2020-02

FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study

Newsome, PN

10.1016/s2468-1253(19)30383-8
The Lancet Gastroenterology & Hepatology
Elsevier BV

Title: FibroScan based FAST (FibroScan-AST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis (NASH) and significant activity and fibrosis: a prospective derivation and global validation study

Philip N Newsome*[1,2,3], Magali Sasso[4], Jonathan J Deeks[1,5], Angelo Paredes[6,7], Jérôme Boursier[8,9], Wah-Kheong Chan[10], Yusuf Yilmaz[11], Sébastien Czernichow[12], Ming-Hua Zheng[13,14], Vincent Wai-Sun Wong[15], Michael Allison[16], Emmanuel Tsochatzis[17], Quentin M Anstee[18,19], David A Sheridan[20], Peter Eddowes[21], Indra N Guha[21], Jeremy F Cobbold[22], Valérie Paradis[23], Pierre Bedossa[18,24], Véronique Miette[4], Céline Fournier[25], Laurent Sandrin[4], Stephen Harrison[26]

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

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

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

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

[5]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, UK

[6]Brooke Army Medical Center, Fort Sam Houston, TX, USA

[7]Uniformed Services University of the Health Sciences, Bethesda, MD, USA

[8]Department of Hepato-Gastroenterology, Angers University Hospital, Angers, France

[9]HIFIH Laboratory UPRES EA3859, SFR4208, Angers University, Angers, France

[10]Gastroenterology and Hepatology Unit, Gastrointestinal Endoscopy Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[11]Liver Research Unit, Institute of Gastroenterology, Marmara University, Istanbul, Turkey

[12]Department of nutrition, Hôpital européen Georges-Pompidou (APHP), University of Paris, Paris, France

[13]NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[14]Institute of Hepatology, Wenzhou Medical University, Wenzhou, China

[15]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China

[16]Department of Medicine, Cambridge Biomedical Research Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[17]UCL institute for liver and digestive health, royal free hospital and UCL, London, UK

[18]Institute of cellular medicine - faculty of medical sciences, Newcastle university, Newcastle upon Tyne, UK

[19]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals, NHS Trust, Newcastle upon Tyne, UK.

[20]Institute of translational and stratified medicine, University of Plymouth, Plymouth, UK

[21]NIHR Nottingham digestive diseases biomedical research unit, university of Nottingham and Nottingham university hospitals, Nottingham, UK

[22]NIHR Oxford Biomedical Research Centre and Oxford Liver Unit, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK

[23]Pathology department, Hôpital Beaujon, APHP, Clichy, France

[24]Liverpat, Paris, France

[25]Echosens, Medical Affairs department, Paris, France

[26]Radcliffe department of medicine, university of Oxford, UK

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

* 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: MS, VM, CF and LS are Echosens employees.

Author contributions: MS, CF, LS, VM, PN, JD were involved in study design and data analysis and data interpretation. Writing was done by PN, MS, CF and JD. Data collection was performed by PN, AP, JB, WKC, YY, SC, MHZ, VWSW, MA, ET, QM, DS, PE, IG, JC, VP, PB, CF. All co-authors reviewed and commented the manuscript and approved the final version.

Abbreviations:

AAR: AST to ALT ratio

AIC: Akaike information criterion

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

FIB-4: fibrosis-4 index

FLIP: fatty liver: inhibition of progression

GGT: gamma-glutamyl transferase

HDL: high-density lipoprotein

IQR: interquartile range

LSM: liver stiffness measurement

MCR: missed case rate

NAFLD: non-alcoholic fatty liver disease

NAS: NAFLD activity score

NASH: non-alcoholic steatohepatitis

NFS: NAFLD fibrosis score

NPV: negative predictive value

PPI: proportion of patients identified

PPV: positive predictive value

ROC: receiver operating characteristic

Se: sensitivity

SFR: screen failure rate

Sp: specificity

TRIPOD: transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

VCTE: vibration-controlled transient elastography

Abstract:

Background The current priority is to identify patients with non-alcoholic steatohepatitis (NASH) at risk of progression to cirrhosis who will be candidates for clinical trials and emerging new pharmacotherapies. To do so we aimed to develop a score to identify patients with NASH, elevated NAFLD activity score (NAS≥4) and advanced fibrosis (F≥2).

Methods A prospective, multicentre study of patients undergoing a liver biopsy for suspicion of NAFLD was conducted in the UK to develop this score. Liver stiffness measurement (LSM) by vibration-controlled transient elastography and controlled attenuation parameter (CAP) measured by FibroScan device were combined with aspartate aminotransferase (AST), alanine transaminase (ALT) or AST:ALT ratio. The best fitting model was identified and internally validated using boot-strapping. Score calibration and discrimination performance were determined in both the derivation dataset and seven international histologically confirmed cohorts.

Findings Between March 2014 and January 2017, 350 patients were prospectively enrolled in the derivation cohort. The most predictive model combined LSM, CAP and AST. Performance was satisfactory in the derivation dataset (C-statistic = 0·80, 95% confidence interval: 0·76-0·85 and was well calibrated). In external validation cohorts, calibration of the score was satisfactory and discrimination was good across the full range of validation cohorts (C-statistics ranging from 0·74 to 0·95, C-statistic = 0·85, 95% confidence interval: 0·83-0·87 in the pooled external validation patients' cohort; n=1026). Cut-offs for a sensitivity and a specificity ≥0.90 were 0.35 and 0.67 respectively in the derivation cohort and lead to a positive predictive value (PPV) of 0·83 and a negative predictive value (NPV) of 0·85. In the external validation cohorts, corresponding PPV ranged from 0·33 to 0·78 and NPV from 0·73 to 1.

Interpretation The FAST score provides an efficient way to non-invasively identify patients at-risk patients with progressive NASH that merit consideration for further treatment.

Funding Echosens.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is rising in prevalence in response to levels of obesity and type 2 diabetes mellitus such that it is now the commonest cause of chronic liver disease globally[1]. Whilst most patients with NAFLD do not progress to advanced fibrosis or cirrhosis, the high prevalence of NAFLD means that in numerical terms significant numbers of patients do develop chronic liver disease, such that NAFLD is now one of main indications for liver transplantation in Europe[2] and the US.[3] Thus, a key challenge is the identification of patients that are at greatest risk of clinical progression and who may benefit from treatment with the emerging range of new pharmacotherapies.[4,5]

Current practice utilises non-invasive markers that risk stratify liver fibrosis[6] or require percutaneous liver biopsy. The former approaches include algorithms[7], serum biomarkers[8] and imaging modalities[9] yet are limited in that they make no determination of the presence and degree of inflammatory liver injury. The presence of non-alcoholic steatohepatitis (NASH) and more profound liver cell injury, as determined by measures of steatosis, lobular inflammation and ballooning[10], is a critical driver of the development of liver fibrosis and will be significant in risk stratification and hence access to pharmacotherapies.

This prospective study set out to develop an algorithm to identify within patients with suspicion of NAFLD, those with NASH, significant liver fibrosis (≥F2) and an elevated NAFLD activity score (NAS≥4). This was undertaken in a derivation cohort before validation in multiple global cohorts (named external validation cohorts hereafter). The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines[11] were followed to report the development and internal as well as external validation of the

prediction model for diagnosis of NASH+NAS≥4+F≥2 (see Supplementary Table 1 for further details).[12]

Patients and methods

Derivation cohort

The derivation cohort was a cross-sectional prospective multicentre study, with a primary objective of assessing the diagnostic accuracy of controlled attenuation parameter (CAP) and secondary objectives of assessing the diagnostic accuracy of liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE), comparing CAP and LSM by VCTE with other non-invasive tests and also developing a score combining LSM by VCTE, CAP and biological markers to diagnose NASH.[13] Patients underwent a liver biopsy for suspicion of NAFLD, commonly due to the presence of abnormal liver enzymes in the presence of an ultrasound scan showing an echobright liver. Liver biopsy was used as the reference standard. Results for the identification of steatosis and fibrosis were recently reported by Eddowes *et al*[13]; this article reports results on the development of a score to identify patients with NASH and significant liver cell injury and fibrosis.

Study and participants

Consecutive patients were recruited between March 2014 and January 2017 in seven tertiary care 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).

Patients were eligible if ≥18 years of age, able to give written informed consent, and scheduled (independently from this study) to have a liver biopsy for investigation of suspected NAFLD within 2 weeks of FibroScan examination (before or after). Patients were also negative for hepatitis B surface antigen, anti-hepatitis C virus antibody, hepatitis C virus RNA, and hepatitis B virus DNA. Patients were excluded in case of: ascites, pregnancy, active implantable medical device (such as pacemaker or defibrillator); liver transplantation, cardiac failure and/or significant valvular disease; haemochromatosis; refusal to undergo liver biopsy or blood tests; alcohol consumption above recommended limits (>14 units/week for women and >21 units/week for men); confirmed diagnosis of active malignancy, or other terminal disease; or participation in another clinical trial within the preceding 30 days.

Patient Characteristics

Age, gender, body mass index (BMI), presence of diabetes, hypertension, and hypercholesterolemia were recorded for each patient. Moreover, a 12-hour fasting blood collection was performed locally and was then shipped to a central laboratory for assessment.

Histopathologic evaluation

Percutaneous liver biopsy was performed on all patients. Specimens were fixed in formalin, embedded in paraffin, and stained with haematoxylin and eosin and Picrosirius red. Slides were analysed independently by two experienced pathologists (PB, VP) blinded to each other's reading and to the patient's clinical and FibroScan data. In case of disagreement, they reviewed the slides together to reach consensus. Steatosis, ballooning, lobular inflammation grades,

fibrosis stage and NAFLD activity score (NAS) were scored using the NASH clinical research network (CRN) scoring system.[10] NASH was diagnosed using the "fatty liver: inhibition of progression" (FLIP) definition (at least grade one for each of steatosis, ballooning and lobular inflammation).[14]

LSM and CAP measurements

In each centre, LSM by VCTE and CAP were measured using FibroScan 502 Touch devices equipped with both M and XL probes (Echosens, Paris, France) by nurses or physicians trained and certified by the manufacturer and blinded to the patient's histological evaluation.[15,16] Probe selection was performed using the automatic probe selection tool embedded in the device software. All patients were asked to fast at least three hours before the examination, placed in the supine position with their right arm fully abducted. Measurements were performed by scanning the right liver lobe through an intercostal space. CAP and LSM by VCTE results were expressed in dB/m and kPa respectively. CAP is an average estimate of ultrasound attenuation at 3·5 MHz. LSM by VCTE is an average estimate of stiffness (Young's modulus) at a shear wave frequency of 50 Hz. Only examinations with at least ten valid individual measurements were deemed valid.

Sample size estimation

Sample size was determined for the primary objective of estimating the accuracy of the CAP parameter to achieve a 5% standard error in the estimates of the area under the receiver operating curve (AUROC) parameter in the subgroup using the XL probe. Assuming two-thirds of recruited patients would use the XL probe and allowing for 30% dropout, the target sample size was set at 450. This sample size was judged adequate to provide robust estimates for

predictive models based on 5 covariates, with over 30 events per variable at an expected prevalence of 50%.

Outcome and predictor variables

The main outcome was the diagnosis of NASH+NAS≥4+F≥2. NASH being defined using the FLIP definition, NAS score and fibrosis stage being scored using the NASH CRN scoring system. The models considered 5 different predictor variables: LSM by VCTE, CAP, and aspartate aminotransferase (AST), alanine transaminase (ALT) or AST to ALT ratio (AAR). It was anticipated that only one of AST, ALT and AAR would be included in the final model.

External validation cohorts

Data from seven clinical studies were gathered to perform the external validation of the score. Data came from different clinical settings and from different geographical origins (North America, Europe, Asia). Five cohorts came from tertiary care liver centres. One cohort came from a bariatric surgery centre and one from a study of screening for NAFLD among patients undergoing a routine colonoscopy. All external validation cohorts' data were collected in the framework of a clinical study for which local ethical committee granted approval. All patients from each study gave written informed consent to participate in the study. Each study was conducted in accordance with the declaration of Helsinki and in agreement with the ICH guidelines on GCP. Enrolment dates of each study are indicated in the Supplementary Table 2 together with external validation cohorts' descriptions. Of note, the Chinese Wenzhou NAFLD, French NAFLD and the Turkish NAFLD studies are still ongoing but the dataset used here are locked at the date of the last inclusion provided in the Supplementary Table 2. In each external validation study, patients were consecutive, FibroScan operators were blinded to patients' clinical data and all liver biopsies were read by expert pathologists who were blinded to patient

clinical data and FibroScan device results. For the two studies that had all patients measured with both M and XL probes (Chinese Hong-Kong and French NAFLD cohorts), the FibroScan examination corresponding to the XL probe was considered if the patient's BMI was greater or equal to 32 kg/m^2; the M probe was considered otherwise.

All external validation cohorts excluded patients with: *(i)* co-morbidities other than metabolic that may have induced liver lesions such as viral hepatitis, drug-induced liver injury, excessive alcohol consumption or human immunodeficiency virus, *(ii)* BMI >32 kg/m^2 if the M probe only was available for the study, *(iii)* less than ten valid measurement for FibroScan, *(iv)* with missing data for the developed score, *(v)* liver biopsy non interpretable or with missing data for the target of the score (NASH+NAS≥4+F≥2), *(vi)* a time interval greater than one year between FibroScan and liver biopsy. In each external validation cohort, patients had their individual lesions of steatosis, lobular inflammation, ballooning grades and fibrosis stage scored according to the NASH CRN scoring system.[10] From these individual items, a diagnosis of NASH was made according to the FLIP definition,[14] and the outcome (NASH+NAS≥4+F≥2) was computed.

Statistical analyses

Score development

The score was developed on the 350 patients in the derivation cohort. Eight patients (2% of the patients) had missing data either for CAP or AST/ALT, and since the proportion of observation with missing data was below 3%, single imputation was performed using stochastic regression imputation.[17,18] The selection of parameters was based on the combination of LSM by VCTE, related to liver fibrosis, and CAP, related to liver steatosis, with factors linked to NASH, inflammation and fibrosis (AST, ALT or their ratio). Parameters were combined into a

multivariable logistic regression model. Akaike's information criterion (AIC) was used to select between AST, ALT and AAR as the optimal parameter to combine to LSM by VCTE and CAP. The relative importance of each parameter was appraised using the Wald test. Nested models were compared using the likelihood ratio test. Optimal parameter transformations were selected using multivariable first degree fractional polynomials.

Internal validation

The model was internally validated using 2000 bootstrap samples.[18] Within each bootstrap iteration, we refitted the model and evaluated the performance in the bootstrap sample (apparent performance) and in the original data (test performance). Performance was assessed in terms of AUROC. The optimism was quantified as the mean differences of the performance estimates, and the shrinkage factor computed and applied to each regression coefficient in the original model to adjust the model for overfitting.

Diagnostic performance of the selected model

Model performance was assessed by calibration and discrimination in both the derivation and validation cohorts. Calibration (the agreement between observed outcomes and prediction) was assessed using calibration plot and a smoothing technique based on locally estimated scatterplot smoothing (Loess).[18] Discrimination was assessed using AUROC (similar to the Harrell's C-statistic). Cut-offs for a sensitivity (Se) ≥0·90 and a specificity (Sp) ≥0·90 were derived in the derivation cohort. When appraising performance at a given cut-offs, Se, Sp, positive predictive value (PPV) and negative predictive value (NPV) were computed together with 95% confidence intervals. Potential risk of bias in each external validation cohort were appraised in Supplementary Table 3. AUROC comparison was performed using Delong test. Statistical analyses were performed using the software R, version 3.4.1.[19]

Role of the funding source

The funder of the study had a role in study design, data collection, data analysis, data interpretation and writing of the report. The corresponding author and the funder had full access to all data in the study and had full responsibility for the decision to submit the publication.

Results

Of 450 potentially eligible participants (Figure 1), 350 were included in the FAST score construction, which was subsequently validated in seven different cohorts (total of 1026 patients). As reported in Table 1 the derivation cohort had broadly similar demographic, metabolic, serological and histological characteristics to patients in the pooled validation cohorts. Prevalence of NASH and NASH+NAS≥4+F≥2 was 69% and 50% respectively in the derivation cohort and 58% and 27% respectively in the pooled validation cohort.

Models combining LSM by VCTE, CAP and AST, ALT or AAR were compared (Supplementary Table 4). AST was determined to be the best parameter to combine with LSM and CAP. Further nested model comparison was performed (Supplementary Table 5) which demonstrated that a model combining LSM, CAP and AST had significantly better predictive properties than models with only one or two of these predictors. This resulted in the following equation for the FAST score:

$$FAST=\frac{e^{-1.65+1.07\times \log(LSM)+2.66*10^{-8}\times CAP^{3}-63.3\times AST^{-1}}}{1+e^{-1.65+1.07\times \log(LSM)+2.66*10^{-8}\times CAP^{3}-63.3\times AST^{-1}}}$$

FAST score was sensitive to each individual histological component (Supplementary Figure 1). As the derived FAST score is the predicted probability from the logistic regression model it is bounded between zero and one, and can be interpreted in a probabilistic manner. Performance in terms of AUROC in the bootstrap sample (apparent performance) and in the original data (test performance) were 0·803 (95% CI 0·758-0·849) and 0·804 (95% CI: 0·790-0·806), respectively, showing very little over-optimism (-6·0x10^{-4} (95% CI -4·3x10^{-2} to 4·7x10^{-2})). Predictive performance of FAST score in terms of discrimination, calibration and diagnostic accuracy (Figure 2) indicated an AUROC of 0·80 (95% CI 0·76-0·85) with satisfactory calibration of predicted probabilities.

Diagnostic performance of FAST score in terms of sensitivity, specificity, positive and negative predictive values is represented in figure 2C for different cut-off values. Moreover, figure 2D illustrates the performance of FAST score as it might be used in the context of identifying patients for therapies or drug trials for NASH. The screen failure rate (SFR) represents the proportion of screened subjects undergoing liver biopsy that would not meet the histological target (NASH+NAS≥4+F≥2) and would therefore not be randomized in trials or considered for treatment. If FAST score was used to identify such patients, the SFR would decrease from an initial 50% with increasing FAST cut-offs as illustrated in figure 2D, although with a rise in the number of patients identified as false negatives for NASH+NAS≥4+F≥2 (missed case rate, MCR). The histological characteristics of misclassified patients are detailed in Supplementary Table 11.

External validation of the score was evaluated alongside calibration plots for each external validation cohort (Figure 3). Calibration was satisfactory for the Chinese Hong-Kong, French and Turkish NAFLD cohorts, which have a prevalence of NASH+NAS≥4+F≥2 similar to the derivation cohort. However, for cohorts with a lower prevalence of the outcome, FAST score over-estimates the probability of having NASH+NAS≥4+F≥2. Corresponding AUROCs are provided (Table 2), with discrimination being good to excellent in all external validation cohort with the exception of the Turkish NAFLD cohort which had modest performance. Best discrimination was observed in the French bariatric surgery cohort with an AUROC of 0·95 (95% CI 0·91-0·99).

Cut-offs for sensitivity and specificity of ≥0·90 were 0·35 and 0·67, respectively in the derivation cohort (full diagnostic performance in Supplementary Table 6), with characteristics for validation cohorts detailed in Table 2. Using the dual cut-off approach, PPV in the

derivation cohort was 0·83 (95% CI 0·75 to 0·87), NPV was 0·85 (0·77 to 0·88) and 39% of the patients are in the grey zone between the two cut-offs. When applying these cut-offs to the external validation cohorts, PPV are in the same order of magnitude with a similar sensitivity in the Chinese Hong-Kong, French and Turkish NAFLD cohorts. The USA screening cohort has a PPV in the same order of magnitude but a lower sensitivity. The Chinese Wenzhou NAFLD with the lowest prevalence of NASH+NAS≥4+F≥2 has a lower PPV but similar sensitivity. NPV is high in all external validation cohorts.

FAST was compared with FIB-4 and NFS for the identification of patients with NASH+NAS≥4+F≥2 in the subgroup of patients from the derivation and external validation cohorts that had all parameters needed to compute those three scores. Corresponding AUROCs and diagnostic performance using the dual-cut-off approach (Supplementary Table 7) were inferior to the FAST score. Indeed, discrimination was significantly higher for the FAST score in the derivation and in the pooled external validation cohort. In addition, using the dual cut-off approach FIB-4 in comparison to FAST yielded a similar number of patients in the grey zone (32 vs 31% respectively in the pooled external validation cohort), slightly higher PPV (0·72 vs 0·69, respectively) but lower NPV (0·83 vs 0·94, respectively) and failed to identify most of the patients with NASH+ NAS≥4+F≥2 (sensitivity of 0·07 vs 0·49, respectively). NFS had a higher grey zone than FAST (49% vs 31%, respectively) with lower PPV (0·50 vs 0·69, respectively) and NPV (0·85 vs 0·94, respectively). Moreover, the addition of metabolic parameters to the score were appraised and did not provide significant improvement in terms of discrimination (Supplementary Table 8).

Discussion

In this prospective cohort study, we present a new simple non-proprietary score which allows for the identification of patients with progressive NASH, and which has been validated in multiple large global cohorts.

There has been considerable debate as to which patients with NASH should be the focus of monitoring and treatment, with recent data confirming that the degree of fibrosis is a major driver of clinical outcomes[20-22]. Our choice of NASH with a NAS≥4 and F≥2 is based on this literature and also many therapeutic studies that demonstrate the presence of elevated necro-inflammatory activity is linked to progressive injury and pharmacological response[23,24].

The FAST score, in keeping with recommended practice, was configured to have two thresholds, a rule-out and a rule-in cut-off. This allowed for classification of >70% of patients in the validation cohorts. Moreover, FAST score has good performance characteristics with a negative likelihood ratio of 0·2 (rule-out cut-off) and a positive likelihood ratio of 5 (rule-in cut-off), ratios which are maintained in the validation groups. Thus, this test has a significant impact on influencing clinical decision making and will be an important adjunct in identifying patients for clinical trials or commencement of pharmacotherapies.

In generating this score, we sought to determine the performance of standard liver blood tests or other widely-used algorithms in identifying patients with progressive NASH. The performance of these parameters in isolation was significantly inferior to the FAST score (Supplementary Table 4), and indeed other than the addition of AST values, there was no evidence that addition of other elements (metabolic syndrome parameters) improved the performance of the FAST score (see Supplementary Table 8).

We, and others, have recently shown that CAP and LSM by VCTE measurements are widely applicable in patients with NASH, with low failure rate (3%) and good performance in determining the degree of liver steatosis and fibrosis, respectively.[13,25] Importantly, we also demonstrated that LSM values only correlated with fibrosis, and were not influenced by other histological parameters or type of probe used.

This study has a number of strengths. Firstly, it is prospective in nature and has undergone extensive validation across multiple large global cohorts. Secondly, FAST score performance is good across the full range of validation cohorts. AUROCs ranged from 0·74 to 0·95, with PPV up to 0·85 and a NPV ranging from 0·73-1 using the dual-cut-offs approach with cut-offs derived in the derivation cohort. Thirdly, the wide availability of FibroScan devices based on VCTE technology, the need for just a serum AST value, its non-invasive nature, its low cost per scan and also modest requirement to attain technical proficiency required to perform the scans, means it can be rolled out easily across most clinical practices. This is further aided by the free availability of the equation which is also available through an app.

The weaknesses of this study are the requirement for a FibroScan device which may impact on some elements of its global uptake, although it is available in more than 90 countries at present. Its use in primary care will require investment in devices and personnel in those areas, although there are many successful examples of such models occurring[6]. Another potential criticism is that our score focusses on patients with ≥F2 fibrosis where there may be a view that the priority is identifying patients with more advanced fibrosis (≥F3). At this stage however clinical trials are aiming for patients with F2 and F3 fibrosis hence this constitutes a reasonable target group. We however, derived cut-offs for the identification of patients with advanced fibrosis (F≥3) in

our cohort (see Supplementary Table 9 and Supplementary Table 10). The performance characteristics are good with moderate likelihood ratios to rule-out and rule-in such patients. Finally, FAST score performs least well in terms of calibration in low prevalence populations, and thus caution should be exercised when interpreting the score in such populations such as primary care. Discrimination performance of the score is however good in these populations. In a future study, recalibration of scores could be considered to correct miscalibration whilst still keeping the same level of discrimination.

In summary, we believe FAST score will allow for the ready identification of an at-risk group of patients with progressive NASH that merit consideration for further treatment.

Figures Legend

Figure 1: Derivation cohort flow chart

Figure 2: Diagnostic performance in the derivation cohort of the FAST score for the diagnostic of NASH+NAS≥4+F≥2

(A) Receiver operating characteristic (ROC) curve plot and area under the ROC curve (95% confidence interval). (B) Calibration plot and calibration intercept and slope (95% confidence interval, grey ribbon). The calibration plot characterises the agreement between observed proportion and predicted probabilities. The intercept compares the mean of all predicted risks with the mean observed risk and indicates the extent that predictions are systematically too low or too high.[18] The slope account for differences in performance in high- or low-risk groups. Calibration of the data is estimated using a smoothed regression line (dotted line) estimated using locally estimated scatterplot smoothing (Loess) which allows to inspect the calibration across the range of predicted value and to determine if there are segments of the range in which the model is poorly calibrated.[18] Triangles represent deciles of subjects (n=50) grouped by similar predicted risk. Here calibration of the score is satisfactory since intercept is not significantly different from 0, slope not significantly different from 1, the flexible calibration curve is close to the ideal calibration (solid line) and its confidence interval zone include the ideal curve. (C) Sensitivity (Se), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) versus FAST score values. Plot of Se, Sp, PPV and NPV versus all possible FAST score values. (D) Screen failure rate (SFR), missed cases rate (MCR) and proportion of patients identified (PPI) versus FAST scores values. Plot of the SFR (equal to 1-PPV) and MCR (equal to 1-Se) versus all possible FAST score values. At a given FAST score cut-offs, it is possible to graphically assess the SFR and MCR together with the proportion on

patients above the FAST score who would be sent to liver biopsy in the context of patients screening in drug trials for NASH.

Figure 3: Calibration plots in external validation cohorts. (A) French bariatric cohort. (B) USA screening cohort. (C): China Hong-Kong NAFLD cohort. (D) China Wenzhou NAFLD cohort. (E): French NAFLD cohort. (F): Malaysian NAFLD cohort. (G): Turkish NAFLD cohort.

In each figure, the solid line represents the ideal calibration. The dotted line represents the calibrations estimated on the data using locally estimated scatterplot smoothing (Loess). The grey ribbon represents its 95% confidence limits. Triangles represent deciles of subjects grouped by similar predicted risk. The distribution of subjects is indicated with spikes at the bottom of the graph (patients with NASH+NAS≥4+F≥2 the x-axis, patients without NASH+NAS≥4+F≥2 below the x-axis). The French (E) and Turkish (G) NAFLD external validation cohort are well calibrated; their calibration curve is nearly linear, their intercept is close to zero (their confidence intervals include zero) and their slope is very close to one (their confidence interval include one). The Chinese Hong-Kong NAFLD cohort (C), has a zone where the risk of being NASH+NAS≥4+F≥2 is overestimated using the FAST score (grey ribbon below the ideal calibration curve) and a zone where the calibration seem adequate (grey ribbon zone include the ideal calibration curve). However, this cohort size is quite small (n=83). The French bariatric surgery (A), USA screening (B) Chinese Wenzhou NALFD (D) and the Malaysian NAFLD (F) cohort have a range of prevalence of NASH+NAS≥4+F≥2 (9% to 20%) which is lower than the derivation cohort. In those four cohorts, the FAST score over-estimates the probability of being NASH+NAS≥4+F≥2. The discrepancy is mainly driven by the intercept (all confidence intervals do not include zero). All slopes are within acceptable range

(the confidence interval include one), except for the French bariatric cohort which seem to be at the limit.

Table 1. Derivation and external validation cohorts' patient characteristics

		Derivation cohort	French bariatric surgery cohort	USA screening cohort	China Hong-Kong NAFLD cohort	China Wenzhou NAFLD cohort	French NAFLD cohort	Malaysian NAFLD cohort	Turkish NAFLD cohort	Pooled external patients cohort
Demographics	**n**	350	110	242	83	104	182	176	129	1026
	Age (years)	54 (45-63)	41 (33-50)	55 (50-60)	55 (46-63)	41 (30-50)	58 (49-66)	52 (46-60)	49 (38-57)	52 (44-60)
	Female	149 (43%)	88 (80%)	97 (40%)	42 (51%)	28 (27%)	65 (36%)	84 (48%)	59 (46%)	463 (45%)
	BMI ($kg \cdot m^{-2}$)	34·2 (29·6-38·6)	43·0 (38·8-47·2)	32·6 (30·0-36·1)	28·9 (26·0-31·9)	25·5 (23·4-27·6)	31·6 (28·6-37·2)	28·1 (25·9-30·0)	33·0 (30·0-36·0)	31·0 (27·7-36·1)
Metabolic	**Diabetes**	176 (50%)	25 (23%)	55 (23%)	54 (65%)	26 (25%)	86 (47%)	90 (51%)	79 (61%)	415 (40%)
	Hypertension	189 (54%)	29 (26%)	113 (47%)	57 (69%)	17 (16%)	–	104 (59%)	69 (53%)	389 (46%)
Blood	**AST (IU/L)**	36 (27-52)	26 (21-39)	22 (18-27)	41 (28-59)	34 (27-52)	36 (28-50)	38 (29-62)	37 (28-59)	32 (23-48)
	ALT (IU/L)	50 (34-72)	37 (31-54)	25 (19-38)	65 (32-97)	48 (32-88)	48 (32-77)	63 (43-104)	54 (34-106)	44 (28-74)
	GGT (IU/L)	57 (34-113)	36 (23-52)	27 (20-40)	56 (37-90)	50 (28-77)	68 (37-131)	74 (41-122)	54 (34-86)	46 (28-85)
	Albumin (g/L)	4·5 (4·3-4·7)	3·9 (3·7-4·1)	4·3 (4·1-4·5)	4·3 (3·9-4·6)	4·8 (4·5-5·0)	4·3 (4·0-4·5)	4·3 (4·1-4·6)	4·6 (4·3-4·8)	4·4 (4·1-4·6)
	Platelets count (x 10^9/L)	239 (199-281)	247 (216-283)	236 (201-285)	225 (178-263)	237 (206-266)	223 (170-269)	272 (228-316)	222 (190-267)	238 (199-284)
	Fasting glucose (mg/dL)	108 (91-142)	88 (79-103)	103 (93-120)	117 (99-141)	92 (86-108)	111 (99-141)	105 (94-128)	109 (96-126)	104 (92-125)
	Triglyceride (mg/dL)	163 (119-213)	130 (100-170)	135 (94-190)	150 (115-221)	178 (116-272)	–	133 (106-168)	169 (116-227)	146 (106-197)
	Total cholesterol (mg/dL)	181 (147-212)	192 (167-221)	190 (158-217)	181 (154-207)	184 (150-217)	–	181 (162-216)	214 (182-242)	190 (161-220)
	HDL cholesterol (mg/dL)	42 (34-50)	50 (40-60)	48 (39-58)	46 (39-50)	36 (33-42)	–	45 (39-52)	44 (39-53)	45 (38-54)
	FIB-4	1·13 (0·78-1·68)	0·69 (0·48-1·11)	0·99 (0·81-1·31)	1·27 (0·96-1·72)	0·91 (0·62-1·20)	1·38 (0·90-1·96)	0·96 (0·65-1·40)	1·17 (0·79-1·59)	1·04 (0·72-1·46)

Fibrosis scores	**NFS**	-1·00 (-2·12-0·08)	-0·80 (-2·17--0·08)	-0·97 (-1·90-0·04)	-0·95 (-2·04--0·16)	-2·77 (-3·61--1·87)	-0·60 (-1·38-0·57)	-2·16 (-3·04--1·17)	-1·12 (-1·80--0·22)	-1·28 (-2·32--0·24)
FibroScan	**FibroScan Probe**	M: 111 (32%) XL: 239 (68%)	M: 10 (9%) XL: 100 (91%)	M: 141 (58%) XL: 101 (42%)	M: 63 (76%) XL: 20 (24%)	M: 104 (100%) XL: 0 (0%)	M: 99 (54%) XL: 83 (46%)	M: 176 (100%) XL: 0 (0%)	M: 68 (53%) XL: 61 (47%)	M: 661 (64%) XL: 365 (36%)
	LSM by VCTE (kPa)	8·9 (6·2-13·9)	5·9 (4·7-8·8)	6·0 (4·7-8·2)	8·8 (6·6-12·2)	5·8 (5·1-6·7)	7·9 (5·9-11·5)	7 (6-10)	11·1 (8·6-14·6)	7·2 (5·3-10·3)
	CAP (dB/m)	342 (307-373)	310 (275-374)	317 (276-360)	319 (290-354)	316 (284-332)	326 (297-369)	323 (289-343)	329 (304-356)	321 (288-355)
Histology	**Length of liver biopsy specimen (mm)**	23 (10)	12 (5)	14 (5)	23 (8)	–	29 (11)	15 (4)	30 (14)	17 (12)
	Length of liver biopsy specimen ≥ 15mm	315 (90%)	50 (45%)	109 (45%)	74 (89%)	–	169 (93%)	102 (58%)	127 (97%)	629 (68%)
	Fibrosis Stage	0: 60 (17%) 1: 80 (23%) 2: 81 (23%) 3: 101 (29%) 4: 28 (8%)	0: 65 (59%) 1: 26 (24%) 2: 9 (8%) 3: 9 (8%) 4: 1 (1%)	0: 131 (54%) 1: 74 (31%) 2: 26 (11%) 3: 11 (5%) 4: 0 (0%)	0: 9 (11%) 1: 23 (28%) 2: 15 (18%) 3: 17 (20%) 4: 19 (23%)	0: 45 (43%) 1: 46 (44%) 2: 8 (8%) 3: 5 (5%) 4: 0 (0%)	0: 28 (15%) 1: 46 (25%) 2: 46 (25%) 3: 53 (29%) 4: 9 (5%)	0: 62 (35%) 1: 73 (41%) 2: 12 (7%) 3: 24 (14%) 4: 5 (3%)	0: 16 (12%) 1: 37 (29%) 2: 30 (23%) 3: 33 (26%) 4: 13 (10%)	0: 356 (35%) 1: 325 (32%) 2: 146 (14%) 3: 152 (15%) 4: 47 (5%)
	Ballooning Grade	0: 78 (22%) 1: 142 (41%) 2: 130 (37%)	0: 64 (58%) 1: 35 (32%) 2: 11 (10%)	0: 127 (52%) 1: 92 (38%) 2: 23 (10%)	0: 35 (42%) 1: 39 (47%) 2: 9 (11%)	0: 28 (27%) 1: 63 (61%) 2: 13 (12%)	0: 44 (24%) 1: 73 (40%) 2: 65 (36%)	0: 58 (33%) 1: 78 (44%) 2: 40 (23%)	0: 5 (4%) 1: 64 (50%) 2: 60 (47%)	0: 361 (35%) 1: 444 (43%) 2: 221 (22%)
	Lobular inflammation Grade	0: 72 (21%) 1: 224 (64%) 2: 50 (14%) 3: 4 (1%)	0: 71 (65%) 1: 33 (30%) 2: 5 (5%) 3: 1 (1%)	0: 110 (45%) 1: 111 (46%) 2: 20 (8%) 3: 1 (0%)	0: 0 (0%) 1: 35 (42%) 2: 45 (54%) 3: 3 (4%)	0: 18 (17%) 1: 66 (63%) 2: 17 (16%) 3: 3 (3%)	0: 38 (21%) 1: 123 (68%) 2: 21 (12%) 3: 0 (0%)	0: 3 (2%) 1: 100 (57%) 2: 67 (38%) 3: 6 (3%)	0: 2 (2%) 1: 51 (40%) 2: 49 (38%) 3: 27 (21%)	0: 242 (24%) 1: 519 (51%) 2: 224 (22%) 3: 41 (4%)
	Steatosis grade	0: 17 (5%) 1: 87 (25%) 2: 108 (31%) 3: 138 (39%)	0: 37 (34%) 1: 27 (25%) 2: 21 (19%) 3: 25 (23%)	0: 56 (23%) 1: 90 (37%) 2: 56 (23%) 3: 40 (17%)	0: 0 (0%) 1: 34 (41%) 2: 30 (36%) 3: 19 (23%)	0: 0 (0%) 1: 44 (42%) 2: 43 (41%) 3: 17 (16%)	0: 12 (7%) 1: 81 (45%) 2: 48 (26%) 3: 41 (23%)	0: 4 (2%) 1: 48 (27%) 2: 92 (52%) 3: 32 (18%)	0: 0 (0%) 1: 18 (14%) 2: 46 (36%) 3: 65 (50%)	0: 109 (11%) 1: 342 (33%) 2: 336 (33%) 3: 239 (23%)
	NAS score ≥4	239 (68%)	36 (33%)	81 (33%)	50 (60%)	47 (45%)	110 (60%)	115 (65%)	120 (93%)	559 (54%)
	NASH	242 (69%)	31 (28%)	92 (38%)	48 (58%)	63 (61%)	122 (67%)	116 (66%)	123 (95%)	595 (58%)
	NASH+NAS≥4+F≥2	174 (50%)	16 (15%)	28 (12%)	36 (43%)	9 (9%)	78 (43%)	36 (20%)	74 (57%)	277 (27%)
	Time between FibroScan and liver biopsy (days)	0 (0) range: -14-12	78 (49-162) range: -328-332	56 (40-84) range: -33-309	1 (-2-1) range: -95-161	0 (0-0) range: -84-9	0 (0-0) range: 0-0	0 (0-0) range: 0-0	35 (16-113) range: -271-360	1 (0-55) range: -328-360

Time between procedures	**Time between FibroScan and blood analyses (days)**	0 (0) range: -1-9	–	9 (0-27) range: -151-217	0 (0-0) range: 0-0	0 (0-0) range: -84-9	0 (0-0) range: 0-0	0 (0-0) range: 0-0	18 (1-113) range: -315-373	0 (0-4) range: -315-373
	Time between liver biopsy and blood analyses (days)	0 (0) range: -15-12	–	-46 (-70--22) range: -309-93	-1 (-1-2) range: -161-95	0 (0-0) range: 0-0	0 (0-0) range: 0-0	0 (0-0) range: 0-0	-12 (-31-36) range: -435-293	0 (-24-0) range: -435-293

Distribution is expressed as median (interquartile range: quartile 1- quartile 3) or figure (percentage). ALT: alanine transaminase, AST: aspartate aminotransferase, BMI: body mass index, CAP: controlled attenuation parameter, FIB-4: fibrosis-4 index, GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, LSM: liver stiffness measurement, NAFLD: non-alcoholic fatty liver disease, NASH: non-alcoholic steatohepatitis, NAS: NAFLD activity score, NFS: NAFLD fibrosis score. The non-alcoholic fatty liver disease activity score and Kleiner scoring system are described in the Supplementary materials (p 2).

Table 2: Diagnostic performance of the FAST score for the diagnostic of NASH+NAS≥4+F≥2 in the derivation and external validation cohorts

	AUROC (95% CI)	n	Prevalence of NASH+NAS≥4+F≥2	Performance using dual cut-off (cut-offs from derivation cohort)		
				rule-out zone	grey zone	rule-in zone
Derivation cohort	0·80 (0·76-0·85)	350	174 (50%)	**FAST<0·35** n=110 (31%) Se=0·90 Sp=0·53 NPV=0·85	**FAST: 0·35-0·67** n=136 (39%)	**FAST≥0·67** n=93 (27%) Sp=0·90 Se=0·48 PPV=0·83
French bariatric surgery cohort	0·95 (0·91-0·99)	110	16 (15%)	**FAST<0·35** n=69 (63%) Se=1·00 Sp=0·73 NPV=1·00	**FAST: 0·35-0·67** n=22 (20%)	**FAST≥0·67** n=19 (17%) Sp=0·93 Se=0·75 PPV=0·63
USA screening cohort	0·86 (0·80-0·93)	242	28 (12%)	**FAST<0·35** n=193 (80%) Se=0·64 Sp=0·86 NPV=0·95	**FAST: 0·35-0·67** n=39 (16%)	**FAST≥0·67** n=8 (3%) Sp=0·99 Se=0·25 PPV=0·78
China Hong-Kong NAFLD cohort	0·85 (0·76-0·93)	83	36 (43%)	**FAST<0·35** n=28 (34%) Se=0·94 Sp=0·55 NPV=0·93	**FAST: 0·35-0·67** n=29 (35%)	**FAST≥0·67** n=25 (30%) Sp=0·89 Se=0·58 PPV=0·81
China Wenzhou NAFLD cohort	0·84 (0·73-0·95)	104	9 (9%)	**FAST<0·35** n=54 (52%) Se=0·89 Sp=0·56 NPV=0·98	**FAST: 0·35-0·67** n=37 (36%)	**FAST≥0·67** n=11 (11%) Sp=0·92 Se=0·44 PPV=0·33
French NAFLD cohort	0·80 (0·73-0·86)	182	78 (43%)	**FAST<0·35** n=67 (37%) Se=0·88 Sp=0·56 NPV=0·87	**FAST: 0·35-0·67** n=69 (38%)	**FAST≥0·67** n=43 (24%) Sp=0·89 Se=0·45 PPV=0·76
Malaysian NAFLD cohort	0·85 (0·78-0·91)	176	36 (20%)	**FAST<0·35**	**FAST: 0·35-0·67**	**FAST≥0·67**

				n=77 (44%) Se=0·94 Sp=0·54 NPV=0·97	n=59 (34%)	n=38 (22%) Sp=0·87 Se=0·58 PPV=0·54
Turkish NAFLD cohort	0·74 (0·65-0·82)	129	74 (57%)	**FAST<0·35** n=26 (20%) Se=0·91 Sp=0·35 NPV=0·73	**FAST: 0·35-0·67** n=57 (44%)	**FAST≥0·67** n=44 (34%) Sp=0·82 Se=0·49 PPV=0·78
Pooled external patients cohort	0·85 (0·83-0·87)	1026	277 (27%)	**FAST<0·35** n=514 (50%) Se=0·89 Sp=0·64 NPV=0·94	**FAST: 0·35-0·67** n=312 (30%)	**FAST≥0·67** n=188 (18%) Sp=0·92 Se=0·49 PPV=0·69

Performance associated with dual cut-off approach is evaluated using the FAST score when the cut-offs are calculated in the derivation cohort and applied in all external validation cohorts. The lower cut-off constitutes a rule-out cut-off and is based on a sensitivity≥0·90 in the derivation cohort. The higher cut-off constitutes a rule-in cut-off and is based on a specificity≥0·90 in the derivation cohort. Individuals with a FAST score in between the rule-out and rule-in cut-offs are in the grey zone. In the rule-out group, the sensitivity is provided together with the specificity and negative predictive value to appraise the rule-out performance of the score. In the rule-in group, the specificity is provided together with the sensitivity and positive predictive value to appraise the rule-in performance of the score. AUROC: area under the receiver operating curve, NALFD: non-alcoholic fatty liver disease, NASH: non-alcoholic fatty liver disease: NAS: NAFLD activity score, NPV: negative predictive value, PPV: positive predictive value, Se: sensitivity, Sp: specificity, USA: United States of America.

References

1. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2017.
2. Haldar D, Kern B, Hodson J, et al. Outcomes of liver transplantation for non-alcoholic steatohepatitis: A European Liver Transplant Registry study. *J Hepatol* 2019.
3. Charlton MR, Burns JM, Pedersen RA, Watt KD, Heimbach JK, Dierkhising RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology* 2011; **141**(4): 1249-53.
4. Tsochatzis EA, Newsome PN. Non-alcoholic fatty liver disease and the interface between primary and secondary care. *Lancet Gastroenterol Hepatol* 2018; **3**(7): 509-17.
5. Sanyal AJ. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nature reviews Gastroenterology & hepatology* 2019; **16**(6): 377-86.
6. Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. *J Hepatol* 2019.
7. 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**(4): 846-54.
8. Guha IN, Parkes J, Roderick P, et al. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology* 2008; **47**(2): 455-60.
9. Sanyal AJ, Friedman SL, McCullough AJ, et al. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration Joint Workshop. *Hepatology* 2015; **61**(4): 1392-405.
10. 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**(6): 1313-21.
11. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015; **162**(1): W1-73.
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**: g7594.
13. Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2019; **156**(6): 1717-30.
14. Bedossa P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* 2014; **60**(2): 565-75.
15. Sandrin L, Fourquet B, Hasquenoph JM, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound in medicine & biology* 2003; **29**(12): 1705-13.
16. 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 in medicine & biology* 2010; **36**(11): 1825-35.
17. Frank E. Harrell J. Regression Modeling Strategies: Springer-Verlag; 2006.
18. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: Springer New York; 2008.

19. R Core Team. R: A Language and Environment for Statistical Computing. In: https://www.R-project.org/, editor. Vienna, Austria: R Foundation for Statistical Computing; 2018.
20. 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**(2): 389-97 e10.
21. 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**(5): 1557-65.
22. Ekstedt M, Hagstrom H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015; **61**(5): 1547-54.
23. Ratziu V, Harrison S, Francque S, et al. Elafibranor, an Agonist of the Peroxisome Proliferator-activated Receptor-alpha and -delta, Induces Resolution of Nonalcoholic Steatohepatitis Without Fibrosis Worsening. *Gastroenterology* 2016.
24. Friedman SL, Ratziu V, Harrison SA, et al. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. *Hepatology* 2018; **67**(5): 1754-67.
25. 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**(1): 134-44.