
B
Diagnostic of F=4
Cutoff for Se=0.90 = 10.9 kPa
(Sp=0.70,Se=0.91)
Youden's index cutoff = 13.6 kPa
(Sp=0.79,Se=0.85)
Cutoff for Sp=0.90 = 20.9 kPa
(Sp=0.90,Se=0.59)
AUROC=0.89 (0.84-0.93)
Sensitivity
Specificity
C


cutoff values (kPa) for FP X-times worse than FN
0
10
20
30
40
0.1
0.2
0.3
0.4
0.5
0.6
0.7
prevalence of F≥F2
FP 2 times worse than FN
FP 5 times worse than FN
FP 10 times worse than FN

A


cutoff values (kPa) for FN X-times worse than FP
0
10
20
30
40
0.02
0.04
0.06
0.08
0.10
prevalence of F=F4
FN 2 times worse than FP
FN 5 times worse than FP
FN 10 times worse than FP

B