Supporting information for:

Diagnostic accuracy of non-invasive tests for diagnosing advanced fibrosis in patients with NAFLD – An individual patient data meta-analysis

Ferenc E. Mózes[1], Jenny Lee[2], Emmanuel A. Selvaraj[1,3,4], Arjun N. A. Jayaswal[1], Michael Trauner[5], Jérôme Boursier[6,7], Céline Fournier[8], Katharina Staufer[5,9,10], Rudolf 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], Philip N. Newsome[19,20,21], Vincent Wai-Sun Wong[22], Victor de Lédinghen[23], Jian-Gao Fan[24], Feng Shen[24], Jeremy F. L. Cobbold[25], Yoshio Sumida[26], Akira Okajima[27], Jörn M. Schattenberg[28], Christian Labenz[29], Won Kim[30], Myoung Seok Lee[31], Johannes Wiegand[32], Thomas Karlas[33], Yusuf Yilmaz[34,35], Guruprasad Padur Aithal[36,37], Naaventhan Palaniyappan[36,37], Christophe Cassinotto[38], Sandeep Aggarwal[39], Harshit Garg[39], Geraldine Ooi[40], Atsushi Nakajima[41], Masato Yoneda[41], Marianne Ziol[42], Nathalie Barget[43], Andreas Geier[44], Theresa Tuthill[45], Julia M. Brosnan[45], Quentin M. Anstee[46], Stefan Neubauer[1], Stephen A. Harrison[1], Patrick M. Bossuyt[2], Michael Pavlides[1,3,4], on behalf of the LITMUS Investigators

Contents

The LITMUS Investigators

Newcastle University	Quentin Anstee Ann Daly Katherine Johnson Olivier Govaere Simon Cockell Dina Tiniakos Pierre Bedossa Fiona Oakley Heather Cordell Chris Day Kristy Wonders
AMC Amsterdam	Patrick Bossuyt Hadi Zafarmand Yasaman Vali Jenny Lee
Hôpital Pitié Salpêtrière, Assistance Publique -Hôpitaux de Paris, and Institute of Cardiometabolism and Nutrition, Paris, France	Vlad Ratziu Karine Clement Raluca Pais
University Medical Center Mainz	Detlef Schuppan Jörn Schattenberg
University of Cambridge	Toni Vidal-Puig Michele Vacca Sergio Rodrigues-Cuenca Mike Allison Ioannis Kamzolas Evangelia Petsalaki
Örebro University	Matej Oresic Tuulia Hyötyläinen Aiden McGlinchey
Center for Cooperative Research in Biosciences	Jose M Mato Oscar Millet
University of Bern	Jean-François Dufour Annalisa Berzigotti
University of Oxford	Michael Pavlides Stephen Harrison Stefan Neubauer Jeremy Cobbold Ferenc Mozes Salma Akhtar
Perspectum	Rajarshi Banerjee Matt Kelly Elizabeth Shumbayawonda Andrea Dennis Charlotte Erpicum Micheala Graham
Servicio Andaluz de Salud, Seville	Manuel Romero-Gómez Emilio Gómez-González Javier Ampuero Javier Castell

	Rocío Gallego-Durán Isabel Fernández Rocío Montero-Vallejo
Nordic Bioscience	Morten Karsdal Elisabeth Erhardtsen Daniel Rasmussen Diana Julie Leeming Mette Juul Fisker Antonia Sinisi Kishwar Musa
Integrated Biobank of Luxembourg	Fay Betsou Estelle Sandt Manuela Tonini
University of Torino	Elisabetta Bugianesi Chiara Rosso Angelo Armandi Fabio Marra (UNIFI) Amalia Gastaldelli (CNR) Gianluca Svegliati (UNIPM)
University Hospital of Angers	Jérôme Boursier
Antwerp University Hospital	Sven Francque Luisa Vonghia
Linköping University	Mattias Ekstedt Stergios Kechagias
University of Helsinki	Hannele Yki-Jarvinen Kimmu Porthan
UMC Utrecht	Saskia van Mil
National & Kapodistrian University of Athens	George Papatheodoridis
Faculdade de Medicina de Lisboa	Helena Cortez-Pinto
Università degli Studi di Milano	Luca Valenti
Università degli Studi di Palermo	Salvatore Petta
Università Cattolica del Sacro Cuore	Luca Miele
University Hospital Würzburg	Andreas Geier
RWTH Aachen University Hospital	Christian Trautwein
University of Nottingham	Guru Aithal
Antaros Medical	Paul Hockings
University Hospitals Birmingham NHS Foundation Trust	Philip Newsome
iXscient	David Wenn
University of Lisbon	Cecília Maria Pereira Rodrigues
Genfit	Pierre Chaumat Rémy Hanf
Intercept Pharma	Aldo Trylesinski
OWL	Pablo Ortiz
Ely-Lilly	Kevin Duffin
Pfizer	Julia Brosnan Theresa Tuthill Euan McLeod
Boehringer-Ingelheim	Judith Ertle Ramy Younes
Somalogic	Rachel Ostroff

	Leigh Alexander
Novo Nordisk	Mette Skalshøi Kjær
Ellegaard Göttingen Minipigs	Lars Friis Mikkelsen
Novartis Pharma AG	Maria-Magdalena Balp Clifford Brass Lori Jennings Miljen Martic Juergen Loeffler
Takeda Development Centre Europe Ltd	Guido Hanauer
AstraZeneca	Sudha Shankar
Echosens	Céline Fournier
Resoundant	Kay Pepin Richard Ehman
Bristol-Myers Squibb	Joel Myers
HistoIndex	Gideon Ho
Allergan	Richard Torstenson
Gilead	Rob Myers
RTI-HS	Lynda Doward

Supporting Methods

According to the manufacturer, probe selection should be driven by skin-to-liver capsule distance (SCD): M probe for SCD < 25 mm and XL probe for 25 mm ≤ SCD < 35 mm. In the latest version of the FibroScan equipment this is done by the Automatic Probe Selection tool. Some investigators have suggested that BMI may be used as a surrogate of SCD, using the M probe if BMI < 30 kg/m^2 and XL probe if BMI ≥ 30 kg/m^2 (28).

For this meta-analysis, if only one VCTE-based liver stiffness measurement was available then this was included in the main analysis irrespective of probe type and BMI. Where two VCTE-based LSM were available (one with each probe), the main analysis included the M-probe measurement for BMI < 30 kg/m^2 and the XL probe measurement for BMI ≥ 30 kg/m^2. Therefore, all LSM cut-offs were determined independent of probe type.

We further conducted sensitivity analysis to investigate the influence of probe selection by excluding patients with BMI ≥ 30 kg/m^2 who had a measurement with the M probe and patients with BMI < 30 kg/m^2 who had measurement with the XL probe.

Supporting Discussion

Rationale for proposing new NIT combinations with higher cut-offs for diagnosis of cirrhosis

Up until now, the literature has focused on the application of non-invasive tests in screening strategies for advanced fibrosis (F3-4). These strategies are useful when applied at the interface of primary and secondary care. Patients assessed using these strategies are classified as low risk, high risk or indeterminate risk of having advanced fibrosis, based on which clinical decisions are made: those with low risk continue to be managed in primary care, those with high risk are referred to secondary care and those with indeterminate risk undergo liver biopsy to determine their risk category.

What is lacking from the literature and what we have tried to answer with our analysis is what happens to patients with high risk of advanced fibrosis that are referred to secondary care. Our view is that they remain an indeterminate group as they can have either F3 or F4 fibrosis stage. Therefore, to distinguish between F3 and cirrhosis (F4) they still need to undergo liver biopsy, as those with liver cirrhosis would be managed differently (ultrasound surveillance for HCC and screening for oesophageal varices is generally indicated in patients with cirrhosis, but not those with F3 fibrosis stage). The identification of patients with cirrhosis would also be important as potential treatments for NASH may be licenced exclusively for patients with or without cirrhosis. We therefore argue that in practice, both the indeterminate and high-risk groups need to have a liver biopsy to establish their disease stage. In the case of those in the indeterminate category, the biopsy is needed to decide whether they merit referral to secondary care, and in the case of those with high risk of advanced fibrosis a biopsy is needed in secondary care to identify those with cirrhosis. We illustrate this point in **Supporting Figure 1a** and in **Figure 3a**, we also show how the FIB4-VCTE combination performs in our cohort.

Our answer to the problem above is a hybrid algorithm, where the lower NIT cut-offs are used to rule out advanced fibrosis, and the upper cut-offs are used to rule in cirrhosis. We provide cut-offs

with 95% and 98% specificity for the diagnosis of cirrhosis. This approach still stratifies patents into 3 risk groups – those with low risk of advanced fibrosis remaining in primary care, those in the indeterminate group needing a biopsy and those with high risk for cirrhosis. We argue that the group with high risk for cirrhosis can be positively diagnosed with cirrhosis without needing to have a biopsy. The net effect is that even though the indeterminate group is larger, fewer patients need to have a biopsy overall. This new approach is illustrated in **Supporting Figure 1b**, with results from our cohort given in **Figures 3b** and **3c**.

Supporting Tables

Supporting Table 1 Definitions of NITs evaluated in the current meta-analysis.

NIT	Definition
LSM by VCTE	An ultrasound probe that can also generate shear waves is placed over the right liver lobe. A low frequency shear wave is then generated by the external vibrator located in the probe, and ultrasound is used to measure the velocity of this shear wave through the liver. This velocity is directly related to liver stiffness.
FIB-4	Age [years] × AST [IU/L] / (platelets [× 10^9/L] x (ALT [U/L])$^{1/2}$)
NFS	-1.675 + 0.037 × age [years] + 0.094 × BMI [kg/m^2] + 1.13 × IFG/diabetes [yes = 1, no = 0] + 0.99 × AST/ALT ratio – 0.013 × platelet [× 10^9/L] – 0.66 × albumin [g/dL]
AST/ALT	AST [IU/L] / ALT [IU/L]
APRI	AST [IU/L] / AST ULN [IU/L] / platelet [× 10^9/L]

Abbreviations: LSM – liver stiffness measurement; VCTE – vibration-controlled transient elastography; FIB-4 – Fibrosis-4 score; NFS – NAFLD fibrosis score; AST/ALT – AST to ALT ratio; APRI – AST to platelet ratio index; ULN – upper limit of normal; IU – international unit; IFG – impaired fasting glucose

Supporting Table 2 Non-invasive test cut-offs to rule-in and rule-out advanced fibrosis in patients with NAFLD

Study ID	Rule out cut-off	Rule-in cut-off
Vibration controlled transient elastography		
Studies testing pre-defined cut-offs (kPa)		
Anstee 2019 (1)	< 9.0	> 11.4
Wong 2019 (2), Papatheodoridi 2021 (ref)	< 10.0	> 15.0^
Petta 2019 (3), Boursier 2019 (4), Petta 2017 (5)	< 7.9	> 9.6*
Cut-offs identified from other primary studies (kPa)		
Tapper 2016 (6)	< 7.9	> 9.8
Eddowes 2019 (7)	< 7.1	> 14.1
Hsu 2019 (8)	< 5.9	> 13.4
Cassinotto 2016 (9)	< 8.2	> 12.5
Papatheodoridi 2021 (ref)		
FIB-4		
Studies testing pre-defined cut-offs		
Anstee 2019 (1), Xun 2012 (10), Petta 2019 (3)	< 1.30	> 2.67[#]
Vilar-Gomez 2018 (11), Sun 2016 (12), McPherson 2010 (13), Srivastava 2019 (14)	< 1.30	> 3.25
Demir 2013 (15)	< 1.45	> 3.25
Cut-offs from other primary studies		
Siddiqui 2019 (16)	< 1.02	> 1.95
NAFLD Fibrosis score		
Studies testing pre-defined cut-offs		
Antsee 2019 (1), Tapper 2016 (6), Vilar-Gomez 2018 (11), Sun 2016 (12), McPherson 2010 (13), Xun 2012 (10), Demir 2013 (15), Petta 2014 (17), Dowman 2011 (18), Petta 2019 (3), Fowell 2020 (19)	< -1.455	> 0.676[%]

^based on BavenoVI (20), *based on Wong (21), [#]from Shah 2009 (22), [%]from Angulo 2007 (23)

Supporting Table 3 Data fields requested from the authors of primary studies of LSM by VCTE

Category	Field	Units or possible values	Proportion of patients in whom reported, %
Study details	Name of first author	-	100.0
	Year of publication	-	100.0
	Country	-	100.0
	Centre	-	
Demographic and anthropometric details	Gender	M/F	100.0
	Age	years	99.9
	Ethnicity	-	38.6
	Height	m	92.4
	Weight	kg	94.9
	Waist circumference	cm	72.3
	Hip circumference	cm	21.8
	Smoking	Current/Ex/Never	10.0
	Presence of type 2 diabetes mellitus	Yes/No	86.4
	Presence of hypertension	Yes/No	48.8
	Presence of hyperlipidaemia	Yes/No	26.0
Laboratory data	Platelet count	$\times 10^9$/l	98.2
	INR	-	35.4
	Bilirubin	μmol/l	55.5
	ALT	IU/L	97.2
	AST	IU/L	96.2
	ALP	IU/L	48.3
	GGT	IU/L	82.2
	Albumin	g/l	67.2
	Sodium	mmol/l	6.7
	Urea	mmol/l	13.7
	Creatine	μmol/l	22.2
	Total cholesterol	mmol/l	62.8
	LDL cholesterol	mmol/l	32.8
	HDL cholesterol	mmol/l	77.6
	Triglycerides	mmol/l	79.3
	CRP	mg/l	7.9
	Fasting glucose	mmol/l	73.0
	Fasting insulin	mU/L	18.0
	HOMA-IR	-	16.8
Biopsy data	Date of biopsy	-	67.0
	Length of biopsy sample	mm	70.6
	Number of portal tracts	-	32.4
	Fibrosis stage	0-4	100.0
	Ballooning	0-2	63.7
	Lobular inflammation	0-3	64.2
	Steatosis	0-3	71.5
	NAS score	0-8	82.9
	Date of scan	-	68.9

	Time between biopsy and scan	days	79.3
	Probe type	M/XL	91.9
Transient elastography details	Number of valid shots	-	59.4
	Median stiffness	kPa	95.7
	IQR	kPa	83.4
	IQR/median	-	83.0
	Success rate	%	77.8

Supporting Table 4 Demographic, biopsy, liver function test and NIT details of the entire cohort and broken down by fibrosis stage

	Entire cohort (n = 5735)	**F0 (n = 1138)**	**F1 (n = 1613)**	**F2 (n = 1262)**	**F3 (n = 1101)**	**F4 (n = 621)**
Females (%)	45	43	44	43	47	50
BMI > 30 kg/m^2 (%)	47	33	45	56	55	51
Waist circumference (cm)	103 (15)	99 (16)	101 (15)	106 (14)	106 (14)	106 (15)
Diabetes (%)	38	28	33	45	62	65
Age (years)*	54 (19)	48 (17)	50 (20)	53 (19)	59 (15)	60 (12)
BMI (kg/m^2)*	30 (7)	28 (7)	29 (7)	31 (7)	31 (7)	30 (7)
Biopsy data						
Steatosis S0/S1/S2/S3 (%)	3/35/36/26	8/45/30/17	2/35/37/26	1/28/39/32	1/28/39/32	3/38/37/22
Ballooning B0/B1/B2 (%)	24/47/29	53/37/10	26/55/19	11/53/36	10/43/47	10/46/44
Inflammation I0/I1/I2/I3 (%)	13/60/24/3	3/60/9/4	13/65/21/1	6/60/31/3	5/53/36/6	8/57/29/6
NAS score+	4 (2)	3 (2)	4 (2)	4 (1)	5 (1)	4 (2)
NASH (%)	50	19	46	64	71	61
Liver function tests						
ALT (IU/L) *	55 (48)	46 (39)	54 (50)	59 (52)	63 (50)	55 (43)
AST (IU/L) *	40 (30)	31 (19)	36 (27)	41 (28)	49 (32)	53 (39)
Platelets (×10^9/l) +	230 (72)	247 (64)	243 (69)	232 (66)	217 (69)	184 (81)
Albumin (g/l) +	43 (9)	43 (8)	43 (7)	43 (5)	43 (6)	43 (20)
GGT (IU/L) *	69 (87)	59 (85)	61 (75)	63 (74)	82 (88)	104 (169)
Total cholesterol (mmol/l) +	5.1 (1.3)	5.2 (1.3)	5.1 (1.2)	5.2 (1.4)	4.9 (1.2)	4.6 (1.3)
HDL cholesterol (mmol/l) +	1.2 (0.5)	1.3 (0.5)	1.3 (0.5)	1.2 (0.5)	1.2 (0.4)	1.2 (0.5)
Triglycerides (mmol/l)*	1.6 (1.1)	1.4 (1.0)	1.6 (1.1)	1.6 (1.0)	1.6 (1.1)	1.5 (1.0)

Fasting glucose (mmol/l) *	5.6 (2.0)	5.3 (1.2)	5.4 (1.8)	5.6 (1.7)	6.3 (2.9)	6.4 (2.8)
Non-invasive tests						
LSM (kPa) *	10.7 (6.1)	5.7 (2.5)	6.7 (3.4)	7.9 (4.3)	11.3 (6.9)	20.9 (16.8)
AST/ALT*	0.8 (0.4)	0.7 (0.4)	0.7 (0.4)	0.7 (0.4)	0.8 (0.4)	0.9 (0.6)
FIB-4+	1.7 (1.2)	1.1 (0.7)	1.3 (1.2)	1.5 (1.1)	2.1 (1.6)	3.3 (2.9)
NFS*	-1.5 (1.7)	-2.3 (2.0)	-2.0 (2.2)	-1.4 (2.2)	-0.8 (1.8)	0.0 (1.8)
APRI*	0.6 (0.4)	0.3 (0.3)	0.4 (0.3)	0.5 (0.4)	0.6 (0.5)	0.8 (0.8)

*Data are reported as median (IQR); +Data are reported as mean (SD).

Supporting Table 5 Details of biopsy and biopsy quality in the entire IPD cohort.

Biopsy details	Entire cohort (n = 5735)	Advanced fibrosis (n = 1722)	Cirrhosis (n = 621)
Time between liver biopsy and LSM by VCTE			
Patients with reported exact time period, %	79 (4549/5735)	80 (1371/1722)	76 (474/621)
Median (IQR) (days)	0 (14)	0 (9)	1 (26)
Length of biopsy sample			
Patients with reported length of biopsy, %	71 (4047/5735)	80 (1369/1722)	80 (495/621)
< 10 mm, %	3 (123/4047)	3 (42/1369)	5 (25/495)
≥ 10 mm and < 20 mm, %	35 (1432/4047)	33 (450/1369)	35 (172/495)
≥ 20 mm, %	62 (2492/4047)	64 (877/1369)	60 (298/495)
Number of portal tracts in biopsy sample			
Patients with reported portal tracts %	32 (1857/5735)	32 (544/1722)	26 (159/621)
< 11, %	54 (1006/1857)	42 (228/544)	47 (74/159)
≥ 11, %	46 (851/1857)	58 (316/544)	54 (85/159)
Patients with both portal tracts and biopsy length reported, %	32 (1854/5735)	32 (543/1722)	26 (159/621)
Biopsy quality			
Intermediate quality (length ≥ 10 mm and < 20 mm), %	46 (849/1854)	41 (220/543)	39 (62/159)
High quality (length ≥ 20 mm and ≥ 11 portal tracts), %	36 (670/1854)	45 (246/543)	39 (62/159)

Data are reported as percentage (number of patient satisfying conditions/total number of patients in subgroup)

Supporting Table 6 Diagnostic performance of non-invasive tests for cirrhosis (F4)

	LSM by VCTE (n = 5489)			**FIB-4 (n = 5393)**			**NFS (n = 3248)**			**APRI (n =5477)**			**AST/ALT ratio (n = 5434)**		
Cirrhosis, %	11			11			11			11			11		
AUC	0.90 (.89-0.91)			0.80 (0.78-0.82)			0.77 (0.75-0.80)			0.72 (0.70-0.74)			0.69 (0.67-0.71)		
Threshold	10.4	<10.2	≥14.9	1.55	<1.13	≥2.66	-1.11	<-1.72	≥0.48	0.58	<0.30	≥1.04	0.82	<0.58	≥1.35
Sensitivity, %	89 (86-91)	90 (8-92)	67 (64-70)	77 (72-80)	90 (87-92)	44 (40-48)	82 (76-85)	90 (86-93)	36 (31-40)	66 (61-69)	90 (87-92)	35 (31-39)	64 (59-67)	90 (87-92)	24 (20-28)
Specificity, %	75 (74-76)	74 (72-75)	90 (89-90)	67 (65-68)	48 (46-49)	90 (89-90)	63 (61-64)	49 (46-50)	90 (88-91)	68 (66-69)	28 (26-28)	90 (89-90)	66 (64-67)	33 (31-33)	90 (89-90)
Misclassified, %	23 (23-24)	24 (24-25)	12 (12-13)	32 (32-33)	48 (47-49)	15 (14-15)	35 (35-36)	47 (46-48)	16 (15-16)	32 (32-33)	66 (65-66)	16 (15-16)	34 (34-35)	61 (61-62)	17 (16-17)

For each non-invasive test thresholds were calculated according to Youden's index (YI), and fixed at 90% sensitivity (90% Se) and 90% specificity (90% Sp). 95% confidence intervals were estimated with 500 bootstrap iterations.

Supporting Table 7 Diagnostic performance of non-invasive tests for advanced fibrosis (F3-4) in a head-to-head comparison of NITs

	LSM by VCTE (n = 3248)			**FIB-4 (n = 3248)**			**NFS (n = 3248)**		
Advanced fibrosis, %	29			29			29		
AUC	0.86 (0.85-0.88)			0.75 (0.73-0.77)			0.73 (0.71-0.75)		
	YI	90% Se	90% Sp	YI	90% Se	90% Sp	YI	90% Se	90% Sp
Threshold	9.1	7.2	11.8	1.45	0.87	2.39	-1.39	-2.55	0.28
Sensitivity, %	77 (74-80)	90 (89-92)	59 (57-63)	69 (66-72)	90 (88-92)	36 (33-39)	75 (72-78)	90 (88-92)	29 (26-32)
Specificity, %	81 (79-82)	61 (59-63)	90 (89-92)	69 (67-71)	38 (36-39)	90 (89-91)	63 (61-65)	36 (33-37)	90 (89-91)
Misclassified, %	21 (19-22)	31 (29-32)	18 (17-20)	31 (29-32)	47 (46-49)	25 (24-27)	34 (34-36)	48 (49-50)	28 (28-29)

Supporting Table 8 Diagnostic performance of non-invasive tests for cirrhosis (F4) in a head-to-head comparison of NITs

	LSM by VCTE (n = 3094)			**FIB-4 (n = 3094)**			**NFS (n = 3094)**		
Cirrhosis, %	11			11			11		
AUC	0.91 (0.89-0.92)			0.78 (0.76-0.81)			0.77 (0.75-0.80)		
	YI	90% Se	90% Sp	YI	90% Se	90% Sp	YI	90% Se	90% Sp
Threshold	10.3	9.7	14.4	1.35	1.08	2.76	-1.11	-1.93	0.46
Sensitivity, %	89 (86-92)	90 (87-93)	68 (63-72)	83 (79-87)	90 (87-93)	42 (37-47)	81 (77-86)	90 (87-93)	35 (29-40)
Specificity, %	78 (76-79)	74 (73-76)	91 (90-92)	59 (57-61)	45 (43-47)	90 (89-91)	64 (62-66)	45 (43-47)	90 (89-91)
Misclassified, %	21 (20-22)	24 (22-25)	12 (10-13)	39 (37-40)	50 (48-52)	15 (14-16)	34 (33-36)	50 (49-52)	16 (15-17)

Supporting Table 9 Diagnostic performance of cut-offs from the literature for LSM by VCTE, FIB-4 and NFS for diagnosing advanced fibrosis.

	LSM by VCTE (n = 5489)						**FIB-4 (n = 5393)**				**NFS (n = 3248)**			
Source	Anstee 2019 (30)		Eddowes 2019 (31)		Wong 2019 (71)		Wong 2010 (21)		Shah 2009 (78)		McPherson 2010 (79)		Angulo 2007 (17)	
Thresholds	<9.9	≥11.4	<7.1	≥14.1	<10	≥15	<7.9	≥9.6	<1.3	≥2.67	<1.3	≥3.25	<-1.455	≥0.676
Sensitivity, %	72 (71-75)	61 (60-64)	91 (90-93)	46 (44-49)	71 (70-74)	41 (39-44)	86 (86-89)	73 (71-76)	74 (72-76)	30 (28-32)	74 (72-76)	20 (18-22)	76 (73-78)	22 (19-24)
Specificity, %	82 (80-83)	87 (86-88)	58 (55-58)	94 (93-94)	82 (81-83)	95 (94-96)	68 (65-68)	81 (79-81)	64 (63-66)	94 (93-94)	64 (63-66)	96 (96-97)	61 (60-64)	94 (93-95)
Misclassified, %	21 (21-22)	21 (20-22)	32 (32-34)	20 (20-21)	21 (21-22)	21 (21-22)	27 (27-29)	21 (21-23)	33 (33-34)	25 (25-26)	33 (33-34)	27 (26-27)	35 (34-36)	28 (27-28)

95% confidence intervals were estimated with 500 bootstrap iterations

Supporting Table 10 Indicative PPV, NPV, false positive and false negative proportions for prevalences found in primary studies for the diagnostic performance of VCTE in identifying advanced fibrosis

Cut-off	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
7.4 kPa	90	89-91	60	59-61	5	11	99	38	1
					10	20	98	36	1
					20	36	96	32	2
					30	**49**	**93**	**28**	**3**
					40	60	90	24	4
					50	69	86	20	5
9.1 kPa	77	75-79	78	76-79	5	16	98	21	1
					10	28	97	20	2
					20	47	93	18	5
					30	**60**	**89**	**15**	**7**
					40	70	84	13	9
					50	78	77	11	12
12.1 kPa	55	52-57	90	89-91	5	22	97	10	2
					10	38	95	9	5
					20	58	89	8	9
					30	**70**	**82**	**7**	**14**
					40	79	75	6	18
					50	85	67	5	23
<7.4 kPa, ≥12.1 kPa	84	81-87	87	85-88	5	25	99	12	1
					10	42	98	12	2
					20	62	96	10	3
					30	**73**	**93**	**9**	**5**
					40	81	89	8	6
					50	87	84	7	8
<9.9 kPa, ≥11.4 kPa (Anstee 2019)	69	67-71	86	85-88	5	21	98	13	2
					10	35	96	13	3
					20	55	92	11	6
					30	**68**	**87**	**10**	**9**
					40	77	81	8	12
					50	83	74	7	16
<7.1, ≥14.1 (Eddowes 2019)	83	80-86	90	88-92	5	30	99	10	1
					10	48	98	9	2
					20	67	95	8	3
					30	**78**	**93**	**7**	**5**
					40	85	89	6	7
					50	89	84	5	9
<10, ≥15 (Wong 2019)	59	57-61	94	93-96	5	34	98	6	2
					10	52	95	5	4
					20	71	90	5	8
					30	**81**	**84**	**4**	**12**
					40	87	77	4	16
					50	91	70	3	21
<7.9, ≥9.6 (Wong 2010)	84	82-87	78	76-80	5	17	99	21	1
					10	30	98	20	2
					20	49	95	18	3
					30	**62**	**92**	**15**	**5**
					40	72	88	13	6

	50	79	83	11	8

*Number of false positives and false negatives for 100 hypothetical cases; Values in bold face correspond to the proportion of patients with advanced fibrosis in the IPDMA cohort

Supporting Table 11 Indicative PPV, NPV, false positive and false negative proportions for prevalences found in primary studies for the diagnostic performance of FIB-4 in identifying advanced fibrosis

Cut-off	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
0.88	90	88-91	39	37-40	5	7	99	58	1
					10	14	97	55	1
					20	27	94	49	2
					30	**39**	**90**	**43**	**3**
					40	50	85	37	4
					50	60	80	31	5
1.44	69	67-72	70	69-72	5	11	98	29	2
					10	20	95	27	3
					20	37	90	24	6
					30	**50**	**84**	**21**	**9**
					40	61	77	18	12
					50	70	69	15	16
2.31	38	36-41	90	89-91	5	17	97	10	3
					10	30	93	9	6
					20	49	85	8	12
					30	**62**	**77**	**7**	**19**
					40	72	69	6	25
					50	79	59	5	31
<1.3, ≥2.67 (Shah 2009)	54	52-56	91	89-92	5	24	97	9	2
					10	40	95	8	5
					20	60	89	7	9
					30	**72**	**82**	**6**	**14**
					40	80	75	5	18
					50	86	66	5	23
<1.3, ≥3.25 (McPherson 2010)	44	42-46	95	93-96	5	32	97	5	3
					10	49	94	5	6
					20	69	87	4	11
					30	**79**	**80**	**4**	**17**
					40	85	72	3	22
					50	90	63	3	28

*Number of false positives and false negatives for 100 hypothetical cases; Values in bold face correspond to the proportion of patients with advanced fibrosis in the IPDMA cohort

Supporting Table 12 Indicative PPV, NPV, false positive and false negative proportions for prevalences found in primary studies for the diagnostic performance of NFS in identifying advanced fibrosis

Cut-off	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
-2.55	90	88-92	36	33-37	5	7	99	61	1
					10	14	97	58	1
					20	26	94	51	2
					30	**38**	**89**	**45**	**3**
					40	48	84	38	4
					50	58	78	32	5
-1.39	75	72-78	63	61-65	5	10	98	35	1
					10	18	96	33	3
					20	34	91	30	5
					30	**46**	**85**	**26**	**8**
					40	57	79	22	10
					50	67	72	19	13
0.28	29	26-32	90	89-91	5	13	96	10	4
					10	24	92	9	7
					20	42	84	8	14
					30	**55**	**75**	**7**	**21**
					40	66	66	6	28
					50	74	56	5	36
<-1.455, ≥0.676 (Angulo 2007)	47	44-50	91	89-93	5	22	97	9	3
					10	37	94	8	5
					20	57	87	7	11
					30	**69**	**80**	**6**	**16**
					40	78	72	5	21
					50	84	63	5	27

*Number of false positives and false negatives for 100 hypothetical cases; Values in bold face correspond to the proportion of patients with advanced fibrosis in the IPDMA cohort

Supporting Table 13 Indicative PPV, NPV, false positive and false negative proportions for prevalences found in primary studies for the diagnostic performance of APRI in identifying advanced fibrosis

Cut-off	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
0.29	90	89-92	29	28-30	5	6	98	67	1
					10	12	96	64	1
					20	24	92	57	2
					30	**35**	**87**	**50**	**3**
					40	46	81	43	4
					50	56	74	36	5
0.49	67	64-69	63	62-65	5	9	97	35	2
					10	17	95	33	3
					20	31	88	30	7
					30	**44**	**82**	**26**	**10**
					40	55	74	22	13
					50	64	66	19	17
0.91	32	30-34	90	89-91	5	14	96	10	3
					10	26	92	9	7
					20	44	84	8	14
					30	**58**	**76**	**7**	**20**
					40	68	67	6	27
					50	76	57	5	34

*Number of false positives and false negatives for 100 hypothetical cases; Values in bold face correspond to the proportion of patients with advanced fibrosis in the IPDMA cohort

Supporting Table 14 Indicative PPV, NPV, false positive and false negative proportions for prevalences found in primary studies for the diagnostic performance of AST/ALT in identifying advanced fibrosis

Cut-off	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
0.51	90	87-91	25	23-26	5	6	98	71	1
					10	12	96	68	1
					20	23	91	60	2
					30	**34**	**85**	**53**	**3**
					40	44	79	45	4
					50	55	71	38	5
0.64	75	73-77	47	45-48	5	7	97	50	1
					10	14	94	48	3
					20	26	88	42	5
					30	**38**	**81**	**37**	**8**
					40	49	74	32	10
					50	59	65	27	13
1.34	16	14-18	90	89-91	5	8	95	10	4
					10	15	91	9	8
					20	29	81	8	17
					30	**41**	**71**	**7**	**25**
					40	52	62	6	34
					50	62	52	5	42

*Number of false positives and false negatives for 100 hypothetical cases; Values in bold face correspond to the proportion of patients with advanced fibrosis in the IPDMA cohort

Supporting Table 15 Diagnostic accuracy of pairs of cut-offs from the literature for NITs for diagnosing advanced fibrosis. Patient proportions used to calculate performance statistics are displayed as ratios.

	LSM by VCTE (n = 5489)					FIB-4 (n = 5393)			NFS (n = 3248)	
Prevalence, %	30					30			29	
AUROC	0.85 (0.84-0.86)					0.76 (0.74-0.77)			0.73 (0.71-0.75)	
Source of thresholds	Anstee 2019 (1)	Eddowes 2019 (7)	Wong 2019 (2)	Wong 2010 (21)	This study	Shah 2009 (22)	McPherson 2010 (13)	This study	Angulo 2007 (23)	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 (1009/1456)	83 (754/905)	59 (674/1145)	84 (1205/1431)	84 (889/1060)	54 (485/901)	44 (328/744)	80 (621/780)	47 (202/429)	74 (270/363)
Specificity, %	86 (3147/3639)	90 (2216/2457)	94 (3165/3351)	78 (2599/3330)	87 (2338/2702)	91 (2423/2668)	95 (2423/2563)	79 (1448/1831)	91 (1423/1562)	78 (821/1050)
Misclassified, %	17 (948/5489)	7 (392/5489)	12 (657/5489)	17 (957/5489)	10 (535/5489)	12 (661/5393)	10 (556/5393)	10 (542/5393)	11 (366/3248)	10 (322/3248)
Indeterminate, %	7 (385/5489)	39 (2127/5489)	18 (993/5489)	13 (728/5489)	31 (1727/5489)	34 (1824/5393)	39 (2086/5393)	52 (2782/5393)	39 (1257/3248)	56 (1835/3248)

95% confidence intervals were estimated with 500 bootstrap replicates.

Supporting Table 16 Derivation of new cut-offs corresponding to 95% specificity for cirrhosis. The cut-offs were calculated in the training cohort and were validated in the validation cohort. Training and validation cohorts were obtained by random sampling from the IPD cohort in a 3:2 ratio.

	LSM by VCTE (n = 5489)		**FIB-4 (n = 5393)**		**NFS (n = 3248)**	
	Training (n = 3290)	Validation (n = 2199)	Training (n = 3254)	Validation (n = 2139)	Training (n = 1963)	Validation (n = 1285)
Proportion of patients with cirrhosis, %	11	10	11	11	10	11
AUC	0.90 (0.89-0.92)	0.90 (0.88-0.92)	0.81 (0.78-0.83)	0.79 (0.77-0.82)	0.78 (0.75-0.81)	0.77 (0.72-0.81)
Threshold	20.4		3.48		1.01	
Sensitivity, %	52 (47-57)	49 (43-56)	33 (28-37)	30 (24-36)	21 (16-27)	28 (21-36)
Specificity, %	95 (95-96)	95 (95-97)	95 (94-96)	96 (95-97)	95 (94-96)	95 (94-96)
Misclassified, %	10 (10-11)	9 (9-10)	12 (12-13)	11 (11-12)	13 (13-14)	13 (13-14)

95% confidence intervals were estimated with 500 bootstrap replicates.

Supporting Table 17 Derivation of new cut-offs corresponding to 98% specificity for cirrhosis. The cut-offs were calculated in the training cohort and were validated in the validation cohort. Training and validation cohorts were obtained by random sampling from the IPD cohort in a 3:2 ratio.

	LSM by VCTE (n = 5489)		**FIB-4 (n = 5393)**		**NFS (n = 3248)**	
	Training (n = 3290)	Validation (n = 2199)	Training (n = 3254)	Validation (n = 2139)	Training (n = 1963)	Validation (n = 1285)
Proportion of patients with cirrhosis, %	11	10	11	11	10	11
AUC	0.90 (0.89-0.92)	0.90 (0.88-0.92)	0.81 (0.78-0.83)	0.79 (0.77-0.82)	0.78 (0.75-0.81)	0.77 (0.72-0.81)
Threshold	27.6		4.63		1.57	
Sensitivity, %	27 (23-32)	29 (22-34)	19 (15-23)	20 (15-26)	12 (8-17)	18 (13-27)
Specificity, %	98 (98-99)	98 (98-99)	98 (97-98)	98 (97-99)	98 (97-99)	98 (97-99)
Misclassified, %	10 (10-11)	9 (9-10)	10 (10-11)	10 (10-11)	11 (11-12)	11 (11-12)

95% confidence intervals were estimated with 500 bootstrap replicates.

Supporting Table 18 Diagnostic performance of combinations of NFS and LSM by VCTE, and FIB-4 and LSM by VCTE tests to reduce need for liver biopsies

	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)	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)
Prevalence, %	30	28	30	28	30	28	30	28	30	28
Threshold for blood-based NIT*	< 1.3, ≥ 3.48	< -1.455, ≥ 1.010	< 1.3, ≥ 3.48	< -1.455, ≥ 1.010	< 1.3, ≥ 3.48	< -1.455, ≥ 1.010	< 1.3, ≥ 4.63	< -1.455, ≥ 1.570	< 1.3, ≥ 4.63	< -1.455, ≥ 1.570
Threshold for VCTE, kPa*	< 7.9, ≥ 16.1	< 7.9, ≥ 16.1	< 7.9, ≥ 20.4	< 7.9, ≥ 20.4	< 8.0, ≥ 20.0	< 8.0, ≥ 20.0	< 7.9, ≥ 27.6	< 7.9, ≥ 27.6	< 8.0, ≥ 28.0	< 8.0, ≥ 28.0
Sensitivity, %	41 (40-43)	41 (39-42)	38 (37-40)	37 (35-38)	38 (37-39)	36 (34-38)	28 (27-29)	25 (24-26)	27 (26-28)	24 (23-25)
Specificity, %	88 (86-89)	88 (87-90)	90 (89-91)	90 (89-92)	90 (89-91)	90 (89-92)	95 (94-97)	96 (95-98)	96 (94-97)	96 (95-98)
PPV, %	45 (43-47)	45 (41-47)	48 (45-50)	46 (43-49)	47 (45-50)	45 (43-49)	57 (54-61)	57 (52-63)	57 (54-61)	57 (52-61)
NPV, %	86 (85-87)	87 (85-88)	86 (85-87)	87 (85-88)	86 (85-87)	86 (85-88)	86 (85-87)	86 (85-88)	86 (85-87)	86 (85-88)
Indeterminate, %	16 (15-17)	17 (16-19)	19 (18-20)	20 (18-21)	18 (17-19)	17 (18-21)	24 (23-25)	25 (23-27)	24 (23-25)	21 (23-26)
Misclassification, %	18 (17-19)	17 (15-19)	16 (15-17)	15 (14-17)	17 (15-18)	14 (14-17)	13 (12-14)	12 (10-13)	13 (12-14)	11 (10-13)
Patients undergoing VCTE, %	40 (39-42)	42 (40-44)	40 (39-42)	42 (40-44)	40 (39-42)	42 (40-44)	44 (42-45)	45 (43-47)	44 (42-45)	45 (43-47)

95% confidence intervals were estimated with 500 bootstrap replicates

*A lower cut-off was used to rule out patients with advanced fibrosis and an upper cut-off was used to rule in patients with cirrhosis. Lower cut-offs were the same as used in **Table 6** of the main manuscript. Upper cut-offs for were calculated to obtain a 95% and 98% specificity in diagnosing cirrhosis in the IPD cohort.

Supporting Table 19 Diagnostic performance of non-invasive tests in subgroup for discriminating advanced fibrosis (F3-F4).

	LSM by VCTE	FIB-4	NFS
Biopsy length < 20 mm (n = 1555)	**0.87 (0.86-0.89)**	**0.80 (0.78-0.83)**	**0.79 (0.75-0.82)**
Biopsy length ≥ 20 mm (n = 2492)	**0.83 (0.82-0.85)**	**0.75 (0.72-0.77)**	**0.72 (0.69-0.75)**
Number of portal tracts < 11 (n = 1006)	**0.86 (0.83-0.88)**	**0.79 (0.75-0.82)**	**0.78 (0.74-0.81)**
Number of portal tracts ≥ 11 (n = 851)	**0.80 (0.77-0.83)**	**0.73 (0.70-0.77)**	**0.68 (0.63-0.72)**
Intermediate quality biopsy (n = 1432)	**0.87 (0.85-0.89)**	**0.79 (0.77-0.82)**	**0.78 (0.74-0.81)**
High quality biopsy (n = 670)	**0.79 (0.75-0.83)**	**0.72 (0.68-0.76)**	**0.67 (0.62-0.73)**
BMI < 25 kg/m^2 (n = 868)	**0.91 (0.89-0.94)**	**0.81 (0.78-0.84)**	**0.76 (0.71-0.81)**$^{\#}$
25 kg/m^2 ≤ BMI < 30 kg/m^2 (n = 2127)	**0.87 (0.85-0.89)**	0.77 (0.75-0.80)	**0.74 (0.71-0.77)***
BMI ≥ 30 kg/m^2 (n = 2710)	**0.81 (0.79-0.83)**	**0.74 (0.72-0.76)**	**0.69 (0.66-0.72)**$^{*,\#}$
Continent – Europe (n = 3560)	0.85 (0.84-0.87)	0.75 (0.73-0.77)	0.72 (0.69-0.75)
Continent - Asia (n = 1278)	0.85 (0.82-0.88)	0.77 (0.73-0.80)	0.76 (0.73-0.80)
Sex – Male (n = 3165)	0.85 (0.83-0.86)	0.76 (0.74-0.78)	**0.75 (0.72-0.77)**
Sex – Female (n = 2570)	0.86 (0.84-0.87)	0.76 (0.73-0.78)	**0.71 (0.68-0.74)**
Presence of T2DM (n = 2191)	**0.81 (0.79-0.83)**	**0.73 (0.71-0.75)**	0.68 (0.65-0.70)
Lack of T2DM (n = 2763)	**0.87 (0.86-0.89)**	**0.77 (0.75-0.79)**	0.71 (0.68-0.74)
ALT < 40 U/L (n = 1656)	0.85 (0.83-0.88)	**0.73 (0.70-0.76)**	0.74 (0.70-0.78)
40 U/L ≤ ALT < 100 U/L (n = 2933)	0.86 (0.85-0.87)	**0.77 (0.76-0.79)**	0.75 (0.73-0.78)
ALT ≥ 100 U/L (n = 984)	0.83 (0.80-0.86)	0.76 (0.73-0.79)	0.77 (0.73-0.81)
AST < 40 U/L (n = 2759)	0.84 (0.82-0.86)	0.73 (0.70-0.75)	**0.76 (0.73-0.78)**
40 U/L ≤ AST < 100 U/L (n = 2385)	0.85 (0.83-0.86)	0.74 (0.72-0.76)	0.72 (0.69-0.75)
AST ≥ 100 U/L (n = 373)	0.86 (0.82-0.90)	0.71 (0.66-0.76)	**0.65 (0.58-0.72)**
ALT < 40 U/L AND AST < 40 U/L (n = 1323)	0.84 (0.81-0.87)	**0.72 (0.68-0.75)**	0.73 (0.69-0.77)
ALT ≥ 40 U/L OR AST ≥ 40 U/L (n = 4235)	0.86 (0.84-0.87)	**0.76 (0.75-0.78)**	0.75 (0.73-0.77)
Age < 43 yrs (n = 1401)	0.81 (0.77-0.84)	**0.65 (0.61-0.70)**	**0.58 (0.52-0.64)**$^{*,\#}$
43 yrs ≤ Age < 54 yrs (n = 1478)	0.84 (0.82-0.86)	0.69 (0.66-0.72)	**0.70 (0.66-0.74)***
54 yrs ≤ Age < 62 yrs (n = 1423)	0.85 (0.83-0.87)	**0.72 (0.69-0.75)**	**0.70 (0.67-0.74)**$^{\#}$
62 yrs ≤ Age (n = 1430)	0.84 (0.81-0.86)	0.70 (0.67-0.72)	0.66 (0.62-0.70)

Bold AUCs within a column and subgroup category are significantly different. Bonferroni's correction was applied when performing multiple comparisons. AUCs marked with * or $^{\#}$ are pairwise significantly different.

Supporting Table 20 Diagnostic performance of non-invasive tests in subgroup for discriminating cirrhosis (F4).

	LSM by VCTE	FIB-4	NFS
Biopsy length < 20 mm (n = 1555)	0.91 (0.88-0.93)	0.84 (0.81-0.86)	**0.83 (0.79-0.87)**
Biopsy length ≥ 20 mm (n = 2492)	0.89 (0.88-0.91)	0.78 (0.76-0.81)	**0.75 (0.71-0.78)**
Number of portal tracts < 11 (n = 1006)	**0.90 (0.87-0.94)**	0.81 (0.76-0.87)	0.76 (0.70-0.83)
Number of portal tracts ≥ 11 (n = 851)	**0.84 (0.81-0.88)**	0.77 (0.72-0.81)	0.71 (0.65-0.77)
Intermediate quality biopsy (n = 1432)	0.91 (0.88-0.93)	0.83 (0.80-0.86)	**0.83 (0.78-0.87)**
High quality biopsy (n = 670)	0.87 (0.83-0.90)	0.87 (0.83-0.90)	**0.69 (0.62-0.76)**
BMI < 25 kg/m^2 (n = 868)	**0.93 (0.91-0.95)$^{\#}$**	**0.84 (0.80-0.88)**	0.77 (0.69-0.84)
25 kg/m^2 ≤ BMI < 30 kg/m^2 (n = 2127)	**0.92 (0.91-0.94)***	0.82 (0.78-0.85)	**0.83 (0.80-0.86)**
BMI ≥ 30 kg/m^2 (n = 2710)	**0.87 (0.85-0.89)$^{*,\#}$**	**0.77 (0.75-0.80)**	**0.73 (0.69-0.76)**
Continent – Europe (n = 3560)	0.90 (0.89-0.92)	0.80 (0.78-0.82)	0.77 (0.74-0.81)
Continent - Asia (n = 1278)	0.92 (0.89-0.94)	0.81 (0.77-0.85)	0.80 (0.75-0.85)
Sex – Male (n = 3165)	0.91 (0.89-0.92)	0.81 (0.78-0.83)	**0.80 (0.77-0.83)**
Sex – Female (n = 2570)	0.89 (0.88-0.91)	0.78 (0.76-0.81)	**0.74 (0.71-0.78)**
Presence of T2DM (n = 2191)	**0.85 (0.83-0.87)**	**0.74 (0.72-0.77)**	**0.70 (0.67-0.70)**
Lack of T2DM (n = 2763)	**0.94 (0.92-0.95)**	**0.85 (0.83-0.88)**	**0.80 (0.76-0.84)**
ALT < 40 U/L (n = 1656)	0.91 (0.89-0.93)	0.79 (0.75-0.83)	0.77 (0.73-0.82)
40 U/L ≤ ALT < 100 U/L (n = 2933)	0.90 (0.88-0.92)	0.80 (0.78-0.83)	0.77 (0.74-0.80)
ALT ≥ 100 U/L (n = 984)	0.90 (0.87-0.93)	0.79 (0.75-0.84)	0.82 (0.76-0.88)
AST < 40 U/L (n = 2759)	0.90 (0.88-0.92)	0.78 (0.75-0.81)	0.80 (0.77-0.84)
40 U/L ≤ AST < 100 U/L (n = 2385)	0.89 (0.87-0.91)	0.78 (0.76-0.81)	0.75 (0.72-0.79)
AST ≥ 100 U/L (n = 373)	0.90 (0.86-0.94)	0.77 (0.71-0.84)	0.75 (0.66-0.84)
ALT < 40 U/L AND AST < 40 U/L (n = 1323)	0.91 (0.89-0.93)	0.76 (0.72-0.81)	0.75 (0.69-0.80)
ALT ≥ 40 U/L OR AST ≥ 40 U/L (n = 4235)	0.90 (0.89-0.91)	0.80 (0.78-0.82)	0.79 (0.76-0.81)
Age < 43 yrs (n = 1401)	**0.97 (0.95-0.99)$^{*,\#,\%}$**	0.82 (0.75-0.88)	0.72 (0.55-0.89)
43 yrs ≤ Age < 54 yrs (n = 1478)	**0.90 (0.87-0.93)***	0.77 (0.72-0.82)	0.74 (0.67-0.81)
54 yrs ≤ Age < 62 yrs (n = 1423)	**0.87 (0.85-0.90)$^{\#}$**	0.75 (0.71-0.78)	0.74 (0.69-0.78)
62 yrs ≤ Age (n = 1430)	**0.86 (0.84-0.89)$^{\%}$**	0.72 (0.69-0.76)	0.66 (0.62-0.71)

Bold AUCs within a column and subgroup category are significantly different. Bonferroni's correction was applied when performing multiple comparisons. AUCs marked with *,$^{\#}$ or $^{\%}$ are pairwise significantly different.

Supporting Table 21 Subgroup analysis on the impact of reliability of liver stiffness measurements (LSM) on diagnostic performance in detecting advanced fibrosis.

	AUROC (95% CI) for diagnosing advanced fibrosis	AUROC (95% CI) for diagnosing cirrhosis
Reliable LSM by VCTE (median LSM < 7.1 kPa OR (median LSM ≥ 7.1 kPa AND IQR/median LSM < 0.30)	**0.86 (0.85-0.87)**	**0.91 (0.90-0.92)**
Unreliable LSM by VCTE (median LSM ≥ 7.1 kPa AND IQR/median LSM > 0.30)	**0.75 (0.70-0.80)**	**0.81 (0.76-0.86)**
Reliable LSM by VCTE (IQR/median LSM < 0.30)	0.86 (0.84-0.87)	0.90 (0.89-0.92)
Unreliable LSM by VCTE (IQR/median LSM ≥ 0.30)	0.84 (0.82-0.86)	0.88 (0.86-0.91)

VCTE – vibration-controlled transient elastography; 95% confidence intervals were estimated using 500 bootstrap iterations. Bold AUCs within a column and subgroup category are significantly different ($p < 0.05$).

Supporting Table 22 Subgroup analysis based on choice of probe type (in patients with data available from both probes) compared to the diagnostic accuracy of LSM by VCTE calculated in the entire IPD cohort.

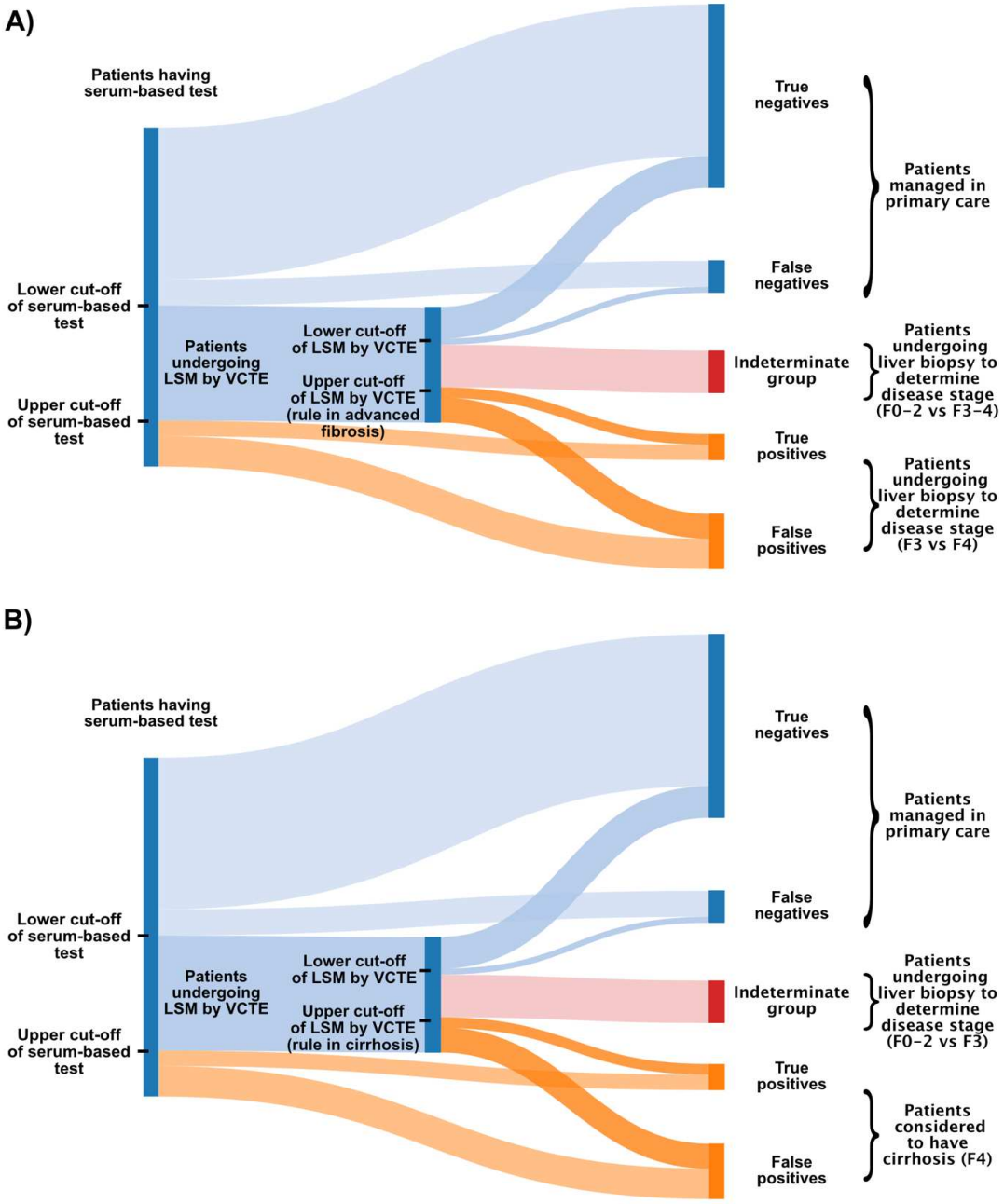
	AUROC (95% CI) for diagnosing advanced fibrosis	AUROC (95% CI) for diagnosing cirrhosis
Entire cohort (n = 5489)	0.85 (0.84-0.86)	0.90 (0.89-0.91)
M probe only (where measurements performed with both probes were performed) (n = 799)	0.84 (0.82-0.87)	0.86 (0.83-0.90)
XL probe only (where measurements performed with both probes were performed) (n = 799)	0.83 (0.80-0.86)	0.87 (0.84-0.90)

Supporting Table 23 Sensitivity analysis on the impact of probe selection on diagnostic performance in detecting advanced fibrosis. Thresholds were calculated from the entire IPD cohort.

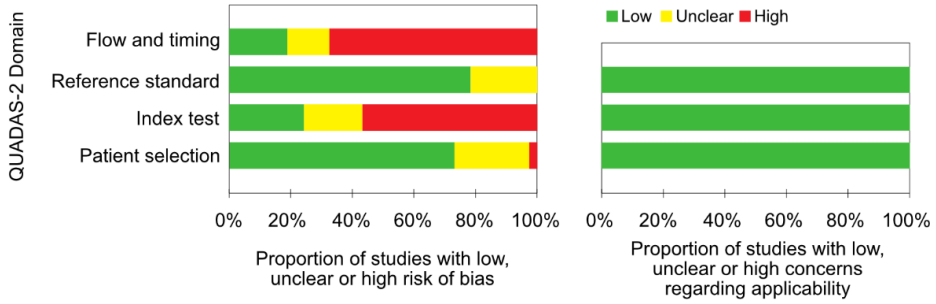
	All patients with LSM (n = 5489)			**Patients with BMI < 30 kg/m^2 and M probe OR BMI ≥ 30 kg/m^2 and XL probe (n = 4464)**		
AUC (95% CI)		0.85 (0.84-0.86)			0.86 (0.85-0.87)	
Thresholds, kPa	9.1	< 7.4	≥ 12.1	9.1	< 7.4	≥ 12.1
Sensitivity, %	77 (75-79)	90 (89-91)	55 (52-57)	75 (72-78)	89 (87-91)	53 (50-56)
Specificity, %	78 (76-79)	60 (59-61)	90 (89-91)	81 (79-82)	65 (63-67)	92 (91-93)
Misclassified, %	22 (22-23)	31 (31-32)	21 (20-21)	21 (20-22)	28 (27-29)	20 (18-21)

95% confidence intervals were estimated with 500 bootstrap replicates.

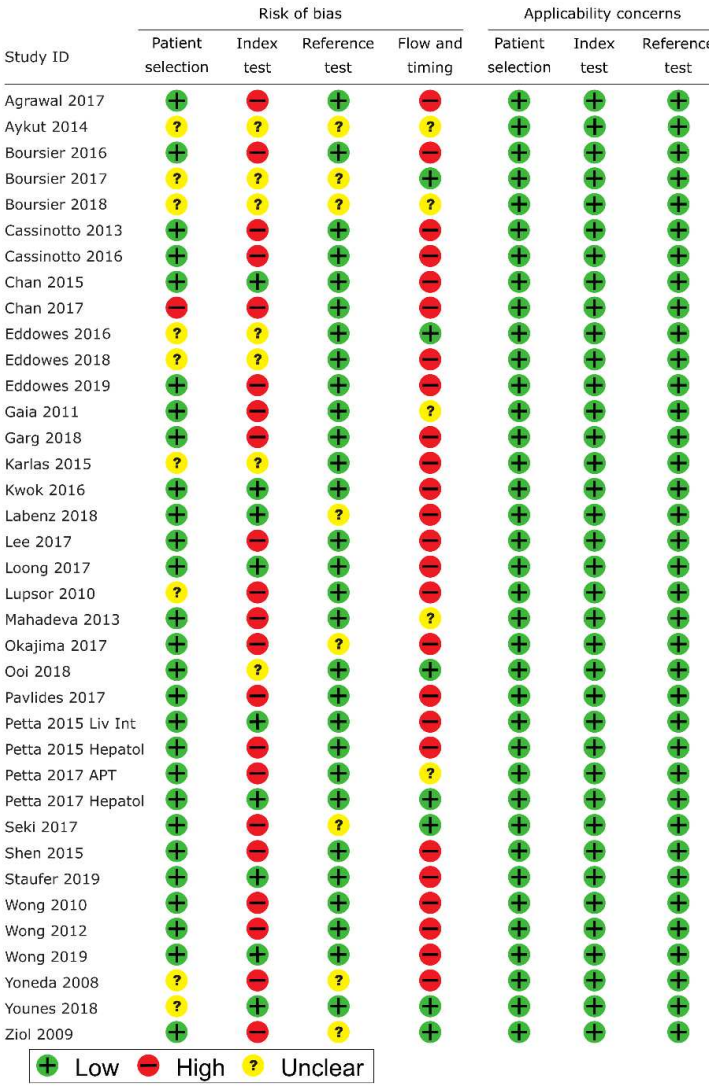
Supporting Figures



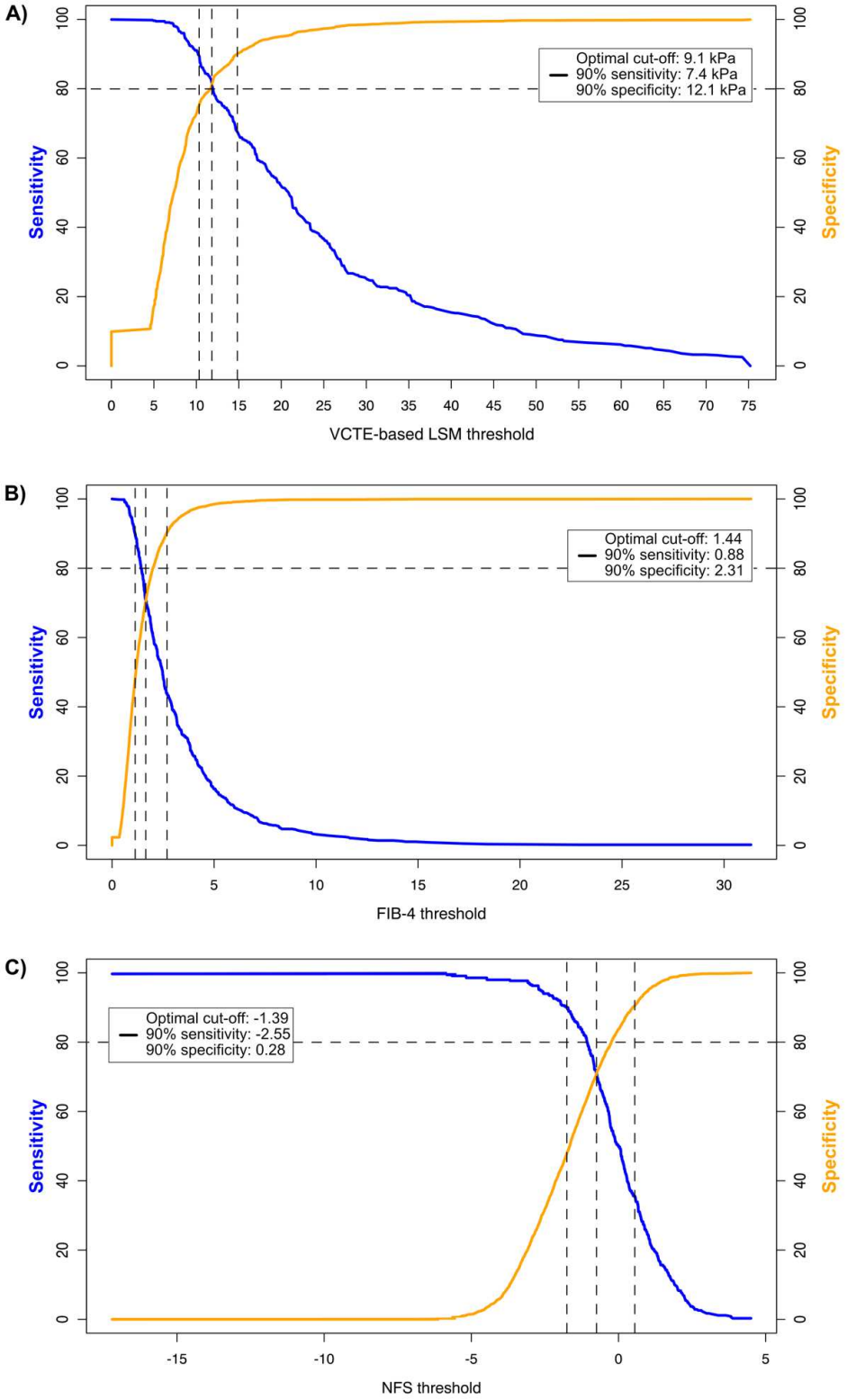
Supporting Figure 1 "Traditional" (A) and newly proposed two-tier algorithms (B) for using non-invasive tests in clinical care. (A) In the traditional application of NITs, patients with NIT vaues below the lower cut-offs are "ruled out" and are managed in primary care. Those with indeterminate NIT values and those "ruled in" with vaues above the upper cut-offs still need to undergo liver biopsy in order to stage their disease. Patients with indeterminate NITs need a liver biopsy to rule out advanced fibrosis, while patients ruled in for advanced fibrosis still need a biopsy to diagnose cirrhosis, as those with cirrhosis are managed differently (they need surveillanve for hepatocellular cancer and screening for oesophageal varices). (B) In the proposed algorithms we use upper cut-off values to rule in cirrhosis, where those who are ruled in are thereby managed as having cirrhosis without the need for liver biopsy. Patients in the indeterminate group still require biopsy to correctly stage their disease.



Supporting Figure 2 Risk of bias and applicability concerns



Supporting Figure 3 Methodological quality summary



Supporting Figure 4 Distribution of sensitivities and specificities over the possible threshold ranges for LSM by VCTE (A), FIB-4 (B) and NFS (C) when considering the diagnosis of cirrhosis. Horizontal dashed lines are representing the minimum acceptable criteria for considering a test as having high sensitivity (≥80%) and high specificity (≥80%).

References

1. 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; 70:1521–1530.

2. Wong VWS, Vergniol J, Wong GLH, et al.: Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* 2010; 51:454–462.

3. 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; 156:1717–1730.

4. Wong VWS, Irles M, Wong GLH, et al.: Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. *Gut* 2019; 68:2057–2064.

5. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ: Comparison of Noninvasive Markers of Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol* 2009; 7:1104–1112.

6. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP: Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut* 2010; 59:1265 LP – 1269.

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; 45:846–854.

8. 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; 114:916–928.

9. 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; 71:389–396.

10. 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; 46:617–627.

11. Tapper EB, Challies T, Nasser I, Afdhal NH, Lai M: The Performance of Vibration Controlled Transient Elastography in a US Cohort of Patients With Nonalcoholic Fatty Liver Disease. *Am J Gastroenterol* 2016; 111:677–84.

12. 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; 17:630-637.e8.

13. 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; 63:1817–1827.

14. Xun YH, Fan JG, Zang GQ, et al.: Suboptimal performance of simple noninvasive tests for advanced fibrosis in Chinese patients with nonalcoholic fatty liver disease. *J Dig Dis* 2012; 13:588–595.

15. Vilar-Gomez E, Chalasani N: Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J Hepatol* 2018:305–315.

16. Sun W, Cui H, Li N, et al.: Comparison of FIB-4 index, NAFLD fibrosis score and BARD score for prediction of advanced fibrosis in adult patients with non-alcoholic fatty liver disease: A meta-

analysis study. *Hepatol Res* 2016; 46:862–870.

17. 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; 71:371–378.

18. Demir M, Lang S, Nierhoff D, et al.: Stepwise Combination of Simple Noninvasive Fibrosis Scoring Systems Increases Diagnostic Accuracy in Nonalcoholic Fatty Liver Disease. *J Clin Gastroenterol* 2013; 47:719–726.

19. Siddiqui MS, Yamada G, Vuppalanchi R, et al.: Diagnostic Accuracy of Noninvasive Fibrosis Models to Detect Change in Fibrosis Stage. *Clin Gastroenterol Hepatol* 2019; 17:1877-1885.e5.

20. 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; 35:1566–1573.

21. Dowman JK, Tomlinson JW, Newsome PN: Systematic review: The diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Aliment Pharmacol Ther* 2011:525–540.

22. Fowell AJ, Fancey K, Gamble K, et al.: Evaluation of a primary to secondary care referral pathway and novel nurse-led one-stop clinic for patients with suspected non-alcoholic fatty liver disease. *Frontline Gastroenterol* 2020; 0:1–6.

23. de Franchis R: Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. *J Hepatol* 2015; 63:743–752.