OPEN ACCESS

Original research

Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis

Ferenc Emil Mózes ,[1] Jenny A Lee,[2] Emmanuel Anandraj Selvaraj,[1,3,4] Arjun Narayan Ajmer Jayaswal,[1] Michael Trauner ,[5] Jerome Boursier ,[6,7] Céline Fournier,[8] Katharina Staufer,[5,9,10] Rudolf E Stauber,[11] Elisabetta Bugianesi ,[12] Ramy Younes ,[13] Silvia Gaia,[12] Monica Lupșor-Platon,[14] Salvatore Petta ,[15] Toshihide Shima,[16] Takeshi Okanoue,[16] Sanjiv Mahadeva ,[17] Wah-Kheong Chan,[17] Peter J Eddowes,[18,19] Gideon M Hirschfield,[20] Philip Noel Newsome ,[18,21,22] Vincent Wai-Sun Wong ,[23] Victor de Ledinghen,[24,25] Jiangao Fan,[26] Feng Shen,[26] Jeremy F Cobbold,[3,4] Yoshio Sumida,[27] Akira Okajima,[28] Jörn M Schattenberg,[29] Christian Labenz ,[29] Won Kim,[30] Myoung Seok Lee,[31] Johannes Wiegand ,[32] Thomas Karlas ,[32] Yusuf Yılmaz ,[33,34] Guruprasad Padur Aithal ,[19,35] Naaventhan Palaniyappan,[19,35] Christophe Cassinotto ,[36] Sandeep Aggarwal,[37] Harshit Garg,[37] Geraldine J Ooi ,[38] Atsushi Nakajima ,[39] Masato Yoneda,[39] Marianne Ziol,[40] Nathalie Barget,[41] Andreas Geier ,[42] Theresa Tuthill,[43] M. Julia Brosnan,[43] Quentin Mark Anstee,[44] Stefan Neubauer,[1] Stephen A. Harrison,[1] Patrick M Bossuyt,[2] Michael Pavlides ,[1,3,4] the LITMUS Investigators

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-324243).

For numbered affiliations see end of article.

Correspondence to
Dr Michael Pavlides, Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, Oxfordshire, UK; michael.pavlides@cardiov.ox.ac.uk

Received 28 January 2021
Revised 23 April 2021
Accepted 29 April 2021

To cite: Mózes FE, Lee JA, Selvaraj EA, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2021-324243

ABSTRACT

Objective Liver biopsy is still needed for fibrosis staging in many patients with non-alcoholic fatty liver disease. The aims of this study were to evaluate the individual diagnostic performance of liver stiffness measurement by vibration controlled transient elastography (LSM-VCTE), Fibrosis-4 Index (FIB-4) and NAFLD (non-alcoholic fatty liver disease) Fibrosis Score (NFS) and to derive diagnostic strategies that could reduce the need for liver biopsies.

Design Individual patient data meta-analysis of studies evaluating LSM-VCTE against liver histology was conducted. FIB-4 and NFS were computed where possible. Sensitivity, specificity and area under the receiver operating curve (AUROC) were calculated. Biomarkers were assessed individually and in sequential combinations.

Results Data were included from 37 primary studies (n=5735; 45% women; median age: 54 years; median body mass index: 30 kg/m^2; 33% had type 2 diabetes; 30% had advanced fibrosis). AUROCs of individual LSM-VCTE, FIB-4 and NFS for advanced fibrosis were 0.85, 0.76 and 0.73. Sequential combination of FIB-4 cut-offs (<1.3; ≥2.67) followed by LSM-VCTE cut-offs (<8.0; ≥10.0 kPa) to rule-in or rule-out advanced fibrosis had sensitivity and specificity (95% CI) of 66% (63–68) and 86% (84–87) with 33% needing a biopsy to establish a final diagnosis. FIB-4 cut-offs (<1.3; ≥3.48) followed by LSM cut-offs (<8.0; ≥20.0 kPa) to rule out advanced fibrosis or rule in cirrhosis had a sensitivity of 38% (37–39) and specificity of 90% (89–91) with 19% needing biopsy.

Conclusion Sequential combinations of markers with a lower cut-off to rule-out advanced fibrosis and a higher cut-off to rule-in cirrhosis can reduce the need for liver biopsies.

Significance of this study

What is already known on this subject?

- Patients with non-alcoholic fatty liver disease (NAFLD) and advanced fibrosis (F3–4) are at risk of disease progression and adverse clinical outcomes.
- Non-invasive tests with predefined cut-offs are used as screening biomarkers to identify those at low risk of advanced fibrosis who can be safely managed in primary care.
- Liver biopsy is still needed in secondary care to further identify those with cirrhosis who would benefit from surveillance for hepatocellular cancer and screening for oesophageal varices.

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is the hepatic manifestation of the metabolic syndrome with high prevalence worldwide.[1] Most patients remain asymptomatic for long periods of time (years/decades) with slowly progressive disease, but

Significance of this study

What are the new findings?

- Existing non-invasive tests cut-offs are validated for their use as screening biomarkers to rule out advanced fibrosis in a study group of 5735 patients.
- The sequential combination of Fibrosis-4 Index (FIB-4) (<1.3; ≥2.67) and liver stiffness measurement by vibration controlled transient elastography (LSM-VCTE) (<8.0 kPa; ≥10.0 kPa) which is increasingly used in routine practice has a false negative rate of 9% for advanced fibrosis.
- The diagnostic performance of LSM-VCTE for advanced fibrosis is influenced by biopsy quality, body mass index and presence of type 2 diabetes.
- An algorithm combining FIB-4 and LSM-VCTE sequentially with lower cut-offs to rule out advanced fibrosis (FIB-4 <1.3; LSM-VCTE <8.0 kPa) and with upper cut-offs to rule-in and positively diagnose cirrhosis without the need for liver biopsy with specificity of 95% (FIB-4 ≥3.48; LSM-VCTE ≥20.0 kPa) or 98% (FIB-4 ≥4.63; LSM-VCTE ≥28.0 kPa) can reduce the need for liver biopsies from 33% to 19% or 24%, respectively.

How might it impact on clinical practice in the foreseeable future?

- The non-invasive test cut-offs for the diagnosis of cirrhosis can be incorporated into clinical practice as they have been validated in a large group of patients.
- Application of these cut-offs can lead to a decrease in the need for liver biopsies in secondary care.

a minority[2] progress to cirrhosis, liver failure and hepatocellular carcinoma (HCC).

NAFLD comprises several histological features ranging from simple steatosis to steatosis with lobular inflammation and ballooned hepatocytes (steatohepatitis), both of which can be accompanied by varying degrees of fibrosis. The currently accepted reference standard for diagnosing NAFLD is liver biopsy as its diagnostic features are based on histology.[3] Liver biopsy, however, is invasive and carries a risk of complications,[4] is limited by sampling variability[5] and high observer dependent variability in pathological reporting.[6 7]

NAFLD is often diagnosed after incidental findings of elevated liver transaminases on blood tests, or liver steatosis or cirrhosis on imaging. One challenge clinicians face is to identify which of these patients are at high risk of progression or clinical outcomes, as they would benefit from specialist follow-up. There is now substantial evidence showing that those with at least advanced fibrosis (F3–4) are at higher risk of liver-related events in later life.[8–10]

A large body of evidence also exists on how non-invasive tests (NITs) could be used to risk-stratify patients for the presence of advanced fibrosis. These approaches usually involve sequential application of two NITs, with the first tier of a simple, inexpensive, serum-based test performed in the community (eg, Fibrosis-4 Index (FIB-4) or NAFLD Fibrosis Score (NFS)), followed by a second tier of liver stiffness measurement (LSM) (eg, vibration controlled transient elastography: VCTE), or a proprietary serum-based test (eg, enhanced liver fibrosis test; ELF). A lower and an upper threshold are usually used in each tier of testing to rule out (those with a NIT result less than the lower threshold) or rule in (those with a NIT result more than the upper threshold) patients at high risk of advanced fibrosis. Patients with indeterminate results in both tiers of testing would need a liver biopsy for risk stratification. The main value of these approaches lies in their high negative predictive value to rule out patients with low risk of advanced fibrosis who can be safely managed in primary care.

Despite the increasing evidence to support these approaches, some aspects of their application require further clarifications. First, there is no consensus on which NIT thresholds to use for this purpose. For example, FIB-4 upper cut-offs of 3.25[11] and 2.67[12] have been described, while other investigators omit the FIB-4 upper cut-off altogether.[13] There is also some uncertainty about the performance of NITs in specific patient subgroups, such as those with diabetes or obesity. Furthermore, for patients who are ruled in as being at high risk of advanced fibrosis (F3–4), liver biopsy is often needed to identify those with cirrhosis who would need surveillance for HCC.[14] Developing approaches that can minimise the need for liver biopsy in secondary care is therefore an area of unmet need.

To address these problems, we conducted an individual patient data meta-analysis (IPDMA) with three main aims: (1) to evaluate the performance of LSM-VCTE and compare it to the performance of FIB-4 and NFS as screening tests to rule out advanced fibrosis; (2) to evaluate NIT combination strategies to minimise the number of cases that would need a liver biopsy in secondary care; (3) to explore factors that influence diagnostic accuracy.

METHODS

This IPDMA was reported in accordance with the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-IPD Statement[15] and was registered as PROSPERO CRD42019157661.

Criteria for considering studies for the IPD meta-analysis

Patients

Studies reporting data on adults (≥18 years) with NAFLD and paired liver histology and LSM-VCTE were eligible. When studies reported study groups of participants with unselected aetiologies, only IPD of those with NAFLD were sought.

Index tests

The index test of main interest was LSM-VCTE performed with FibroScan (Echosens, France). Results for serum-based biomarkers NSF,[16] FIB-4,[17] aspartate aminotransferase (AST) to alanine aminotransferase (ALT) ratio[18] and AST-to-platelet ratio index (APRI)[19]) were also computed where data were available. Online supplemental table 1 summarises the definition of NITs considered in this IPDMA.

Universally accepted cut-offs for diagnosing different groups of fibrosis stages do not exist (several suggested cut-offs are presented in online supplemental table 2). For LSM-VCTE, <7.9 kPa and ≥9.6 kPa are the most used for respectively ruling out and in, advanced fibrosis.[20]

Reference standard

Only studies reporting histological classification of liver fibrosis based on the non-alcoholic steatohepatitis Clinical Research Network (NASH CRN) staging system were considered.[21]

Target conditions

Advanced fibrosis (F3–4) and cirrhosis (F4) were the target conditions of interest. To fulfil the aims of the study, cut-offs

were selected to rule out or rule in advanced fibrosis, and to rule out advanced fibrosis or rule in cirrhosis.

Study design

All study designs were considered if they were reporting on patients with NAFLD undergoing both liver biopsy and LSM-VCTE within 6 months. No language restrictions were applied.

Establishing collaborations

Authors of eligible studies were contacted by email and reminders were sent if a response was not received within 2 weeks. Only data from studies that received ethical approval were used. Additional ethical approval was not sought for the meta-analysis as only anonymised data were provided.

Data verification

Range checks of measurement values provided for individual patients were carried out and authors were asked to provide clarifications where necessary. Missing data were queried until received or confirmed as unavailable. Missing data were handled in the analysis by pairwise deletion.

LSM-VCTE with median stiffness ≥7.1 kPa and IQR-to-median LSM ratio >30% were considered unreliable.[22] These were included in the main analysis and were later compared in a subgroup analysis to reliable measurements, to assess whether they can be reliably used to diagnose advanced fibrosis.

Authors were provided with a template table of required data (online supplemental table 3) and were asked to deduplicate data were possible. We also checked for duplicate entries and where identified these were removed.

Data analysis

Quality and bias assessment

The quality of studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2).[23]

IPD meta-analysis

The original data sets were merged, a study identification variable was added, and descriptive statistical analysis of the data sets was conducted. Dichotomous variables are displayed as percentages. Continuous variables are reported as means with SD, or medians with IQRs according to the distribution of the data.

Analyses were done per protocol, as we did not have information on failed LSM-VCTE. To express the diagnostic performance of NITs, non-parametric, empirical receiver operating characteristic (ROC) curves were constructed for the target conditions of interest. Diagnostic performance was expressed as the area under the ROC curve (AUROC) with 95% CI, based on De Long's method. AUROCs were compared using De Long's test statistic.

Thresholds to maximise the Youden index (ie, sensitivity+specificity−1), for 90% sensitivity, and for 90% specificity were reported. The diagnostic performance of previously published cut-offs was also evaluated. Sequential combinations of serum biomarkers and LSM-VCTE were evaluated, by computing sensitivity, specificity and proportions of misclassified and indeterminate patients.

Positive and negative predictive values (PPV and NPV) were estimated for prevalences within the range of those reported in the original studies. The number of false positive and false negative results for 100 theoretical cases was also reported.

The main analysis was conducted to maximise data for each NIT. For a valid comparison of the performance of NITs, a separate analysis was conducted in the subgroup of patients where all three of VCTE, FIB-4 and NFS were available in each participant.

To fulfil the aim of developing testing strategies that reduce the number of patients in need of a liver biopsy, lower cut-offs for ruling out advanced fibrosis and upper cut-offs for ruling in cirrhosis were used. The rationale for this approach is illustrated in online supplemental figure 1. The upper cut-offs for identifying cirrhosis were chosen at 95% and 98% specificity in a derivation set and tested in a validation set. Derivation and validation sets were obtained by random sampling from the IPD study group in a 3:2 ratio. These upper cut-offs were combined with lower cut-offs from the literature for ruling out advanced fibrosis and the algorithm was tested in the whole IPD study group. For ease of reference, we also examined the cut-offs of 8 kPa and 10 kPa (corresponding to the most common VCTE cut-offs in the literature of 7.9 kPa and 9.6 kPa rounded to the nearest integer) and also rounded our cirrhosis cut-offs to the nearest integer to facilitate application in clinical practice.

Only test-positive and test-negative patients were included in the calculation of diagnostic performance indices, and patients in the indeterminate group were excluded from calculations.

Subgroup analysis was performed according to biopsy length (<20 mm, ≥20 mm), number of portal tracts in biopsy samples (<11, ≥11), biopsy quality (intermediate: 10 mm ≤length <20 mm; high: length ≥20 mm and ≥11 tracts), age (four quartiles), sex, body mass index (BMI; BMI <25 kg/m^2, 25 kg/m^2≤ BMI <30 kg/m^2, BMI ≥30 kg/m^2), presence of type 2 diabetes mellitus (T2DM), continent of provenance (Europe, Asia), probes used (M, XL), reliability criteria for LSM-VCTE (reliable (median LSM <7.1 kPa or median LSM ≥7.1 kPa and IQR/median LSM <0.30) vs unreliable (median LSM ≥7.1 kPa and IQR/median LSM ≥0.30)[22] ; reliable (IQR/median LSM <0.30) vs unreliable (IQR/median LSM ≥0.30)), and aminotransferase levels (ALT or AST<40, 40≤ALT or AST<100, ALT or AST≥100; ALT<40 and AST<40, ALT≥40 or AST≥40.

All statistical analyses were performed using R (V.1.2.1335, R Foundation for Statistical Computing, Vienna, Austria) with the pROC package[24 25]; 95% CIs were calculated using 500 stratified bootstrap replicates using the boot package.[26 27]

VCTE probe types

The analysis to account for probe type is described in the online supplemental materials.

Patient and public involvement

Patients and the public were not involved in the conduct of this study as there was no direct patient participation in the study.

RESULTS

Search process and data collection

Ten thousand three hundred ninety-two articles were identified in a search performed for a larger systematic review evaluating the diagnostic performance of LSM-VCTE and other index tests for the staging of fibrosis and diagnosis of NASH in adult patients with NAFLD. After removing duplicates, and screening titles, abstracts, and full texts, 59 studies examining VCTE were identified. The authors of 37 studies shared useable data (figure 1). Authors of more than one study supplied data in a single dataset and, overall, we received 30 data sets including data from 6571 patients. After removing duplicates (n=628) and patients with

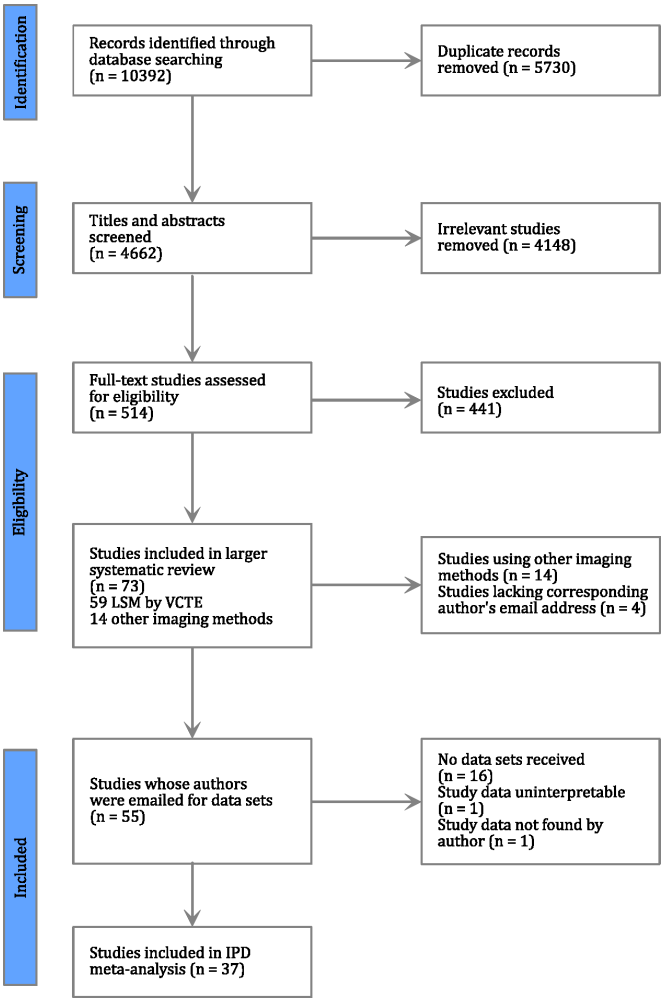


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart illustrating the identification and selection process for studies finally included in this individual patient data meta-analysis. IPD, individual patient data; LSM, liver stiffness measurement; VCTE, vibration controlled transient elastography.

missing biopsy (n=14) or LSM-VCTE (n=194) data, the final dataset consisted of 5735 unique patients.

Study and population characteristics

The characteristics of the 30 data sets are summarised in table 1. Studies were conducted in Europe (67%), Asia (40%) and Australia (3%). Data availability is shown in online supplemental table 3. FIB-4 and NFS were determined in 5393 (94%) and 3248 (57%) cases, respectively. Median age was 54 years, 2570 (45%) patients were women, 33% had diabetes and 43% had BMI ≥30 kg/m^2. Overall, 30% had advanced fibrosis and 11% had cirrhosis. Details of the IPD study group are included in table 2, and online supplemental tables 4 and 5.

Study quality

The methodological quality of the studies assessed with the QUADAS-2 tool is summarised in online supplemental figures 2 and 3. Only one study had low risk of bias or low applicability concerns in all QUADAS-2 domains.[28] The flow and timing domain were judged to have high risk or unclear risk of bias in 65% of studies, as these either excluded technical failures from their final diagnostic performance analysis or did not report them.

Validating the diagnostic performance of LSM by VCTE and serum-based tests for detecting advanced fibrosis

LSM-VCTE, FIB-4, NFS, APRI and AST/ALT had corresponding AUROCs of 0.85, 0.76, 0.73, 0.70, 0.64 for identifying advanced fibrosis (table 3), and 0.90, 0.80, 0.78, 0.72, 0.69 for the identification of cirrhosis (online supplemental table 6). LSM-VCTE performed significantly better ($p<10^{-15}$) in detecting both advanced fibrosis and cirrhosis than all serum-based tests. This relationship was preserved when performing a head-to-head comparison of LSM-VCTE, FIB-4 and NFS in the same group of patients (online supplemental tables 7 and 8).

When considering cut-offs from the literature, we evaluated lower and higher cut-offs separately. For any given test, as would be expected, low thresholds yielded higher sensitivity and high thresholds were associated with higher specificity (online supplemental table 9). Indicative PPV and NPV are also provided for the range of prevalences (5%–50%) reported in the primary studies (online supplemental tables 10–14).

APRI and AST/ALT ratio had only modest diagnostic performance for advanced fibrosis (AUROC ≤0.70, table 3), and were therefore not considered further.

None of the thresholds regarded in isolation resulted in both a high sensitivity (≥80%) and high specificity (≥80%) (figure 2, table 3, online supplemental tables 9 and 15, and online supplemental figure 4). Therefore, we explored the use of a lower and an upper cut-off. LSM-VCTE literature cut-offs performed well in only two cases (<7.1 kPa and ≥14.1 kPa: 83% sensitivity, 90% specificity; and <7.9 kPa and ≥9.6 kPa: 84% sensitivity, 78% specificity), while for other LSM-VCTE, NFS and FIB-4 thresholds a high specificity was observed (FIB-4: 91% for <1.3 and ≥2.67, 95% for <1.3, ≥3.25) but sensitivity was <60% (table 4). In addition, the proportion of indeterminate cases was >30% for serum-based NITs. Threshold pairs derived from the IPD study group did not reduce the proportion of misclassified and indeterminate patients seen with literature-based threshold pairs (table 4).

We further evaluated the performance of LSM-VCTE, FIB-4 and NFS to diagnose advanced fibrosis in sequential combinations of serum-based NITs and LSM-VCTE. When selecting threshold combinations for FIB-4 and NFS available in the literature (<1.3 & ≥2.67,<1.3 & ≥3.25 for FIB-4;<−1.455 & ≥0.676 for NFS) and pairing them with the best threshold pair for LSM-VCTE (<7.9 kPa & ≥9.6 kPa, identified as the one with highest sensitivity and lowest indeterminate proportion), the proportion of patients in the indeterminate group was 5%. While both the FIB-4+LSM VCTE and NFS+LSM VCTE sequential combinations had specificity >80%, their sensitivity was ≤80% (table 5). A better sensitivity was reached by using thresholds derived from the IPD study group (<0.88 & ≥2.31 for FIB-4;<−2.55 & ≥0.28 for NFS), but the proportion of indeterminate cases was near 20% in those cases and the proportions of patients needing LSM-VCTE was also larger than when using literature cut-offs (table 5).

Algorithms to minimise the need for liver biopsy

In the derivation set, the cut-offs for 95% and 98% specificity for the diagnosis of cirrhosis were respectively 20.4 kPa and 27.6 kPa for LSM-VCTE, 3.48 and 4.63 for FIB-4 and 1.01 and 1.57 for NFS. These cut-offs performed similarly in the validation set (online supplemental tables 16 and 17).

Table 1 Details of individual patient data included in this meta-analysis

Data set ID	Country	Study design	Number of participants (n)	Age (year)	BMI (kg/m^2)	WC (cm)	M/F	Recruitment interval	Hardware used	Probe used
Agrawal *et al*[49]	UK	MC, P, CS	25	47.8 (19–70)	27.7 (15.8–35.7)	95.4 (39–120)	18/7	2009–2012	–	M
Aykut *et al*[50]	Turkey	SC, P, CS	88	46.0 (24–62)	30.3 (18.3–41.8)	101.5 (70–143)	50/38	–	FibroScan 502 Touch	M
Boursier *et al*[51–53]	France	MC, P, CS	1063	56.1 (18–83)	31.6 (16.7–55.5)	108.3 (58–174)	613/450	–	–	M or XL
Cassinotto *et al* [54]	France	SC, P, CS	61	55.9 (22–81)	30.1 (16.7–46.6)	103.6 (72–125)	40/21	2010–2012	–	M and XL
Cassinotto *et al* [55]	France	MC, P, CS	286	56.6 (18–80)	32.2 (20.3–57.4)	109.8 (68–168)	171/115	2011–2015	–	M and XL
Chan *et al*[56]	Malaysia	SC, P, CS	146	50.4 (18–73)	29.4 (6.9–41.2)	98.3 (79–127)	80/66	2012–2013	FibroScan 502 Touch	M
Chan *et al*[57]	Malaysia, Hong Kong	MC, P, CC	153	54.0 (24–76)	29.9 (20.1–44.8)	98.4 (69–141)	68/85	–	FibroScan 502 Touch	M and XL
Eddowes *et al*[31 58]	UK	MC, P, CS	358	53.3 (19–77)	34.2 (19.5–53.2)	117.2 (65–158)	206/152	–	–	M or XL
Eddowes *et al*[59]	UK	MC, P, CS	50	50.2 (18–73)	33.6 (23.6–47.8)	109.4 (89–132)	28/22	2014–2015	–	M or XL
Gaia *et al*[60]	Italy	SC, P, CS	68	46.8 (28–65)	28.0 (21.2–40.2)	–	48/20	2007–2009	–	M
Garg *et al*[61]	India	SC, P, CS	76	38.2 (20–65)	45.2 (32.3–73.8)	–	16/60	2014–2016	FibroScan 502 Touch	XL
Karlas *et al*[62]	Germany	SC, P, CS	41	45.7 (28–64)	47.7 (33.7–60.1)	–	13/28	–	FibroScan 502	XL
Labenz *et al*[63]	Germany	SC, P, CS	126	47.4 (20–73)	31.6 (23.2–50.4)	–	72/54	–	FibroScan 402	M or XL
Lee *et al*[64]	Korea	SC, P, CS	94	55.5 (19–82)	27.2 (19.1–36.3)	–	41/53	2014–2015	–	M or XL
Lupsor *et al*[65]	Romania	SC, P, CS	72	42.4 (20–69)	29.7 (21.0–41.5)	102.4 (60–124)	51/21	2007–2009	–	M
Mahadeva *et al*[66]	Malaysia	SC, P, CS	131	49.9 (23–73)	28.7 (18.6–43.1)	93.5 (43–128)	66/65	2009–2010	–	M
Okajima *et al*[67]	Japan	SC, P, CS	173	56.3 (18–81)	27.2 (16.5–40.3)	–	84/89	2013–2015	–	M
Ooi *et al*[68]	Australia	MC, P, CS	82	44.5 (18–67)	46.2 (29.1–74.0)	136.5 (101–192)	23/59	2015–2016	–	M or XL
Pavlides *et al*[69]	UK	SC, P, CS	70	53.5 (25–77)	34.5 (23.0–57.3)	112.5 (80–149)	42/28	2011–2015	–	M or XL
Petta *et al*[70 71]	Italy	MC, P&R, CS	234	45.5 (15–78)	28.2 (15.7–40.7)	99.4 (69–126)	169/65	2008–2013	–	M
Petta *et al*[28]	France, Hong Kong, Italy	MC, P, CS	260	54.6 (15–87)	29.4 (16.5–46.6)	100.9 (74–148)	122/138	–	–	M
Petta *et al*[72]	Italy	MC, P, CS	474	45.5 (19–77)	29.2 (15.2–49.5)	99.6 (47–164)	275/199	–	–	M
Seki *et al*[73]	Japan	SC, P, CS	181	57.7 (16–82)	27.1 (16.9–38.1)	95.1 (71–117)	91/90	2013–2015	–	M
Shen *et al*[74]	China	MC, P, CS	101	59.0 (16–67)	27.0 (20.1–37.3)	92.9 (75–120)	74/27	2012–2014	FibroScan 502	M
Staufer[75]	Austria	MC, P, CS	186	49.6 (19–83)	32.5 (19.0–56.9)	–	106/80	2011–2016	FibroScan 502 Touch	M or XL
Wong *et al*[76–79]	Hong Kong, France	MC, P, CS	464	53.8 (20–83)	30.5 (17.3–48.0)	102.0 (71–148)	201/263	2009–2017	–	M and XL
Wong *et al*[20]	Hong Kong, France	MC, P, CS	273	51.6 (21–77)	28.8 (16.5–54.0)	96.2 (65–144)	147/126	2003–2009	–	M
Yoneda *et al*[80]	Japan	MC, P, CS	97	52.1 (19–76)	26.5 (17.9–38.5)	–	41/56	<2008	–	M
Younes *et al*[81]	Italy	MC, P, CS	289	44.8 (15–78)	28.8 (17.5–41.7)	98.9 (47–128)	199/90	–	–	M
Ziol *et al*[82]	France	SC, P, CS	13	49.3 (39–60)	29.4 (23.8–34.6)	–	10/3	2003–2005	–	–

–, Data not available; BMI, body mass index; CC, case-control; CS, cross-sectional; F, females; M, males; MC, multicentre; P, prospective; R, retrospective; SC, single-centre; WC, waist circumference.

Algorithms combining FIB-4 (lower cut-off of 1.3 as described in the literature and upper cut-offs of 3.48 and 4.63 as described above) and LSM by VCTE (lower cut-off rounded to 8.0 kPa and upper cut-offs rounded to 20.0 kPa and 28.0 kPa, as described above) were then compared with the traditional way of applying these tests, also with rounded cut-offs for LSM by VCTE (8 kPa and 10 kPa) (figure 3). This approach increased the number of patients requiring a LSM

Table 2 Demographic details of the entire cohort, and patients without (F0–2) and with (F3–4) advanced fibrosis

	Entire cohort (N=5735)	F0–2 (N=4013)	F3–4 (N=1722)
Females (%)	45	43	48
BMI ≥30 kg/m^2 (%)	43	45	53
Waist circumference (cm)	103 (15)	102 (15)	106 (14)
Diabetes (%)	33	30	58
Age (years)*	54 (19)	50 (19)	59 (14)
BMI (kg/m^2)*	30 (7)	29 (8)	30 (7)
Biopsy data			
Steatosis			
S0/S1/S2/S3 (%)	3/35/36/26	3/36/36/25	2/32/38/28
Ballooning			
B0/B1/B2 (%)	24/47/29	30/49/21	10/45/45
Inflammation			
I0/I1/I2/I3 (%)	13/60/24/3	17/62/20/1	5/55/34/6
NAS score†	4 (2)	4 (2)	5 (1)
NASH (%)	50	43	67
Liver function tests			
ALT (IU/L)*	55 (48)	53 (48)	60 (48)
AST (IU/L)*	40 (30)	36 (25)	50 (34)
Platelets (×10^9/L)†	230 (72)	241 (67)	205 (75)
Albumin (g/L)†	43 (9)	43 (7)	43 (13)
GGT (IU/L)*	69 (87)	62 (78)	87 (102)
NITs			
LSM (kPa)*	10.7 (6.1)	6.7 (3.5)	13.3 (12.0)
FIB-4*	1.7 (1.2)	1.1 (0.9)	1.9 (1.7)
NFS†	−1.5 (1.7)	−1.9 (1.6)	−0.6 (1.8)
APRI*	0.6 (0.4)	0.4 (0.3)	0.6 (0.6)
AST/ALT*	0.8 (0.4)	0.7 (0.4)	0.8 (0.5)

*Data are reported as median (IQR).
†Data are reported as mean (SD).
ALT, alanine aminotransferase; APRI, AST-to-platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index ; FIB-4, Fibrosis-4 Index; GGT, gamma-glutamyltransferase; LSM, liver stiffness measurement; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; NFS, NAFLD (non-alcoholic fatty liver disease) Fibrosis Score; NIT, non-invasive test.

(from 34% to 40% and 44%) but decreased the number of patients needing liver biopsy (from 33% to 19% and 24% when using the 95% and 98% specificity cut-offs, respectively) (online supplemental table 18 and figure 3).

Subgroup and sensitivity analyses

In subgroup analysis for the diagnosis of advanced fibrosis (online supplemental table 19), NITs performed better in patients with lower BMI (AUROCs LSM-VCTE: 0.91, p<0.005; FIB-4: 0.81, p<0.001; NFS: 0.76, p<0.025), without T2DM (LSM-VCTE: 0.87, p<10^{-6}; FIB-4: 0.77, p<0.01), and with biopsies shorter than 20 mm (LSM-VCTE: 0.87, p<0.005; FIB-4: 0.80, p<0.001; NFS: 0.79, p<0.05), or with fewer than 11 portal tracts (LSM-VCTE: 0.86, p=0.01; FIB-4: 0.79, p=0.04; NFS: 0.78, p<0.005). Diagnostic performance was also lower in patients in the youngest age quartile (<43 years, AUROC: 0.58, p<0.001) and in women (AUROC: 0.71, p=0.03) for NFS, while continent of provenance did not have a significant effect for any NITs. In patients with normal levels of ALT (ALT<40) FIB-4 performed worse (AUROC: 0.73) than in patients with ALT≥40 and ALT<100 (AUROC: 0.77, p<0.01). NFS

Table 3 Diagnostic performance of non-invasive tests for advanced fibrosis (F3–F4)

	LSM by VCTE (n=5489)			FIB-4 (n=5393)			NFS (n=3248)			APRI (n=5477)			AST/ALT (n=5434)		
Advanced fibrosis, %	30			30			29			30			30		
AUROC	0.85 (0.84–0.86)			0.76 (0.74–0.77)			0.73 (0.71–0.75)			0.70 (0.69–0.72)			0.64 (0.62–0.65)		
	YI	90% Se	90% Sp	YI	90% Se	90% Sp	YI	90% Se	90% Sp	YI	90% Se	90% Sp	YI	90% Se	90% Sp
Threshold	9.1	7.4	12.1	1.44	0.88	2.31	−1.39	−2.55	0.28	0.49	0.29	0.91	0.64	0.51	1.34
Sensitivity, %	77 (75–79)	90 (89–91)	55 (52–57)	69 (67–72)	90 (88–91)	38 (36–41)	75 (72–78)	90 (88–92)	29 (26–32)	67 (64–69)	90 (89–92)	32 (30–34)	75 (73–77)	90 (87–91)	16 (14–18)
Specificity, %	78 (76–79)	60 (59–61)	90 (89–91)	70 (69–72)	39 (37–40)	90 (89–91)	63 (61–65)	36 (33–37)	90 (89–91)	63 (62–65)	29 (28–30)	90 (89–91)	47 (45–48)	25 (23–26)	90 (89–91)
Misclassified, %	22 (22–23)	31 (31–32)	21 (20–21)	30 (30–31)	46 (46–47)	26 (25–26)	34 (34–36)	48 (49–50)	28 (28–29)	36 (36–37)	53 (53–54)	27 (27–28)	45 (45–46)	56 (56–57)	32 (32–33)

For each non-invasive test thresholds were selected according to Youden's index (YI), and fixed at 90% sensitivity (90% Se) and 90% specificity (90% Sp). 95% CIs were estimated with 500 bootstrap replicates.
ALT, alanine aminotransferase; APRI, AST-to-platelet ratio index; AST, aspartate aminotransferase; FIB-4, Fibrosis-4 Index; LSM, liver stiffness measurement; NFS, NAFLD (non-alcoholic fatty liver disease) Fibrosis Score; VCTE, vibration controlled transient elastography.

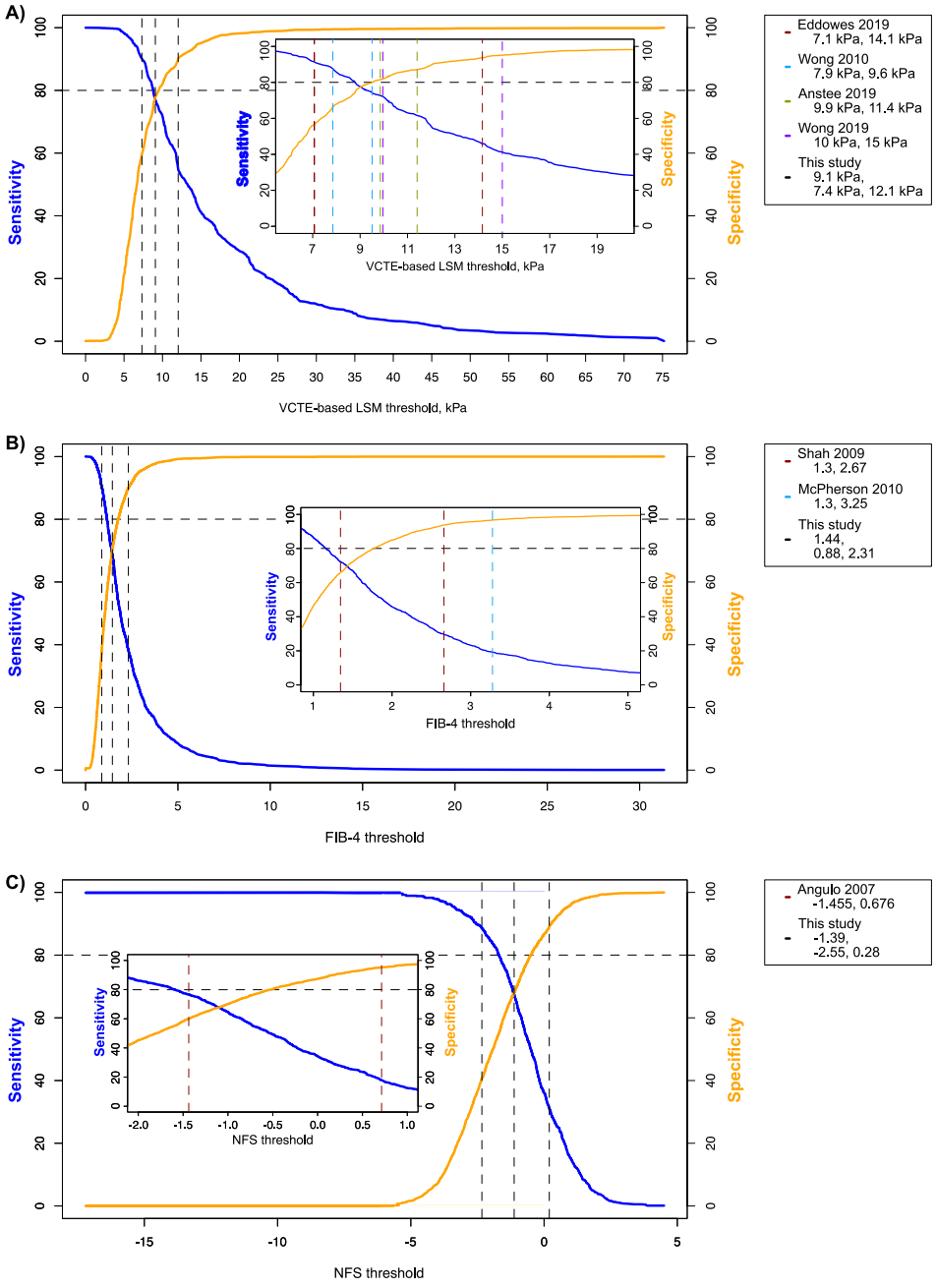


Figure 2 Distribution of sensitivities and specificities over the possible threshold ranges for liver stiffness measurement (LSM) by vibration controlled transient elastography (VCTE) (A), Fibrosis-4 Index (FIB-4) (B) and NAFLD (non-alcoholic fatty liver disease) Fibrosis Score (NFS) (C) when considering the diagnosis of advanced fibrosis. Insets show the distribution of cut-offs identified from the literature. Horizontal dashed lines are representing the minimum acceptable criteria for considering a test as having high sensitivity (≥80%) and high specificity (≥80%).

performed better in patients with AST<40 (AUROC: 0.76), than in patients with AST≥100 (AUROC: 0.65, p<0.01). FIB-4 performed better in patients with at least one abnormal aminotransferase measurement (AUROC: 0.72, p=0.014). For cirrhosis, the trends were similar, except that for the diagnosis of cirrhosis, LSM by VCTE performed better in the youngest age group (AUROC: 0.97, $p<10^{-4}$) and NIT diagnostic performance was independent of aminotransferase levels (online supplemental table 20).

The diagnostic performance of LSM-VCTE was significantly lower in patients with unreliable LSMs ($p<10^{-8}$; both for advanced fibrosis and cirrhosis) when applying the Boursier-criteria,[22] but not when only considering IQR/median LSM <0.30. The proportion of unreliable results was 12% both in the advanced fibrosis and cirrhosis groups (online supplemental table 21).

There was no difference in the diagnostic performance of LSM-VCTE between the M and XL probes in the subgroup of patients who had undergone LSM by both probes (online supplemental table 22).

In a sensitivity analysis of patients with LSM matched to BMI (only M probe measurements if BMI <30 kg/m^2 and only XL probe measurements if BMI ≥30 kg/m^2), there was no significant difference between the diagnostic performance of LSM-VCTE when comparing to the entire IPD study group (online supplemental table 23).

Table 4 Diagnostic accuracy of pairs of cut-offs from the literature for liver stiffness measurement (LSM) by vibration controlled transient elastography (VCTE), Fibrosis-4 Index (FIB-4) and NAFLD (non-alcoholic fatty liver disease) Fibrosis Score (NFS) for diagnosing advanced fibrosis

	LSM by VCTE (n=5489)				FIB-4 (n=5393)			NFS (n=3248)	
Advanced fibrosis, %	30				30			29	
AUROC	0.85 (0.84–0.86)				0.76 (0.74–0.77)			0.73 (0.71–0.75)	
Source	Anstee *et al*[29]	Eddowes *et al*[31]	Wong *et al*[76]	Wong *et al*[20]	This study	Shah *et al*[83]	McPherson *et al*[84]	This study	Angulo *et al*[16]	This study
Thresholds	<9.9, ≥11.4	<7.1, ≥14.1	<10, ≥15	<7.9, ≥9.6	<7.4, ≥12.1*	<1.3, ≥2.67	<1.3, ≥3.25	<0.88, ≥2.31*	<−1.455, ≥0.676	<−2.55, ≥0.28*
Sensitivity, %	69 (67–71)	83 (80–86)	59 (57–61)	84 (82–87)	84 (81–87)	54 (52–56)	44 (42–46)	80 (76–83)	47 (44–50)	74 (70–79)
Specificity, %	86 (85–88)	90 (88–92)	94 (93–96)	78 (76–80)	87 (85–88)	91 (89–92)	95 (93–96)	79 (77–81)	91 (89–93)	78 (76–81)
Misclassified, %	17 (16–19)	7 (6–8)	12 (11–13)	17 (16–19)	10 (9–11)	12 (11–13)	10 (9–11)	10 (9–11)	11 (10–13)	10 (8–11)
Indeterminate, %	7 (6–8)	39 (37–40)	18 (17–19)	13 (12–14)	31 30–33)	34 (33–35)	39 (37–40)	52 (50–53)	39 (37–41)	56 (54–59)

*Cut-offs determined from the individual patient data study group. Lower cut-offs correspond to a lower limit of 90% sensitivity, upper cut-offs correspond to a lower limit of 90% specificity. 95% CIs were determined with 500 bootstrap replicates.
AUROC, area under the receiver operating characteristic.

DISCUSSION

Through an extensive collaboration network with authors of primary studies we were able to collect the largest dataset of its kind ever to be reported on. This includes a diverse set of study groups from Europe, Asia, and Australia, 30% of whom had advanced fibrosis. We believe that our findings are therefore relevant for patients typical of secondary care in these territories and may be applied in the development of new strategies or in the consolidation of existing practices in evaluating patients for referral to secondary care.

A few studies evaluated the diagnostic performance of LSM-VCTE and other NITs, but most report on fewer than 500 patients. One similarly large study reported on patients screened for inclusion in clinical trials, where the prevalence of advanced fibrosis was 71%,[29] making it difficult to make generalisations about its applicability in routine practice or compare its results to ours. A smaller study with 1073 patients with NAFLD of whom 29% had advanced fibrosis[30] examined the diagnostic performance of LSM by VCTE. The authors of that study reported AUC and specificity values similar to our findings, however they reported increased sensitivity. Other smaller studies reported similar prevalence of advanced fibrosis and similar AUROCs for LSM-VCTE.[31–34]

Overall, the diagnostic performance of LSM-VCTE for advanced fibrosis was good (AUROC=0.85), while that of FIB-4 and NFS in the same group was moderate (AUROC=0.76 for FIB-4, AUROC=0.73 for NFS). None of the studied NITs had both sufficiently high sensitivity and specificity (≥80%) when used with single cut-offs. Diagnostic performance was higher for detecting cirrhosis, as reported in previous studies.[31 35 36] LSM-VCTE had the highest sensitivity and specificity, both in the case of a single cut-off (9.1 kPa obtained by maximising the Youden index; 77% and 78%) and for two cut-offs (<7.4 kPa & ≥12.1 kPa; 84% and 87%). Of the LSM-VCTE cut-off pairs tested,<7.1 kPa and ≥14.1 kPa, first published by Eddowes *et al*.,[31] performed well for advanced fibrosis, with sensitivity of 83% and specificity of 90%, but with a proportion of 39% of patients ending up with an indeterminate result, similar to 41% indeterminate patients reported in the original paper.[31]

LSM-VCTE thresholds identified in our study group (<9.1 kPa; <7.4 kPa & ≥12.1 kPa) were similar to thresholds reported in the literature (<9.9 kPa; <7.1 kPa & ≥14.1 kPa, <7.9 kPa & ≥9.6 kPa). However, thresholds for FIB-4 (<1.44; <0.88 & ≥2.31) and NFS (<−1.39; <−2.55 & ≥0.28) defined in our IPD study group spanned a wider range than those reported in the literature (<1.3 & ≥2.67 or <1.3 & ≥3.25 for FIB-4;<−1.455 & ≥0.676 for NFS).

Our findings are in line with the existing literature suggesting that sequential combinations of NITs increase sensitivity and specificity.[29] Additionally, we have found NFS+LSM VCTE and FIB-4+LSM VCTE combinations to have similar sensitivity and specificity as recently reported by Boursier *et al*.[37] Such combined testing strategies can reduce the number of indeterminate cases and reduce the costs associated with liver biopsies.

Furthermore, we propose an approach that could minimise the need for liver biopsies further, by using upper cut-offs with 95% and 98% specificity for the identification of cirrhosis. The rationale for this approach is explained in the online supplemental discussion. When using the 95% specificity cut-off, the proportion of patients needing liver biopsy decreases from 33% to 19% (figure 3). However, in this approach, 345 of 656 patients 'ruled-in' as having cirrhosis do not have histologically diagnosed cirrhosis. While this may seem like a high proportion

Table 5 Diagnostic performance of combinations of NAFLD (non-alcoholic fatty liver disease) Fibrosis Score (NFS) and liver stiffness measurement (LSM) by vibration controlled transient elastography (VCTE), and Fibrosis-4 Index (FIB-4) and LSM by VCTE tests to diagnose patients with advanced fibrosis

	FIB-4 & LSM by VCTE (n=5159)	NFS & LSM by VCTE (n=3094)	FIB-4 & LSM by VCTE (n=5159)	NFS & LSM by VCTE (n=3094)	FIB-4 & LSM by VCTE (n=5159)	NFS & LSM by VCTE (n=3094)
Advanced fibrosis, %	30	28	30	28	30	28
Thresholds for blood-based NIT	<0.88, ≥2.31*	<−2.55, ≥0.28*	<1.3, ≥2.67†	<−1.455, ≥0.676†	<1.3, ≥2.67†	<−1.455, ≥0.676†
Thresholds for LSM by VCTE, kPa	<7.4, ≥12.1*	<7.4, ≥12.1*	<7.9, ≥9.6†	<7.9, ≥9.6†	<8.0, ≥10.0†	<8.0, ≥10.0†
Sensitivity, %	80 (77–83)	77 (74–81)	67 (64–69)	65 (62–68)	66 (63–68)	64 (62–67)
Specificity, %	81 (79–83)	83 (81–85)	85 (84–87)	86 (84–88)	86 (84–87)	86 (84–88)
PPV, %	62 (60–65)	61 (58–64)	66 (64–68)	63 (61–67)	66 (64–68)	64 (61–67)
NPV, %	91 (90–92)	91 (89–93)	86 (85–87)	87 (85–88)	86 (85–87)	86 (85–88)
Indeterminate, %	18 (17–19)	20 (18–21)	5 (4–5)	5 (5–6)	5 (4–6)	5 (5–6)
Misclassification, %	16 (14–17)	15 (13–17)	19 (18–21)	19 (17–21)	19 (18–20)	19 (17–21)
Patients undergoing LSM by VCTE, %	51 (50–53)	56 (54–59)	34 (32–35)	38 (36–40)	34 (33–35)	38 (37–40)

95% CIs were estimated with 500 bootstrap replicates.
*Thresholds were determined from the individual patient data study group as corresponding to 90% sensitivity (lower value) and 90% specificity (upper value)
†Threshold were determined from the literature. For LSM by VCTE, a threshold pair yielding the highest sensitivity and specificity while having the smallest proportion of indeterminate cases in diagnosing advanced fibrosis was chosen.
NIT, non-invasive test; NPV, negative predictive value; PPV, positive predictive value.

of patients with false positive results, this must be interpreted in the light of two factors. First, the limitations of liver biopsy could mean that these patients are falsely classified as not having cirrhosis histologically. Furthermore, patients without cirrhosis on histology and with high NIT values could have equivalent risks as patients with cirrhosis on histology. For example, it is known from the hepatitis C literature[38] that patients without cirrhosis on liver biopsy but with a high FIB-4 (>3.25) still had a significant risk of developing HCC after hepatitis C treatment, demonstrating that NITs can have added benefit beyond the histological diagnosis of cirrhosis alone. The rate of false positive results for cirrhosis can be decreased by choosing cut-offs with higher specificity, but this will come at the expense of doing more biopsies. Despite this encouraging result, this is an area where more information is needed, particularly longitudinal data comparing the prognostic value of LSM-VCTE and other NITs against histology, and ultimately, the cost effectiveness of the various cut-offs would need to be evaluated.

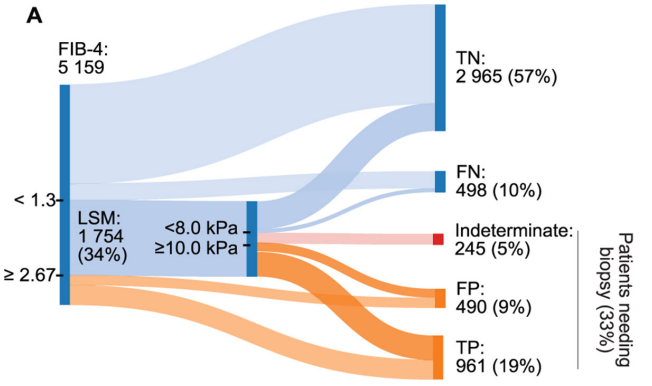
Surprisingly, subgroup analyses showed that the diagnostic accuracy of NITs was better in cases with poor biopsy quality. This finding is difficult to explain but a similar observation was reported previously in a large group of patients screened for clinical trials.[29] The use of local biopsy reports as reference standard and the well-known observer-dependent variability of biopsy interpretation, even among expert pathologists,[7] are factors that may have contributed to our finding. Spectrum bias was excluded as a source of this finding due to a near-identical proportion of patients in both the advanced fibrosis and cirrhosis group having short biopsies (online supplemental table 5).

Subgroup analysis showed better diagnostic performance of NITs in patients with lower BMI,[39 40] and patients without diabetes, in keeping with other studies.[41 42] This effect is likely to be primarily driven by BMI as there is thought to be a causal association between BMI and T2DM. NIT performance was impacted by age, with all NITs performing worse in the younger quartile of our study group for advanced fibrosis, but the trend was reversed for cirrhosis where NITs performed better in those younger than 43 years of age. The age dependence of FIB-4 and NFS is expected, as age is one of the parameters included in the algorithms, and has indeed been previously described.[13 43] It is, however, difficult to explain why performance of NITs is better in the younger age group for the diagnosis of cirrhosis.

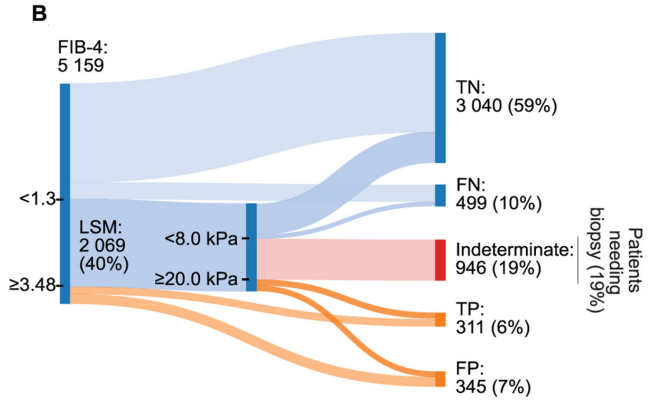
Our study has several strengths, including the large size of the IPD study group and composition with prevalence of advanced fibrosis of 30%, which makes it relevant to routine practice. Furthermore, the proportion of unreliable VCTE measurements in our study was 12%, in keeping with the literature.[22] However, we acknowledge some limitations. We did not have any data from the USA and very few studies from Australia, so the results could not be globally applicable, due to differences in BMI across study populations. In addition, due to the nature of our study, we had to use the locally provided histology results possibly introducing bias. Furthermore, we covered a large chronological period, during which LSM-VCTE application underwent significant changes, initially with the introduction of the XL probe, followed by the advice to measure skin-to-capsule distance (SCD) and the introduction of the Automatic Probe Selection tool. There was therefore some heterogeneity in the performance of LSM-VCTE, with early studies using only the M probe to assess all patients, while only a subset of studies assessed SCD to guide probe selection. Furthermore, one third of the included studies was carried out in France, as the technology used for LSM by VCTE originates from there. Lastly, our data confirm that LSM-VCTE had superior accuracy to serum-based tests, and this is independent of probe type, sex, ALT, AST, and participants' continent of origin. There was, however, some dependence on the presence of T2DM, BMI and for the detection of cirrhosis, and we did not check for subgroup-specific cut-offs, but these should be explored in future studies.

Our study examined some of the most widely available NITs. While it cannot be considered exhaustive, it can be regarded as the benchmark against which newer NITs can be tested. This is particularly important as new tests are continuously being developed (FibroTest-FibroSURE, ActiTest,[44] ELF[45]). Furthermore, newer tests are also needed for patients with 'at risk' NASH (NASH+F2–3) who would be candidates for clinical trials or treatments, once approved therapies become available (FAST score,[46] NIS4,[47] cTAG[48]).

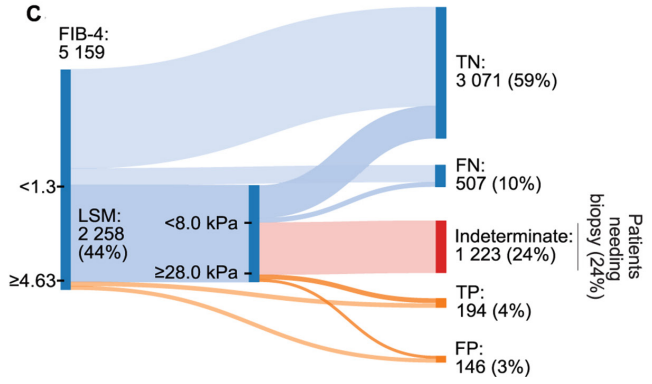
In conclusion, our study provides further validation of the use of sequential combination of FIB-4 and LSM-VCTE to rule out patients with NAFLD and advanced fibrosis who can be managed in primary care. We have shown how the use of upper cut-offs to rule in cirrhosis in combination with lower cut-offs to



		FIB-4	LSM
< Lower cutoff	TN	2311 (44%)	654 (13%)
	FN	396 (8%)	102 (2%)
≥ Upper cutoff	TP	459 (9%)	502 (10%)
	FP	239 (4%)	251 (5%)



		FIB-4	LSM
< Lower cutoff	TN	2311 (44%)	729 (14%)
	FN	396 (8%)	103 (2%)
≥ Upper cutoff	TP	169 (3%)	142 (3%)
	FP	214 (4%)	131 (3%)



		FIB-4	LSM
< Lower cutoff	TN	2311 (44%)	760 (15%)
	FN	396 (8%)	111 (2%)
≥ Upper cutoff	TP	102 (2%)	92 (2%)
	FP	92 (2%)	54 (1%)

Figure 3 Sankey diagrams showing the distribution of patients in true positive, true negative, false positive, false negative and indeterminate groups for a sequential combination of Fibrosis-4 Index (FIB-4) and liver stiffness measurement,(LSM) by vibration controlled transient elastography (VCTE) when using different thresholds for each testing tier. A lower threshold was used to rule out patients without advanced fibrosis and an upper threshold ruled in patients with advanced fibrosis when applying both tests (A). In an alternative model, a lower threshold was used to rule out patients without advanced fibrosis, but the upper threshold ruled in only patients with cirrhosis (B, C). Two different pairs of thresholds were chosen for this hybrid strategy: the lower cut-off for both FIB-4 and LSM by VCTE were determined from the literature; upper cut-offs were both determined as corresponding to 95% specificity in detecting cirrhosis (B) or both corresponding to 98% specificity in detecting cirrhosis (C). In the application of the algorithm described in (A) 33% of patients would need to have a liver biopsy for the diagnosis of cirrhosis (those in the indeterminate group to rule out advanced fibrosis and those in the rule in group to identify cirrhosis). With the application of an upper cut-off to rule in cirrhosis without the need of biopsy, only patients in the indeterminate group need to have a biopsy. The latter strategy results in fewer patients undergoing biopsy (18% and 24% depending on the threshold used). Tables next to each panel contain the number and proportion of patients in each of the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) groups for FIB-4 and LSM by VCTE.

rule out advanced fibrosis can lead to a reduction in the number of patients who would need to undergo liver biopsy.

Author affiliations
[1]Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[2]Department of Epidemiology and Data Science, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
[3]Translational Gastroenterology Unit, University of Oxford, Oxford, UK
[4]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust and the University of Oxford, Oxford, UK
[5]Division of Gastroenterology and Hepatology, Department of Medicine III, Medical University of Vienna, Vienna, Austria
[6]Laboratoire HIFIH, UPRES EA 3859, SFR ICAT 4208, Universite d'Angers, Angers, Pays de la Loire, France
[7]Service d'Hepato-Gastroenterologie et Oncologie Digestive, Centre Hospitalier Universitaire d'Angers, Angers, Pays de la Loire, France
[8]Echosens SA, Paris, Île-de-France, France

[9]Department of Visceral Surgery and Medicine, Inselspital University Hospital Bern, Bern, Switzerland
[10]Department of Surgery, Division of Transplantation, Medical University of Vienna, Vienna, Austria
[11]Department of Internal Medicine, Medical University of Graz, Graz, Austria
[12]Medical Sciences, University of Turin, Torino, Italy
[13]Boehringer Ingelheim International GmbH, Ingelheim, Rheinland-Pfalz, Germany
[14]Department of Ultrasonography, University of Medicine and Pharmacy, Regional Institute of Gastroenterology and Hepatology "Prof. Dr. Octavian Fodor", Cluj Napoca, Romania
[15]Section of Gastroenterology and Hepatology, PROMISE, Palermo, Italy
[16]Hepatology Center, Saiseikai Suita Hospital, Suita, Osaka, Japan
[17]Faculty of Medicine, Department of Medicine, University of Malaya, Kuala Lumpur, Wilayah Persekutuan, Malaysia
[18]NIHR Biomedical Research Centre, University of Birmingham, Birmingham, UK
[19]NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK
[20]Toronto Centre for Liver Disease, University Health Network, Toronto, Ontario, Canada
[21]Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK
[22]Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[23]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, Hong Kong
[24]Centre d'Investigation de la Fibrose Hépatique, Hopital Haut-Leveque, Pessac, France
[25]INSERM1053, Universite de Bordeaux, Talence, Aquitaine, France
[26]Department of Gastroenterology, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[27]Department of Internal Medicine, Aichi Medical University, Nagakute, Aichi, Japan
[28]Department of Gastroenterology, Koseikai Takeda Hospital, Kyoto, Japan
[29]Department of Internal Medicine I, University Medical Centre of the Johannes Gutenberg-University Mainz, Mainz, Rhineland-Palatinate, Germany
[30]Department of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul, South Korea
[31]Department of Radiology, Seoul National University Seoul Metropolitan Government Boramae Medical Center, Dongjak-gu, Seoul, The Republic of Korea
[32]Department of Medicine II, Leipzig University Medical Center, Leipzig, Sachsen, Germany
[33]Department of Gastroenterology, Marmara University School of Medicine, Istanbul, Turkey
[34]Institute of Gastroenterology, Marmara University, Istanbul, Turkey
[35]Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham, UK
[36]Diagnostic and Interventional Radiology, University Hospital Centre Montpellier, Montpellier, Languedoc-Roussillon, France
[37]Department of Surgical Disciplines, AIIMS, New Delhi, Delhi, India
[38]Department of Surgery, Monash University, Prahran, Victoria, Australia
[39]Department of Gastroenterology and Hepatology, Yokohama City University, Yokohama, Kanagawa, Japan
[40]Service d'Anatomie Pathologique et Centre de Ressources Biologiques, Hopital Jean Verdier, Paris, France
[41]Centre de Ressources Biologiques, Hopitaux Universitaires Paris-Seine-Saint-Denis, Bondy, Île-de-France, France
[42]Division of Hepatology, University Hospital Wurzburg, Wurzburg, Bayern, Germany
[43]Internal Medicine Research Unit, Pfizer Inc, Cambridge, Massachusetts, USA
[44]Translational and Clinical Research Institute, Faculty of Medicine, Newcastle University, Newcastle upon Tyne, UK

Correction notice This article has been corrected since it published Online First. The funding statement and table 1 have been updated and an author has been added.

Collaborators The LITMUS Investigators: Quentin Anstee, Ann Daly, Katherine Johnson, Olivier Govaere, Simon Cockell, Dina Tiniakos, Pierre Bedossa, Fiona Oakley, Heather Cordell, Chris Day, Kristy Wonders (Newcastle University); Patrick Bossuyt, Hadi Zafarmand, Yasaman Vali, Jenny Lee (AMC Amsterdam); Vlad Ratziu, Karine Clement, Raluca Pais (Hôpital Pitié Salpêtrière, Assistance Publique -Hôpitaux de Paris, and Institute of Cardiometabolism and Nutrition, Paris, France); Detlef Schuppan, Jörn Schattenberg (University Medical Center Mainz); Detlef Schuppan, Jörn Schattenberg (University Medical Center Mainz); Toni Vidal-Puig, Michele Vacca, Sergio Rodrigues-Cuenca, Mike Allison, Ioannis Kamzolas, Evangelia Petsalaki (University of Cambridge); Matej Oresic, Tuulia Hyötyläinen, Aiden McGlinchey (Örebro University); Jose M Mato, Oscar Millet (Center for Cooperative Research in Biosciences); Jean-François Dufour, Annalisa Berzigotti (University of Bern); Michael Pavlides, Stephen Harrison, Stefan Neubauer, Jeremy Cobbold, Ferenc Mozes, Salma Akhtar (University of Oxford); Rajarshi Banerjee, Matt Kelly, Elizabeth Shumbayawonda, Andrea Dennis, Charlotte Erpicum, Micheala Graham (Perspectum); Manuel Romero-Gómez, Emilio Gómez-González, Javier Ampuero, Javier Castell, Rocío Gallego-Durán, Isabel Fernández, Rocío Montero-Vallejo (Servicio Andaluz de Salud, Seville); Morten Karsdal, Elisabeth Erhardtsen, Daniel Rasmussen, Diana Julie Leeming, Mette Juul Fisker, Antonia Sinisi, Kishwar Musa (Nordic Bioscience); Fay Betsou, Estelle Sandt, Manuela Tonini (Integrated Biobank of Luxembourg); Elisabetta Bugianesi, Chiara Rosso, Angelo Armandi, Fabio Marra (UNIFI), Amalia Gastaldelli (CNR), Gianluca Svegliati (UNIPM) (University of Torino); Jérôme Boursier (University Hospital of Angers); Sven Francque; Luisa Vonghia (Antwerp University Hospital); Mattias Ekstedt, Stergios Kechagias (Linköping University); Hannele Yki-Jarvinen, Kimmu Porthan (University of Helsinki); Saskia van Mil (UMC Utrecht); George Papatheodoridis (National & Kapodistrian University of Athens); Helena Cortez-Pinto (Faculdade de Medicina de Lisboa); Luca Valenti (Università degli Studi di Milano); Salvatore Petta (Università degli Studi di Palermo); Luca Miele (Università Cattolica del Sacro Cuore); Andreas Geier (University Hospital Würzburg); Christian Trautwein (RWTH Aachen University Hospital); Guru Aithal (University of Nottingham); Paul Hockings (Antaros Medical); Philip Newsome (University Hospitals Birmingham NHS Foundation Trust); David Wenn (iXscient); Cecília Maria Pereira Rodrigues (University of Lisbon); Pierre Chaumat, Rémy Hanf (Genfit); Aldo Trylesinski (Intercept Pharma); Pablo Ortiz (OWL); Kevin Duffin (Ely-Lilly); Julia Brosnan, Theresa Tuthill, Euan McLeod (Pfizer); Judith Ertle, Ramy Younes (Boehringer-Ingelheim); Rachel Ostroff, Leigh Alexander (Somalogic); Mette Skalshøi Kjær (Novo Nordisk); Lars Friis Mikkelsen (Ellegaard Göttingen Minipigs); Maria-Magdalena Balp, Clifford Brass, Lori Jennings, Miljen Martic, Juergen Loeffler (Novartis Pharma AG); Guido Hanauer (Takeda Development Centre Europe Ltd); Sudha Shankar (AstraZeneca); Céline Fournier (Echosens); Kay Pepin, Richard Ehman (Resoundant); Joel Myers (Bristol-Myers Squibb); Gideon Ho (HistoIndex); Richard Torstenson (Allergan); Rob Myers (Gilead); Lynda Doward (RTI-HS).

Contributors FEM, EAS, ANAJ, MT, JB, AG, TT, JMB, QMA, SN, SAH, PMB, and MP contributed to the planning and design of the study. MT, JB, CF, KS, RES, EB, RY, SG, MLP, SP, TS, TO, SM, WKC, PJE, PNN, VWSW, VL, JGF, FS, JFC, YS, AO, JMS, CL, WK, MSL, JW, TK, YY, GPA, NP, CC, SA, HG, GO, AN, MY, MZ, and NB collected and provided individual patient data. FEM, JAL, PMB, and MP performed statistical analyses and data interpretation. FEM and MP wrote the first draft of the manuscript. All coauthors have approved the final version of the manuscript.

Funding This individual patient data meta-analysis is being conducted as part of the imaging study in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) study. The LITMUS study is a large multicentre study aiming to evaluate biomarkers on Non-Alcoholic Fatty Liver Disease. The LITMUS study is funded by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under Grant Agreement 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Competing interests AG reports personal fees from AbbVie, personal fees from Alexion, personal fees from Bayer, personal fees from BMS, personal fees from CSL Behring, personal fees from Gilead, grants and personal fees from Intercept, personal fees from Ipsen, personal fees from Merz, grants and personal fees from Novartis, personal fees from Pfizer, personal fees from Sanofi-Aventis, personal fees from Sequana, grants and personal fees from Falk, personal fees from MSD, during the conduct of the study. ANAJ reports other from Perspectum, during the conduct of the study. CF reports other from Echosens, during the conduct of the study. JFC reports personal fees from AstraZeneca, personal fees from Novo Nordisk, personal fees from Intercept, personal fees from Alnylam, during the conduct of the study. JB reports other from Pfizer, during the conduct of the study. JMS reports personal fees from BMS, personal fees from Boehringer Ingelheim, personal fees from Echosens, personal fees from Genfit, personal fees from Gilead Sciences, personal fees from Intercept Pharmaceuticals, personal fees from Madrigal, personal fees from Novartis, personal fees from Pfizer, personal fees from Roche, personal fees from Sanofi, personal fees from Falk Foundation, personal fees from MSD, grants from Gilead Sciences, during the conduct of the study. JW reports grants from Echosens, during the conduct of the study. Dr. Pavlides reports other from Perspectum, during the conduct of the study. MT reports personal fees from Bristol-Myers Squibb, personal fees from Falk Foundation, personal fees from Gilead, personal fees from Intercept, personal fees from Merck Sharp & Dohme, personal fees from Albireo, personal fees from Boehringer Ingelheim, personal fees from BiomX, personal fees from Falk Pharma GmbH, personal fees from GENFIT, personal fees from Jannsen, personal fees from Novartis, personal fees from Phenex, personal fees from Regulus and Shire, grants from AbbVie, grants from Falk, grants from Gilead, grants from Intercept, grants from Albireo, grants from CymaBay, grants from Merck Sharp & Dohme, grants from Takeda, during the conduct of the study; in addition, MT has a patent norUDCA issued. PNN reports personal fees from Bristol-Myers Squibb, personal fees from Gilead, personal fees from Boehringer Ingelheim, personal fees from Pfizer, personal fees from Novo Nordisk, personal fees from Poxel, grants from Pharmaxis, grants from Boehringer Ingelheim, grants from Echosens, grants from Novo Nordisk, during the conduct of the study. QMA reports grants from AbbVie, grants and personal fees from Allergan/Tobira, grants from AstraZeneca, grants from GlaxoSmithKline, grants from Glympse Bio, grants and personal fees from Novartis Pharma, grants and personal fees from Pfizer, grants from Vertex, personal fees from

Abbott Laboratories, personal fees from Acuitas Medical, personal fees from Blade, personal fees from BNN Cardio, personal fees from Cirius, personal fees from CymaBay, personal fees from EcoR1, personal fees from Eli Lilly, personal fees from Galmed, personal fees from Genfit, personal fees from Gilead, personal fees from Grunthal, personal fees from HistoIndex, personal fees from Indalo, personal fees from Imperial Innovations, personal fees from Intercept Pharma Europe, personal fees from Inventiva, personal fees from IQVIA, personal fees from Janssen, personal fees from Kenes, personal fees from Madrigal, personal fees from MedImmune, personal fees from Metacrine, personal fees from NewGene, personal fees from NGMBio, personal fees from North Sea Therapeutics, personal fees from Novo Nordisk, personal fees from Poxel, personal fees from ProSciento, personal fees from Raptor Pharma, personal fees from Servier, personal fees from Viking Therapeutics, personal fees from Abbott Laboratories, personal fees from BMS, personal fees from Clinical Care Options, personal fees from Falk, personal fees from Fishawack, personal fees from Integritas Communications, personal fees from MedScape, other from IMI2 LITMUS consortium, during the conduct of the study. SH reports grants and personal fees from Akero, grants and personal fees from Axcella, grants and personal fees from Cirius, grants and personal fees from CiVi Biopharma, grants and personal fees from CymaBay, grants and personal fees from Galectin, grants from Galmed, grants and personal fees from Genfit, grants and personal fees from Gilead Sciences, grants and personal fees from Hepion Pharmaceuticals, grants and personal fees from Hightide Therapeutics, grants and personal fees from Intercept, grants and personal fees from Madrigal, grants and personal fees from Metacrine, grants and personal fees from NGM Bio, grants and personal fees from North Sea Therapeutics, grants and personal fees from Novartis, grants and personal fees from Novo Nordisk, grants and personal fees from Poxel, grants and personal fees from Sagimet, grants and personal fees from Viking, personal fees from Altimmune, personal fees from Alentis, personal fees from Arrowhead, personal fees from Canfite, personal fees from Echosens, personal fees from Enyo, personal fees from Fibronostics, personal fees from Foresite Labs, personal fees from Fortress Biotech, personal fees from HistoIndex, personal fees from Kowa, personal fees from Prometic, personal fees from Ridgeline, personal fees from Terns, during the conduct of the study. SM reports personal fees from Echosens, during the conduct of the study. SN reports other from Perspectum, during the conduct of the study. SP reports personal fees from AbbVie, personal fees from Gilead, personal fees from Intercept, personal fees from Pfizer, during the conduct of the study. TK reports grants from Echosens, during the conduct of the study. TT reports other from Pfizer, during the conduct of the study. VdL reports personal fees from Bristol-Myers Squibb, personal fees from Gilead Sciences, personal fees from AbbVie, personal fees from Pfizer, personal fees from Echosens, personal fees from Intercept Pharmaceuticals, personal fees from MSD, personal fees from Myr-Pharma, personal fees from Supersonic Imagine, personal fees from Tillotts, during the conduct of the study. Dr. Wong reports personal fees from AbbVie, personal fees from 3V-BIO, personal fees from Allergan, personal fees from Boehringer Ingelheim, personal fees from Center for Outcomes Research in Liver Diseases, grants and personal fees from Gilead, personal fees from Intercept, personal fees from Echosens, personal fees from Hanmi Pharmaceutical, personal fees from Novartis, personal fees from Pfizer, personal fees from Merck, personal fees from Novo Nordisk, personal fees from Perspectum, personal fees from ProSciento, personal fees from Sagimet Biosciences, personal fees from TARGET PharmaSolutions, personal fees from Terns, personal fees from BMS, during the conduct of the study. WK reports personal fees from Samil, personal fees from Boehringer Ingelheim, personal fees from Ildong, personal fees from LG Chemistry, personal fees from Gilead Sciences, personal fees from HK inno.N, personal fees from GreenCross, personal fees from Bukwang, personal fees from Standigm, personal fees from PharmaKing, personal fees from KOBIOLABS, personal fees from Eisai, personal fees from Zydus, personal fees from Novo Nordisk, grants from Gilead, grants from Ildong, grants from GreenCross, grants from Bukwang, grants from PharmaKing, grants from Roche, grants from Galmed, grants from Novartis, grants from Pfizer, grants from Springbank, grants from Altimmune, grants from MSD, grants from BMS, grants from Dicerna, grants from Enyo, grants from Hitachi-Aloka, other from KOBIOLABS, other from Lepidyne, during the conduct of the study. YY reports grants from Biocodes, grants and personal fees from Gilead Sciences, personal fees from Bilim Pharmaceuticals, personal fees from Pharmactive Pharmaceutical, personal fees from Sanovel Pharmaceuticals, personal fees from Galmed, personal fees from Zydus, personal fees from Novo Nordisk, during the conduct of the study. MP, ANAJ and SN are shareholders of Perspectum, Oxford, UK. CF is employed by Echosens, France. MT received speaker fees from Bristol-Myers Squibb (BMS), Falk Foundation, Gilead, Intercept and Merck Sharp & Dohme (MSD); advisory board fees from Albireo, Boehringer Ingelheim, BiomX, Falk Pharma GmbH, GENFIT, Gilead, Intercept, Jannsen, MSD, Novartis, Phenex, Regulus and Shire; travel grants from AbbVie, Falk, Gilead, and Intercept; and research grants from Albireo, CymaBay, Falk, Gilead, Intercept, MSD, and Takeda. He is also coinventor of patents on the medical use of norUDCA filed by the Medical University of Graz. SP was speaker and/or Advisor for AbbVie, Gilead, Intercept and Pfizer. PNN received grant and research support from Pharmaxis, Boehringer Ingelheim, Echosens and Novo Nordisk and consulting fees from BMS, Boehringer Ingelheim, Gilead, Novo Nordisk, Pfizer, and Poxel on behalf of the University of Birmingham. VL reports consultancy for AbbVie, BMS, Echosens, Gilead Sciences, Intercept Pharmaceuticals, MSD, Myr-Pharma, Pfizer, Supersonic Imagine and Tillotts. SM received honorarium fees from Echosens. JFC received consultancy, advisory board, and speaker fees from Astra Zeneca, NovoNordisk, Intercept and Alnylam. JMS reports consultancy for BMS, Boehringer Ingelheim, Echosens, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Madrigal, Novartis, Pfizer, Roche, Sanofi; received research funding from Gilead Sciences and was on the speakers bureau for Falk Foundation MSD Sharp & Dohme GmbH. WK has served as a speaker and consultant of Gilead, Boehringer-Ingelheim, Samil, Ildong, LG Chemistry, HK inno.N, GreenCross, Bukwang, Standigm, PharmaKing, KOBIOLABS, Eisai, Zydus, and Novonordisk, received grants from Gilead, Ildong, GreenCross, Bukwang, Pharmaking, Roche, Galmed, Novartis, Pfizer, Springbank, Altimmune, MSD, BMS, Dicerna, Enyo, and Hitachi-Aloka, and owns stocks in KOBIOLABS and Lepidyne. TK and JW received unrestricted research grants from Echosens, Paris France. TK participated in a clinical advisory board meeting. YY received research grant from Biocodex, Gilead Sciences, speaker fees for Gilead Sciences, Bilim Pharmaceuticals, Pharmactive Pharmaceutical, Sanovel Pharmaceuticals, and served as advisory board member for Galmed, Zydus, NovoNordisk. AG reports consultancy for AbbVie, Alexion, Bayer, BMS, CSL Behring, Gilead, Intercept, Ipsen, Merz, Novartis, Pfizer, Sanofi-Aventis, Sequana; received research funding from Intercept, Falk, Novartis and was on the speakers bureau for AbbVie, Alexion, BMS, CSL Behring, Falk Foundation, Gilead, Intercept, MSD, Merz, Novartis, Sequana. VWSW has served as a consultant or advisory board member for 3V-BIO, AbbVie, Allergan, Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens, Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk, Perspectum Diagnostics, Pfizer, ProSciento, Sagimet Biosciences, TARGET PharmaSolutions, and Terns; and a speaker for AbbVie, Bristol-Myers Squibb, Echosens, and Gilead Sciences. He has also received a research grant from Gilead Sciences for fatty liver research. QMA is coordinator of the IMI2 LITMUS consortium and he reports research grant funding from Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex; consultancy on behalf of Newcastle University for Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Viking Therapeutics; and speaker fees from Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, MedScape. SAH has research grants from Akero, Axcella, Cirius, CiVi Biopharma, Cymabay, Galectin, Galmed, Genfit, Gilead Sciences, Hepion Pharmaceuticals, Hightide Therapeutics, Intercept, Madrigal, Metacrine, NGM Bio, Northsea Therapeutics, Novartis, Novo Nordisk, Poxel, Sagimet, Viking. He has received consulting fees from Akero, Altimmune, Alentis, Arrowhead, Axcella, Canfite, Cirius, CiVi, Cymabay, Echosens, Enyo, Fibronostics, Foresite Labs, Fortress Biotech, Galectin, Genfit, Gilead Sciences, Hepion, HIghtide, HistoIndex, Intercept, Kowa, Madrigal, Metacrine, NGM, Northsea, Novartis, Novo Nordisk, Poxel, Prometic, Ridgeline, Sagimet, Terns, and Viking.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Anonymised individual patient data are available upon reasonable request and with the agreement of the authors of original studies.

ORCID iDs
Ferenc Emil Mózes http://orcid.org/0000-0002-1361-4349
Michael Trauner http://orcid.org/0000-0002-1275-6425
Jerome Boursier http://orcid.org/0000-0002-7282-1436
Elisabetta Bugianesi http://orcid.org/0000-0002-0502-4381
Ramy Younes http://orcid.org/0000-0003-2302-5318
Salvatore Petta http://orcid.org/0000-0002-0822-9673
Sanjiv Mahadeva http://orcid.org/0000-0001-5824-0590
Philip Noel Newsome http://orcid.org/0000-0001-6085-3652
Vincent Wai-Sun Wong http://orcid.org/0000-0003-2215-9410

Christian Labenz http://orcid.org/0000-0001-8390-9663
Johannes Wiegand http://orcid.org/0000-0001-9233-4064
Thomas Karlas http://orcid.org/0000-0002-8109-8526
Yusuf Yılmaz http://orcid.org/0000-0003-4518-5283
Guruprasad Padur Aithal http://orcid.org/0000-0003-3924-4830
Christophe Cassinotto http://orcid.org/0000-0001-5136-4742
Geraldine J Ooi http://orcid.org/0000-0002-4540-408X
Atsushi Nakajima http://orcid.org/0000-0002-6263-1436
Andreas Geier http://orcid.org/0000-0002-9626-5083
Michael Pavlides http://orcid.org/0000-0001-9882-8874

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* 2018;15:11–20.
2 Loomba R, Wong R, Fraysse J, *et al*. Nonalcoholic fatty liver disease progression rates to cirrhosis and progression of cirrhosis to decompensation and mortality: a real world analysis of Medicare data. *Aliment Pharmacol Ther* 2020;51:1149–59.
3 European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. *Obes Facts* 2016;9:65–90.
4 Thampanitchawong P, Piratvisuth T. Liver biopsy:complications and risk factors. *World J Gastroenterol* 1999;5:301–4.
5 Ratziu V, Charlotte F, Heurtier A, *et al*. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005;128:1898–906.
6 Standish RA, Cholongitas E, Dhillon A, *et al*. An appraisal of the histopathological assessment of liver fibrosis. *Gut* 2006;55:569–78.
7 Davison BA, Harrison SA, Cotter G, *et al*. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol* 2020;73:1322–32.
8 Ekstedt M, Franzén LE, Mathiesen UL, *et al*. Long-Term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology* 2006;44:865–73.
9 Ekstedt M, Hagström H, Nasr P, *et al*. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015;61:1547–54.
10 Taylor RS, Taylor RJ, Bayliss S, *et al*. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology* 2020;158:e12:1611–25.
11 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;71:371–8.
12 Moolla A, Motohashi K, Marjot T, *et al*. A multidisciplinary approach to the management of NAFLD is associated with improvement in markers of liver and cardio-metabolic health. *Frontline Gastroenterol* 2019;10:337–46.
13 Davyduke T, Tandon P, Al-Karaghouli M, *et al*. Impact of Implementing a "FIB-4 First" Strategy on a Pathway for Patients With NAFLD Referred From Primary Care. *Hepatol Commun* 2019;3:1322–33.
14 European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO)Marchesini G, Day CP, Dufour JF. EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016;64:1388–402.
15 Stewart LA, Clarke M, Rovers M, *et al*. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;313:1657.
16 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.
17 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.
18 Sheth SG, Flamm SL, Gordon FD, *et al*. AST/ALT ratio predicts cirrhosis in patients with chronic hepatitis C virus infection. *Am J Gastroenterol* 1998;93:44–8.
19 Lin Z-H, Xin Y-N, Dong Q-J, *et al*. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology* 2011;53:726–36.
20 Wong VW-S, Vergniol J, Wong GL-H, *et al*. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* 2010;51:454–62.
21 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.
22 Boursier J, Zarski J-P, de Ledinghen V, *et al*. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013;57:1182–91.
23 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529.
24 Team RC. *R: a language and environment for statistical computing*, 2020.
25 Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
26 Canty A, Ripley B. *Boot: bootstrap R (S-Plus) functions. R package version 1.3-24*, 2019.
27 Davison AC, Hinkley D V. *Bootstrap methods and their application*. Cambridge University Press, 1997.
28 Petta S, Wong VW-S, Cammà 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;65:1145–55.
29 Anstee QM, Lawitz EJ, Alkhouri N, *et al*. Noninvasive tests accurately identify advanced fibrosis due to NASH: baseline data from the stellar trials. *Hepatology* 2019;70:1521–30.
30 Papatheodoridi M, Hiriart JB, Lupsor-Platon M, *et al*. Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease. *J Hepatol* 2021;74:1109–16.
31 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;156:1717–30.
32 Inadomi C, Takahashi H, Ogawa Y, *et al*. Accuracy of the enhanced liver fibrosis test, and combination of the enhanced liver fibrosis and non-invasive tests for the diagnosis of advanced liver fibrosis in patients with non-alcoholic fatty liver disease. *Hepatol Res* 2020;50:682–92.
33 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* 2019;17:156–63.
34 Hsu C, Caussy C, Imajo K, *et al*. Magnetic resonance vs transient elastography analysis of patients with nonalcoholic fatty liver disease: a systematic review and pooled analysis of individual participants. *Clin Gastroenterol Hepatol* 2019;17:630–7.
35 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;283:418–28.
36 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.
37 Boursier J, Guillaume M, Leroy V, *et al*. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. *J Hepatol* 2019;71:389–96.
38 Ioannou GN, Feld JJ. What are the benefits of a sustained virologic response to direct-acting antiviral therapy for hepatitis C virus infection? *Gastroenterology* 2019;156:446–60.
39 Petta S, Wai-Sun Wong V, Bugianesi E, *et al*. Impact of obesity and alanine aminotransferase levels on the diagnostic accuracy for advanced liver fibrosis of noninvasive tools in patients with nonalcoholic fatty liver disease. *Am J Gastroenterol* 2019;114:916–28.
40 Joo SK, Kim W, Kim D, *et al*. Steatosis severity affects the diagnostic performances of noninvasive fibrosis tests in nonalcoholic fatty liver disease. *Liver Int* 2018;38:331–41.
41 Alkayyali T, Qutranji L, Kaya E, *et al*. Clinical utility of noninvasive scores in assessing advanced hepatic fibrosis in patients with type 2 diabetes mellitus: a study in biopsy-proven non-alcoholic fatty liver disease. *Acta Diabetol* 2020;57:613–8.
42 Eren F, Kaya E, Yilmaz Y. Accuracy of Fibrosis-4 index and non-alcoholic fatty liver disease fibrosis scores in metabolic (dysfunction) associated fatty liver disease according to body mass index: failure in the prediction of advanced fibrosis in lean and morbidly obese individuals. *Eur J Gastroenterol Hepatol*.
43 McPherson S, Hardy T, Dufour J-F, *et al*. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *Am J Gastroenterol* 2017;112:740–51.
44 Ratziu V, Massard J, Charlotte F, *et al*. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2006;6:6.
45 Lichtinghagen R, Pietsch D, Bantel H, *et al*. The enhanced liver fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. *J Hepatol* 2013;59:236–42.
46 Newsome PN, Sasso M, Deeks JJ, *et al*. 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. *Lancet Gastroenterol Hepatol* 2020;5:362–73.
47 Harrison SA, Ratziu V, Boursier J, *et al*. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol Hepatol* 2020;5:970–85.
48 Dennis A, Mouchti S, Kelly M, *et al*. A composite biomarker using multiparametric magnetic resonance imaging and blood analytes accurately identifies patients with non-alcoholic steatohepatitis and significant fibrosis. *Sci Rep* 2020;10:1–11.
49 Agrawal S, Hoad CL, Francis ST, *et al*. Visual morphometry and three non-invasive markers in the evaluation of liver fibrosis in chronic liver disease. *Scand J Gastroenterol* 2017;52:107–15.
50 Aykut UE, Akyuz U, Yesil A, *et al*. A comparison of FibroMeter™ NAFLD score, NAFLD fibrosis score, and transient elastography as noninvasive diagnostic tools for hepatic fibrosis in patients with biopsy-proven non-alcoholic fatty liver disease. *Scand J Gastroenterol* 2014;49:1343–8.

51 Boursier J, Vergniol J, Guillet A, *et al*. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *J Hepatol* 2016;65:570–8.
52 Boursier J, Vergniol J, Lannes A, *et al*. The combination of Fibroscan with blood markers in the fibrometerVCTE significantly reduces the use of liver biopsy for the assessment of advanced fibrosis in non-alcoholic fatty liver disease. *J Hepatol* 2017;66:S161–2.
53 Boursier J, Lannes A, Shili S. The new fibrometervcte outperforms recommended liver fibrosis tests in NAFLD. *Hepatology* 2018;68.
54 Cassinotto C, Lapuyade B, Aït-Ali A, *et al*. Liver fibrosis: noninvasive assessment with acoustic radiation force impulse elastography--comparison with FibroScan M and XL probes and FibroTest in patients with chronic liver disease. *Radiology* 2013;269:283–92.
55 Cassinotto C, Boursier J, de Lédinghen V, *et al*. Liver stiffness in nonalcoholic fatty liver disease: a comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. *Hepatology* 2016;63:1817–27.
56 Chan W-K, Nik Mustapha NR, Mahadeva S. A novel 2-step approach combining the NAFLD fibrosis score and liver stiffness measurement for predicting advanced fibrosis. *Hepatol Int* 2015;9:594–602.
57 Chan W-K, Nik Mustapha NR, Wong GL-H, *et al*. Controlled attenuation parameter using the FibroScan® 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.
58 Eddowes PJ, Newsome PN, Anstee Q. Staging fibrosis and excluding advanced fibrosis in patients with NAFLD: comparison of non-invasive markers in an interim analysis from a prospective multicentre study. *Hepatology* 2016;64.
59 Eddowes PJ, McDonald N, Davies N, *et al*. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. *Aliment Pharmacol Ther* 2018;47:631–44.
60 Gaia S, Carenzi S, Barilli AL, *et al*. Reliability of transient elastography for the detection of fibrosis in non-alcoholic fatty liver disease and chronic viral hepatitis. *J Hepatol* 2011;54:64–71.
61 Garg H, Aggarwal S, Shalimar S. 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.
62 Karlas T, Dietrich A, Peter V, *et al*. Evaluation of transient elastography, acoustic radiation force impulse imaging (ARFI), and enhanced liver function (ELF) score for detection of fibrosis in morbidly obese patients. *PLoS One* 2015;10:e0141649.
63 Labenz C, Huber Y, Kalliga E, *et al*. Predictors of advanced fibrosis in non-cirrhotic non-alcoholic fatty liver disease in Germany. *Aliment Pharmacol Ther* 2018;48:1109–16.
64 Lee MS, Bae JM, Joo SK, *et al*. Prospective comparison among transient elastography, supersonic shear imaging, and ARFI imaging for predicting fibrosis in nonalcoholic fatty liver disease. *PLoS One* 2017;12:e0188321.
65 Lupsor M, Badea R, Stefanescu H, *et al*. Performance of unidimensional transient elastography in staging non-alcoholic steatohepatitis. *J Gastrointestin Liver Dis* 2010;19:53–60.
66 Mahadeva S, Mahfudz AS, Vijayanathan A. Performance of transient elastography (te) and factors associated with discordance in nonalcoholic fatty liver disease. *J Dig Dis* 2013;14.
67 Okajima A, Sumida Y, Taketani H, *et al*. Liver stiffness measurement to platelet ratio index predicts the stage of liver fibrosis in non-alcoholic fatty liver disease. *Hepatol Res* 2017;47:721–30.
68 Ooi GJ, Earnest A, Kemp WW, *et al*. Evaluating feasibility and accuracy of non-invasive tests for nonalcoholic fatty liver disease in severe and morbid obesity. *Int J Obes* 2018;42:1900–11.
69 Pavlides M, Banerjee R, Tunnicliffe EM, *et al*. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. *Liver Int* 2017;37:1065–73.
70 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.
71 Petta S, Vanni E, Bugianesi E, *et al*. The combination of liver stiffness measurement and NAFLD fibrosis score improves the noninvasive diagnostic accuracy for severe liver fibrosis in patients with nonalcoholic fatty liver disease. *Liver Int* 2015;35:1566–73.
72 Petta S, Wong VW-S, Cammà C, *et al*. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. *Aliment Pharmacol Ther* 2017;46:617–27.
73 Seki K, Shima T, Oya H, *et al*. Assessment of transient elastography in Japanese patients with non-alcoholic fatty liver disease. *Hepatol Res* 2017;47:882–9.
74 Shen F, Zheng R-D, Shi J-P, *et al*. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. *Liver Int* 2015;35:2392–400.
75 Staufer K, Halilbasic E, Spindelboeck W, *et al*. Evaluation and comparison of six noninvasive tests for prediction of significant or advanced fibrosis in nonalcoholic fatty liver disease. *United European Gastroenterol J* 2019;7:1113–23.
76 Wong VW-S, Irles M, Wong GL-H, *et al*. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. *Gut* 2019;68:2057–64.
77 Kwok R, Choi KC, Wong GL-H, *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.
78 Loong TC-W, Wei JL, Leung JC-F, *et al*. Application of the combined FibroMeter vibration-controlled transient elastography algorithm in Chinese patients with non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 2017;32:1363–9.
79 Wong VW-S, Vergniol J, Wong GL-H, *et al*. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. *Am J Gastroenterol* 2012;107:1862–71.
80 Yoneda M, Yoneda M, Mawatari H, *et al*. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). *Dig Liver Dis* 2008;40:371–8.
81 Younes R, Rosso C, Petta S, *et al*. Usefulness of the index of NASH - ION for the diagnosis of steatohepatitis in patients with non-alcoholic fatty liver: An external validation study. *Liver Int* 2018;38:715–23.
82 Ziol M, Kettaneh A, Ganne-Carrié N, *et al*. Relationships between fibrosis amounts assessed by morphometry and liver stiffness measurements in chronic hepatitis or steatohepatitis. *Eur J Gastroenterol Hepatol* 2009;21:1261–8.
83 Shah AG, Lydecker A, Murray K, *et al*. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009;7:1104–12.
84 McPherson S, Stewart SF, Henderson E, *et al*. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut* 2010;59:1265–9.