Supplement to:

Diagnostic accuracy of elastography, and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis

Emmanuel A. Selvaraj[1,2,3]*, Ferenc E. Mózes[1]*, Arjun N. A. Jayaswal[1]*, Mohammad H. Zafarmand[4], Yasaman Vali[4], Jenny Lee[4], Christina K. Levick[1], Liam A. J. Young[1], Naaventhan Palaniyappan[5], Guruprasad P. Aithal[5], Manuel R. Gomez[6], Chang-Hai Liu[6,7], M. Julia Brosnan[8], Theresa A. Tuthill[8], Quentin M. Anstee[9], Stefan Neubauer[1], Stephen Harrison[1], Patrick M. Bossuyt[4], Michael Pavlides[1,2,3], on behalf of the LITMUS investigators

*joint first authors

Table of 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
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 Panu Luukkonen
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

Supplementary results

Narrative synthesis of MRI techniques

Two studies assessed cT_1 measured by LMS[1, 2], one evaluated deMILI[3] and one tested DWI[4]. Amongst the MRI techniques, only DWI was used to assess any stage of fibrosis (≥F1). DWI-derived parameters performed poorly in diagnosing any fibrosis stage (≥F1); however, the NPV for each parameter was acceptable (>80%). deMILI predicted significant fibrosis with AUC of 0.94 and 0.85 in a training and validation cohort, respectivey. LMS cT_1 was used to predict significant fibrosis (AUC 0.73 and 0.78), advanced fibrosis (AUC 0.73) and cirrhosis (AUC 0.85).

The diagnostic performance of all three MRI-based methods was assessed for distinguishing NASH from simple steatosis. LMS cT_1 had an AUC of 0.69 and 0.80 from two studies[2, 1], and deMILI had AUCs of 0.88 and 0.83 in the training and validation cohorts respectively.[3] DWI had AUCs of 0.74, 0.68, and 0.61 for the pure molecular diffusion coefficient, perfusion-related diffusion coefficient, and the perfusion fraction.[4] Forest plots for each target condition and each MRI modality are presented in **Figs. S23-S2**.

Supplementary tables

Table S1 PRISMA-DTA checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	5
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	7
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	7
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	8
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	9
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	9-10
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	10
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Table S2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	9-10
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	11
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	9
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	11
Diagnostic accuracy	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	11

measures			
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	12
Section/topic	**#**	**PRISMA-DTA Checklist Item**	**Reported on page #**
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	12-13
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	12
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	14 Fig.1
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	14-15 Table 1
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	14 Figs. S1-S5
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figs. 2-6 Figs.S6-S21
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Table 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	18
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	19-22
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	22-23
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	21-22
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	Title page

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163. For more information, visit: **www.prisma-statement.org**

Table S2 Descriptions of index tests.

Index tests		Details of technique
Ultrasound	VCTE	Vibration-controlled transient elastography (VCTE; FibroScan® (Echosens, Paris, France) measures liver stiffness by measuring shear wave velocity. Vibrations of mild amplitude and low frequency (50 Hz) are transmitted by a small transducer on the end of an ultrasound probe, inducing a shear wave that propagates through the liver. The probe also has a transducer on the end that can measure the velocity of the shear wave (in meters per second). The shear wave velocity can then be converted into liver stiffness, which is expressed in kilopascals.[5]
	pSWE	Acoustic radiation force impulse (ARFI) is a type of shear wave elastography (SWE) that is incorporated into a diagnostic ultrasound system and is performed using an ultrasound probe to emit a single pulse of shear wave using an excitation method known as ARFI.[6] The propagation velocity of the shear waves is measured at single point measurement and therefore also known as point SWE (pSWE). Only studies that used the Virtual Touch Quantification (VTQ) on Siemens platform are included in the present study.
	2DSWE	Two-dimensional (2D) SWE uses supersonic shear imaging (SSI) technique to measure acoustically-generated tissue shear wave propagation speeds to derive estimates of liver stiffness in real-time.[6] It has the advantage of simultaneous anatomic B-mode US imaging, thus allows selection of a larger liver parenchymal region of interest devoid of blood vessels and larger than pSWE. Only studies that used the Aixplorer® platform are included in the present study.
Magnetic resonance	MRE	Magnetic resonance elastography (MRE; Resoundant, Rochester, USA) measures liver stiffness. A mechanical shear wave generator is installed outside the scanner room which is connected to a plastic circular disc that is attached to the patient in a position overlying the liver during the scan. Oscillating signals produced by the generator are transmitted to the plastic disc and then through the liver as acoustic waves. Specific MR sequences are then used to visualise the propagation of these waves through the liver.

		Dedicated software is used to analyse these data and produce shear wave elastograms that can be used to measure liver stiffness.[7]
	LMS-cT1	Liver*MultiScan*™ (LMS; Perspectum Diagnostics, Oxford, UK) measures multiple MRI parameters (T_1, T_2^* and proton density fat fraction (PDFF). Central to this technology is the correction of the T_1 relaxation time, as measured by the shortened modified Look-Locker inversion recovery (shMOLLI) technique[8], for iron. T_1 is an intrinsic property of tissues that changes to reflect alterations in extracellular fluid[9]. T_1 is, however, confounded by the presence of iron. In LMS, the measured T_1 is corrected for iron (the concentration of which is estimated from T_2^*), to produce an "iron corrected T_1 (cT_1)", a parameter that has been reported to improve diagnostic accuracy.[10]
	DWI	Diffusion weighted imaging (DWI) uses MRI acquisition and analysis techniques to track diffusion of water in tissues. Quantitative measures of diffusion can be produced by measuring the magnitude (apparent diffusion coefficient; ADC) and directionality (fractional anisotropy) of diffusion. The accumulation of steatosis, inflammation and fibrosis can lead to changes in water diffusion[11] and these can be measured using various DWI techniques.
	deMILI	Detection of metabolic liver injury (deMILI) MRI uses optical analysis of magnetic resonance images to define the NASHMRI (0-1) and FibroMRI (0-1) measures of NASH and liver fibrosis respectively. Image acquisition does not require injection of intravenous contrast and includes just three sequences: single-shot fast spin echo T_2-weighted (SSFE-T_2), fast short inversion time inversion recovery (FAST-STIR), 2D fast-field-echo T1-weighted (gradient echo) (2D-FFE-T1).[3]

Table S3 Electronic search strategy.

Index test	Search terms
VCTE	((((("transient elastography"[Title/Abstract] OR "liver stiffness"[Title/Abstract] OR "vibration controlled transient elastography"[Title/Abstract] OR VCTE[Title/Abstract] OR fibroscan[Title/Abstract]))) OR "Elasticity Imaging Techniques"[Mesh])) AND (("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))
MRE	((("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))) **AND** ((MRE[Title/Abstract] OR "MR elastography"[Title/Abstract] OR "magnetic resonance elastography"[Title/Abstract]))
pSWE	((((("point shear wave elastography" OR pSWE OR "acoustic radiation force impulse" OR ARFI))) OR "Elasticity Imaging Techniques"[Mesh])) AND (("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR

	"nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))
2DSWE	((((("2D shear wave elastography" OR 2DSWE OR "acoustic radiation force impulse" OR ARFI))) OR "Elasticity Imaging Techniques"[Mesh])) AND (("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))
MRI	((("Magnetic Resonance Imaging"[Mesh]) OR ((T1 OR "iron corrected T1" OR "cT1" OR "Liver Inflammation and Fibrosis score" OR LIF OR "multiparametric MR" OR "multi-parametric MR")))) AND (("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))
	((("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic

steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))) AND (((("diffusion weighted imaging"[Title/Abstract] OR DWI[Title/Abstract] OR "intravoxel incoherent motion"[Title/Abstract] OR IVIM[Title/Abstract] OR "Diffusion tensor imaging"[Title/Abstract] OR DTI[Title/Abstract]))) OR "Diffusion Magnetic Resonance Imaging"[Mesh])

(((((deMILI OR "detection of metabolic liver injury" OR "NASHMRI" OR "fibroMRI"))) OR "Magnetic Resonance Imaging"[Mesh])) AND (("Non-alcoholic Fatty Liver Disease"[Mesh]) OR (("non-alcoholic fatty liver"[Title/Abstract] OR "non alcoholic fatty liver"[Title/Abstract] OR "nonalcoholic fatty liver"[Title/Abstract] OR "non-alcoholic steatohepatitis"[Title/Abstract] OR "non alcoholic steatohepatitis"[Title/Abstract] OR "nonalcoholic steatohepatitis"[Title/Abstract] OR steatohepatitis[Title/Abstract] OR NASH[Title/Abstract] OR NAFLD[Title/Abstract])))

Table S4 Proportion of technical failures of index tests in individual studies.

Study ID	Number of patients, n	Technical failure, %
VCTE		
Agrawal 2017 (UK)	25	-
AlJuboori 2018 (USA)[‡]	60	-
Anstee 2019 (UK)	F0-2: 284 F3-4: 1323	-
Attia 2016 (Germany)	M probe (overweight): XL probe (obese):	1.1 3.3
Aykut 2014 (Turkey)	88	-
Boursier 2016 (France)	452	14.1
Boursier 2019 (France)	Training:625 Validation:313	-
Cardoso 2019 (Brazil)	M probe: 81 XL probe: 81	4.7 0
Cassinotto 2013 (France)	M probe: 48 XL probe: 49	21.3 5.8
Cassinotto 2016 (France)	M probe: 223	23.4
Chan 2015 (Malaysia)	Training: 101 Validation: 46	3.8 4.2
Chan 2017 (Malaysia)	M probe: 60 XL probe: 60	3.3 1.7
Clet 2018 (UK)[‡]	176	-
Das 2012 (India)	25	9.6
Eddowes 2018 (UK)	50	6.0
Eddowes 2019 (UK)	373	2.7
Ergelen 2016 (Turkey)	63	1.6
Forlano 2017 (UK)[‡]	238	-
Gaia 2011 (Italy)	72	-
Gaia 2015 (Italy)	58	1.9
Gallego-Durán 2016 (Spain)	126	-
Garg 2018 (India)	XL probe: 76	12.1
Hee 2014 (Australia)[‡]	98	-
Imajo 2016 (Japan)	M probe: 142	10.6
Inadomi 2020 (Japan)	224	-
Kao 2020 (Taiwan)	Training: 73 Validation: 50	-
Karlas 2015 (Germany)	Bariatric: 41 Non-bariatric: 45	48.8 6.3
Kumar 2013 (India)	Non-cirrhotic: 120 Cirrhotic: 85 (M probe)	10.1
Kwok 2016 (Hong Kong)	94	1.8
Labenz 2018 (Germany)	126	9.4

Lee 2016 (South Korea)	183	1.6
Lee 2017 (South Korea)	94	21.3
Lee 2019 (South Korea)	184	-
Lee 2020 (South Korea)	130	2.5
Leong 2020 (Malaysia)	100	8.0
Loong 2017 (Hong Kong)	215	11.5
Lupsor 2010 (Romania)	M probe: 72	9.7
Mahadeva 2013 (Malaysia)	M probe: 131	8.4
Morrison 2020 (USA)‡	162	-
Myers 2012 (Canada)	75	-
Naveau 2014 (France)	100	19.1
Naveau 2017 (France)	Retrospective: 194 Prospective: 123	-
Oeda 2020 (Japan)	Training: 96	21.3
	Validation: 103	14.2
	M probe: 104	14.8
	XL probe: 109	10.7
Okajima 2017 (Japan)	M probe: 163	5.8
Ooi 2018 (Australia)	XL probe: 66	19.5
Pais 2011 (France)	208	10.7
Park 2017 (USA)	97	6.7
Pavlides 2017 (UK)	71	46.5
Petta 2011 (Italy)	M probe: 146	13.6
Petta 2015_Liv Int (Italy)	Training: 179 (M probe)	12.7
	Validation: 142 (M probe)	23
Petta 2015_Hepatol (Italy)	M probe: 253	17.3
Petta 2017_APT (Italy)	761	-
Petta 2017_Hepatol (Italy)	324	-
Rosso 2016 (Italy)	105	8.3
Seki 2017 (Japan)	171	-
Shen 2015 (China)	M probe: 101	9.7
Shima 2020 (Japan)	M probe: 249	10.4
Staufer 2019 (Austria)	140	6.7
Tapper 2016 (USA)	M probe: 120	26.8
Wong 2010 (Hong Kong)	M probe: 246	10.2
Wong 2012 (Hong Kong)	193	9.8 M: 10 XL: 2
Wong 2018 (Hong Kong)	496	M: 21.2 XL: 12.7
Yoneda 2008 (Japan)	M probe: 97	4.9
Younes 2018 (Italy)	292	2.4
Ziol 2009 (France)	13	-
MRE		
Chen 2011 (USA)	58	0.0

Costa-Silva 2018 (Brazil)	49	0.0
Cui 2015 (USA)	102	0.0
Cui 2016 (USA)	125	0.8
Imajo 2016 (Japan)	142	0.0
Kim 2013 (USA)	142	-
Kim 2020 (South Korea)	47	-
Lee 2020 (South Korea)	130	1.3
Loomba 2013 (USA)‡	52	-
Loomba 2014 (USA)	117	-
Loomba 2016 (USA)	99	0.0
Park 2017 (USA)	104	0.0
pSWE		
Attia 2016 (Germany)	Overweight: 61 Obese: 26	1.0
Cassinotto 2013 (France)	60	1.6
Cassinotto 2016 (France)	236	18.9
Cui 2016 (USA)	125	2.4
Fierbinteanu-Braticevici 2013 (Romania)	Simple steatosis: 21 NASH: 43	0.0
Joo 2018 (South Korea)	315	8.5
Karlas 2015 (Germany)	Bariatric: 41 Non-bariatric: 48	9.8 0.0
Lee 2017 (South Korea)	83	11.7
Medellin 2019 (Canada)	51	42.6
Palmeri 2011 (USA)	135	21.5
Zhang 2014 (China)	67	-
2DSWE		
Cassinotto 2016 (France)	232	20.3
Lee 2017 (South Korea)	83	26.6
Takeuchi 2018 (Japan)	71	2.7
Ozturk 2020 (USA)	116	12.1
MRI		
Eddowes 2018 (UK)	50	5.6
Gallego-Duran 2016 (Spain)	126	0.0
Parente 2015 (Brazil)	59	17.5
Pavlides 2017 (UK)	71	5.3

‡ Abstracts
- Not reported or unable to derive

Table S5 Details of pre-defined cut-offs for index tests.

Study ID	Lower limit	Upper limit	Single cut-off	Source of pre-defined cut-offs
VCTE (kPa)				
Any fibrosis (F0 vs F1-4)				
Ooi 2018 (Australia)	-	-	5.25	Kwok R et al. APT 2014[12]
Significant fibrosis (F0-1 vs F2-4)				
Eddowes 2018 (UK)	-	-	5.8, 7.0, 7.9, 9.0	Wong VW et al. Hepatology 2010[13]
Morrison 2020 (USA)‡	-	-	8.5	unclear
Ooi 2018 (Australia)	-	-	7.0	Kwok R et al. APT 2014[12]
Staufer 2019 (Austria)	-	-	8.2	Eddowes PJ et al. Gastroenterology 2019[14]
Wong 2018 (Hong Kong)	5.00	-	-	de Franchis R et al. J Hepatol 2015[15]
Advanced fibrosis (F0-2 vs F3-4)				
Anstee 2019 (UK)	9.9	11.4	-	unclear
Hee 2014 (Australia)‡	7.9(M)	9.6(M)	-	unclear
	7.2(XL)	9.3(XL)	-	
Kwok 2016 (Hong Kong)	-	-	9.6-11.4 (M) 9.3-10.9 (XL)	Wong VW et al. Hepatology 2010[13] Wong VW et al. AJG 2012[16]
Labenz 2018 (Germany)	-	-	12.0	Vuppalanchi R et al. Hepatology. 2018[17] Siddiqui MS et al. CGH. 2018[18]
Loong 2017 (Hong Kong)	7.9	9.6	-	Wong VW et al. Hepatology 2010[13]
Morrison 2020 (USA)‡	-	-	9.5	unclear
Ooi 2018 (Australia)	-	-	10.3	Kwok R et al. APT 2014[12]
Pais 2011 (France)‡	7.9	9.6	-	Wong VW et al. Hepatology 2010[13]
Petta 2015_Liv Int (Italy)	7.9	9.6	-	Wong VW et al. Hepatology 2010[13]
Petta 2017_APT (Italy)	7.9	9.6	-	Wong VW et al. Hepatology 2010[13]
Staufer 2019 (Austria)	-	-	9.7	Eddowes PJ et al. Gastroenterology 2019[14]
Wong 2018 (Hong Kong)	10.0	15.0	-	de Franchis R et al. J Hepatol 2015[15]
Cirrhosis				
Kwok 2016 (Hong Kong)	-	-	11.5 (M)	Wong VW et al. Hepatology 2010[13]
			11.0 (XL)	Wong VW et al. AJG 2012[16]
Morrison 2020 (USA)‡	-	-	12.0	

Wong 2018 (Hong Kong)	10.0	15.0	-	de Franchis R et al. J Hepatol 2015[15]
MRE (kPa)				
Advanced fibrosis				
Cui 2015 (USA)	-	-	3.64	Loomba R et al. Hepatology. 2014[19]
pSWE (m/s)				
Significant fibrosis				
Medellin 2019 (Canada)	-	-	1.20 1.34	Barr RG et al. Radiology 2015[20] Friedrich-Rust M et al. EJR 2012[21]
Advanced fibrosis				
Medellin 2019 (Canada)	-	-	2.20 1.55	Barr RG et al. Radiology 2015[20] Friedrich-Rust M et al. EJR2012[21]
Cirrhosis				
Medellin 2019 (Canada)	-	-	2.20 1.80	Barr RG et al. Radiology 2015[20] Friedrich-Rust M et al. EJR2012[21]
Iron corrected T1 measured by Liver*MultiScan*™ (ms)				
Significant fibrosis				
Eddowes 2018 (UK)	-	-	822 875	Banerjee R et al. J Hepatol. 2014[22] Pavlides M et al. J Hepatol. 2016[23]

‡Abstracts

Table S6 Details of liver biopsy in individual studies.

Study ID	Needle gauge, G	Biopsy length, mm	Number of portal tracts, n	Other details
VCTE				
Agrawal 2017 (UK)	16	29	-	
AlJuboori 2018 (USA)	-	-	-	Intraoperative
Anstee 2019 (UK)	-	-	-	
Attia 2016 (Germany)	16	>10	>6	
Aykut 2014 (Turkey)	-	>20	>11	
Boursier 2016 (France)	-	26	-	
Boursier 2019 (France)	-	27	-	
Cardoso 2020 (Brazil)	16	>15	-	
Cassinotto 2013 (France)	-	22.5-26.8	20.9-21.3	
Cassinotto 2016 (France)	-	26.8	-	
Chan 2015 (Malaysia)	18	14.7-15.3	8.3-8.4	
Chan 2017 (Malaysia)	18	14	7	
Clet 2018 (UK)	-	-	-	
Das 2012 (India)	-	27.3	9	
Eddowes 2018 (UK)	16	25	≥11	
Eddowes 2019 (UK)	-	23	-	
Ergelen 2016 (Turkey)	16	>20	>11	
Forlano 2017 (UK)	-	-	-	
Gaia 2011 (Italy)	-	25.2	-	
Gaia 2015 (Italy)	-	>20	-	
Gallego-Durán 2016 (Spain)	-	20	≥15	
Garg 2018 (India)	16	>15	>6	
Hee 2014 (Australia)	-	-	-	
Imajo 2016 (Japan)	16	21.3	14.3	
Inadomi 2020 (Japan)	16	-	-	Intraoperative
Kao 2020 (Taiwan)	-	-	-	
Karlas 2015 (Germany)	-	-	>50	
Kumar 2013 (India)	18	-	-	
Kwok 2016 (Hong Kong)	16	20	8	
Labenz 2018 (Germany)	16	-	-	
Lee 2016 (South Korea)	-	>15	-	
Lee 2017 (South Korea)	-	-	-	
Lee 2019 (South Korea)	-	≥15	≥10	
Lee 2020 (South Korea)	18	≥20	-	
Leong 2020 (Malaysia)	18	14	6	
Loong 2017 (Hong Kong)	16	22	-	
Lupsor 2010 (Romania)	14	11	11	
Mahadeva 2013 (Malaysia)	18	13		
Morrison 2020 (USA)	-	-	-	
Myers 2012 (Canada)	-	>15	>6	

Naveau 2014 (France)	14	≥10	≥10	
Naveau 2017 (France)	14	> 10	-	
Oeda 2020 (Japan)	16	≥20	-	
Okajima 2017 (Japan)	16	19.6	> 6	Intraoperative
Ooi 2018 (Australia)	-	-	-	Intraoperative
Pais 2011 (France)	-	25	-	
Park 2017 (USA)	-	-	-	
Pavlides 2017 (UK)	18/19	18	10	Intraoperative
Petta 2011 (Italy)	-	17	>10	
Petta 2015_Liv Int (Italy)	-	16.4	-	
Petta 2015_Hepatol (Italy)	16	17-20	-	
Petta 2017_APT (Italy)	-	27.3	-	
Petta 2017_Hepatol (Italy)	-	>15	>10	
Rosso 2016 (Italy)	-	-	>11	
Seki 2017 (Japan)	16	≥20	-	
Shen 2015 (China)	18	≥16	≥6	
Shima 2020 (Japan)	16	>20	-	
Staufer 2019 (Austria)	-	≥15	-	
Tapper 2016 (USA)	-	-	-	
Wong 2010 (Hong Kong)	16	21	-	
Wong 2012 (Hong Kong)	16	24	-	
Wong 2018 (Hong Kong)	16	26	11	
Yoneda 2008 (Japan)	18	>20	>7	
Younes 2018 (Italy)	-	-	-	
Ziol 2009 (France)	-	-	-	
MRE				
Chen 2011 (USA)	-	-	-	
Costa-Silva 2018 (Brazil)	-	-	-	
Cui 2015 (USA)	-	-	-	
Cui 2016 (USA)	-	-	-	
Imajo 2016 (Japan)	16	21.3	14.3	
Kim 2013 (USA)	-	-	-	
Kim 2020 (South Korea)	18	-	-	
Lee 2020 (South Korea)	18	≥20	-	
Loomba 2013 (USA)	-	-	-	
Loomba 2014 (USA)	-	23.8	13.7	
Loomba 2016 (USA)	-	23.2	13.7	
Park 2017 (USA)	-	-	-	
pSWE				
Attia 2016 (Germany)	16	>10	>6	
Cassinotto 2013 (France)	-	26.8	-	
Cassinotto 2016 (France)	-	25.4	20.9	
Cui 2016 (USA)	-	-	-	
Fierbinteanu-Braticevici 2013 (Romania)	-	22	>8	
Joo 2018 (South Korea)	-	26	10	

Karlas 2015 (Germany)	-	-	>50	
Lee 2017 (South Korea)	-	-	-	Intraoperative
Medellin 2019 (Canada)	-	15-20	-	
Palmeri 2011 (USA)	-	-	-	
Zhang 2013 (China)	16	>15	-	
2D-SWE				
Cassinotto 2016 (France)	-	26.8	-	
Lee 2017 (South Korea)	-	-	-	
Takeuchi 2018 (Japan)	16	>20	>7	
Ozturk 2020 (USA)	-	-	-	
MRI				
Eddowes 2018 (UK)	16	25	≥11	
Gallego-Duran 2016 (Spain)	-	17.5	≥15	
Parente 2015 (Brazil)	16	>20	-	
Pavlides 2017 (UK)	18/19	18	10	

Table S7 Calculated sensitivities and specificities at cut-offs identified from the literature for the detection of advanced fibrosis by VCTE and their corresponding PPVs and NPVs for different prevalences derived using the multiple-thresholds model

Cut-off, kPa	Se, %	95% CI, %	Sp, %	95% CI, %	Prevalence, %	PPV, %	NPV, %	FP*	FN*
7.1 (Eddowes 2019)	87.9	81.4-92.3	66.7	58.3-74.2	5	12.2	99.1	32	1
					10	22.7	98.0	30	1
					20	39.7	95.7	27	2
					30	53.1	92.8	23	4
					40	63.8	89.2	20	5
					50	72.5	84.6	17	6
7.9 (Wong 2010)	84.7	77.5-89.9	71.6	64.2-78.0	5	13.6	98.9	27	1
					10	24.9	97.7	26	2
					20	42.7	94.9	23	3
					30	56.1	91.6	20	5
					40	66.6	87.5	17	6
					50	74.9	82.4	14	8
9.6 (Wong 2010)	75.7	67.0-82.7	80.5	74.9-85.1	5	17.0	98.4	19	1
					10	30.1	96.8	18	2
					20	49.3	93.0	16	5
					30	62.5	88.6	14	7
					40	72.1	83.3	12	10
					50	79.5	76.8	10	12
9.9 (Anstee 2019)	73.8	64.8-81.2	81.8	76.4-86.2	5	17.6	98.3	17	1
					10	31.1	96.6	16	3
					20	50.4	92.6	15	5
					30	63.5	87.9	13	8
					40	73.0	82.4	11	10
					50	80.2	75.8	9	13
10 (Wong 2019)	73.2	64.1-80.6	82.2	76.9-86.6	5	17.8	98.3	17	1
					10	31.4	96.5	16	3
					20	50.7	92.5	14	5
					30	63.8	87.7	12	8
					40	73.3	82.1	11	11
					50	80.5	75.4	9	13
11.4 (Anstee 2019)	62.9	52.8-72.1	87.4	82.9-90.9	5	20.8	97.8	12	2
					10	35.7	95.5	11	4
					20	55.6	90.4	10	7
					30	68.2	84.6	9	11
					40	76.9	78.0	8	15
					50	83.3	70.2	6	19
14.1 (Eddowes 2019)	40.6	29.9-52.2	93.8	90.5-96.0	5	25.7	96.8	6	3
					10	42.2	93.4	6	6
					20	62.1	86.3	5	12
					30	73.8	78.6	4	18
					40	81.4	70.3	4	24
					50	86.8	61.2	3	30
15 (Wong 2019)	33.5	23.3-45.5	95.2	92.2-97.1	5	26.7	96.5	5	3
					10	43.5	92.8	4	7
					20	63.4	85.1	4	13
					30	74.8	77.0	3	20
					40	82.2	68.2	3	27
					50	87.4	58.9	2	33

*Number of false positives and false negatives for 100 hypothetical cases

Table S8 Probe type (M or XL) as a covariate of the diagnostic performance of VCTE.

Fibrosis group	M probe	XL probe	χ^2	p-value
Any fibrosis (F>0)	8	2	2.930	0.231
Significant fibrosis (F>1)	21	6	2.138	0.343
Advanced fibrosis (F>2)	28	6	2.927	0.231
Cirrhosis (F=4)	11	6	1.091	0.579

Table S9 Study origin as a covariate of the diagnostic performance of VCTE.

Fibrosis	Europe	Asia	Australia	America	χ^2	p-value
Any fibrosis (F>0)	3	9	1	1	4.449	0.616
Significant fibrosis (F>1)	16	20	1	3	4.400	0.819
Advanced fibrosis (F>2)	22	21	2	2	5.266	0.729
Cirrhosis (F=4)	11	13	0	1	9.764	0.135

Table S10 Sensitivity analysis of the effect allowing less than 3 months' time between liver biopsy and VCTE.

	Interval between biopsy and VCTE <3 months			**All studies**		
	Se [%]	**Sp [%]**	**sAUC**	**Se [%]**	**Sp [%]**	**sAUC**
Any fibrosis (F>0)	77 (72-82)	71 (61-79)	0.81 (0.77-0.85)	78 (73-82)	72 (65-79)	0.82 (0.78-0.85)
Significant fibrosis (F>1)	77 (72-82)	72 (66-77)	0.81 (0.78-0.85)	80 (76-83)	73 (68-77)	0.83 (0.80-0.87)
Advanced fibrosis (F>2)	79 (75-82)	80 (77-82)	0.86 (0.84-0.88)	80 (77-83)	77 (74-80)	0.85 (0.83-0.87)
Cirrhosis (F=4)	72 (63-80)	89 (85-92)	0.89 (0.82-0.93)	76 (70-82)	88 (85-91)	0.88 (0.84-0.93)

Table S11 Subgroup analysis for the comparison of diagnostic performance of VCTE performed by M probe and XL probe.

	M probe			XL probe			
	Se [%]	**Sp [%]**	**sAUC**	**Se [%]**	**Sp [%]**	**sAUC**	**p**
Any fibrosis (F>0)	81 (76-86)	68 (57-78)	0.83 (0.78-0.89)	71 (49-86)	73 (29-94)	0.76 (0.64-0.84)	0.231
Significant fibrosis (F>1)	80 (72-86)	72 (64-79)	0.82 (0.78-0.86)	81 (72-87)	62 (51-72)	0.78 (0.67-0.87)	0.343
Advanced fibrosis (F>2)	80 (76-83)	78 (75-81)	0.86 (0.81-0.88)	74 (61-83)	79 (68-86)	0.83 (0.80-0.85)	0.231
Cirrhosis (F=4)	69 (59-77)	89 (85-93)	0.85 (0.71-0.93)	71 (54-84)	92 (88-94)	0.93 (0.78-0.95)	0.579

Table S12 Comparison of our results with previously published meta-analyses.

	Studies included (n)	Patients (n)		F0 vs F1-4	F0-1 vs F2-4	F0-2 vs F3-4	F0-3 vs F4
VCTE							
Present study	≥F1: 14	1064	sAUC	0.82	0.82	0.85	0.89
	≥F2: 37	2763	Se	78	80	80	76
	≥F3: 44	4219	Sp	72	73	77	88
	F4: 22	337					
Hsu 2019[24]	3	230	sAUC	0.82	0.87	0.84	0.84
			Se	66	76	77	80
			Sp	67	80	78	81
Jiang 2018[25]	≥F2: 10	1735	sAUC	-	0.85	0.92	0.94
	≥F3: 11		Se	-	77	79	90
	F4: 11		Sp	-	80	89	91
Xiao 2017[26†]	M ≥F2: 21	3165	sAUC	-	-	-	-
	M ≥F3: 22	3090	Se	-	68-92	69-89	78-97
	M F4: 20	2692	Sp	-	57-84	66-78	78-91
	XL ≥F2: 4	654	sAUC	-	-	-	-
	XL ≥F3: 3	579	Se	-	76	75	88
	XL F4: 4	654	Sp	-	65	75	82
Hashemi 2016[27]	≥F1: 3	199	sAUC	-	-	-	-
	≥F2: 6	567	Se	84	88	94	96
	≥F3: 7	698	Sp	78	78	91	92
	F4: 4	546					
Kwok 2014[12]	≥F2: 7	800	sAUC	-	-	-	-
	≥F3: 8	854	Se	-	79	85	92
	F4: 6	639	Sp	-	75	85	92
MRE							
Present study	≥F1: 6	391	sAUC	0.87	0.91	0.92	0.90
	≥F2: 6	209	Se	71	78	83	81
	≥F3: 10	214	Sp	85	89	89	90
	F4: 5	41					
Liang 2020[28]**	12	910	sAUC	0.89	0.93	0.93	0.95
			Se	77	87	89	94
			Sp	90	86	84	75
Hsu 2019[24]	3	230	sAUC	0.87	0.92	0.93	0.94
			Se	71	85	83	80
			Sp	73	85	83	86
Xiao 2017[26†]	≥F2: 3	≥F2: 384	sAUC	-			
	≥F3: 5	≥F3: 628	Se	-	73	86	87
	F4: 3	F4: 384	Sp	-	91	91	93
Singh 2016[29]	9	232	sAUC	0.86	0.87	0.90	0.91
			Se	75	79	83	88
			Sp	77	81	86	87
pSWE							
Present study	≥F1: 4	276	sAUC	0.77	0.86	0.91	0.90
	≥F2: 9	805	Se	64	69	80	76
	≥F3: 11	1209	Sp	76	85	88	88
	F4: 8	759					

Jiang 2018[25‡]	≥F2: 6 ≥F3: 9 F4: 7	982	sAUC	-	0.86	0.94	0.95
			Se	-	70	89	89
			Sp	-	84	88	91
Liu 2015[30‡]	7	723	sAUC	-	0.90	-	-
			Se	-	82	-	-
			Sp	-	85	-	-

†This study included some patients with viral hepatitis, and some children

‡This study included data from cohorts of patients that included some children (Osaki et al. 2010) and a duplicate cohort to Zhang et al. 2014 (Li et al. 2016)

*This study included data from cohorts of patients with mixed liver disease aetiologies, data from biopsies reported with histological scores other than the NASH CRN, and data from a cohort of patients that included some children (Osaki et al. 2010)

**This study included data from biopsies reported with histological scores other than NASH CRN (e.g. METAVIR)

sAUC – summary area under the curve, Se – sensitivity, Sp – specificity. Sensitivity and specificity are reported as %.

Supplementary figures

Study ID	Risk of bias: Patient selection	Risk of bias: Index test	Risk of bias: Reference test	Risk of bias: Flow and timing	Applicability concerns: Patient selection	Applicability concerns: Index test	Applicability concerns: Reference test
Agrawal 2017	+	−	+	−	+	+	+
AlJuboori 2018	?	?	+	?	+	+	+
Anstee 2019	+	?	+	?	+	+	+
Attia 2016	+	−	+	−	+	+	+
Aykut 2014	?	?	?	?	+	+	+
Boursier 2016	+	−	+	−	+	+	+
Boursier 2019	?	+	+	+	+	+	+
Cardoso 2019	+	−	+	+	+	+	+
Cassinotto 2013	+	−	+	−	+	+	+
Cassinotto 2016	+	−	+	−	+	+	+
Chan 2015	+	+	+	−	+	+	+
Chan 2017	−	−	+	−	+	+	+
Clet 2018	?	−	+	?	+	+	+
Das 2012	?	−	+	?	+	+	+
Eddowes 2018	?	?	+	−	+	+	+
Eddowes 2019	+	−	+	−	+	+	+
Ergelen 2016	+	−	+	−	+	+	+
Forlano 2017	?	−	?	?	+	+	+
Gaia 2011	+	−	+	?	+	+	+
Gaia 2015	−	−	+	−	+	+	+
Gallego-Durán 2016	?	+	+	+	+	+	+
Garg 2018	+	−	+	−	+	+	+
Hee 2014	?	?	?	?	+	+	+
Imajo 2016	?	−	?	−	+	+	+
Inadomi 2020	?	−	+	−	+	+	+
Kao 2020	+	−	+	+	+	+	+
Karlas 2015	?	?	+	−	+	+	+
Kumar 2013	+	−	+	−	+	+	+
Kwok 2016	+	+	+	−	+	+	+
Labenz 2018	+	+	?	−	+	+	+
Lee 2016	?	−	+	−	+	+	+
Lee 2017	+	−	+	−	+	+	+
Lee 2019	?	−	+	+	+	+	+
Lee 2020	?	−	+	+	+	+	+
Leong 2020	+	−	+	+	+	+	+
Loong 2017	+	+	+	−	+	+	+
Lupsor 2010	?	−	+	−	+	+	+
Mahadeva 2013	+	−	+	?	+	+	+
Morrison 2020	+	?	?	?	+	+	+
Myers 2012	?	−	+	−	+	+	+
Naveau 2014	+	−	+	−	+	+	+
Naveau 2017	+	−	+	+	+	+	+
Oeda 2020	+	−	+	−	+	+	+
Okajima 2017	+	−	?	−	+	+	+
Ooi 2018	+	?	+	+	+	+	+
Pais 2011	?	?	?	?	+	+	+
Park 2017	+	−	+	−	+	+	+
Pavlides 2017	+	−	+	−	+	+	+
Petta 2011	+	−	+	−	+	+	+
Petta 2015 Liv Int	+	+	+	−	+	+	+
Petta 2015 Hepatol	+	−	+	−	+	+	+
Petta 2017 APT	+	−	+	?	+	+	+
Petta 2017 Hepatol	+	+	+	+	+	+	+
Rosso 2016	?	−	+	−	+	+	+
Seki 2017	+	−	?	+	+	+	+
Shen 2015	+	−	+	−	+	+	+
Shima 2020	?	−	−	−	+	+	+
Staufer 2019	+	+	+	−	+	+	+
Tapper 2016	+	−	+	−	+	+	+
Wong 2010	+	−	+	−	+	+	+
Wong 2012	+	−	+	−	+	+	+
Wong 2018	+	+	+	−	+	+	+
Yoneda 2008	?	−	?	−	+	+	+
Younes 2018	?	+	+	+	+	+	+
Ziol 2009	+	−	?	+	+	+	+

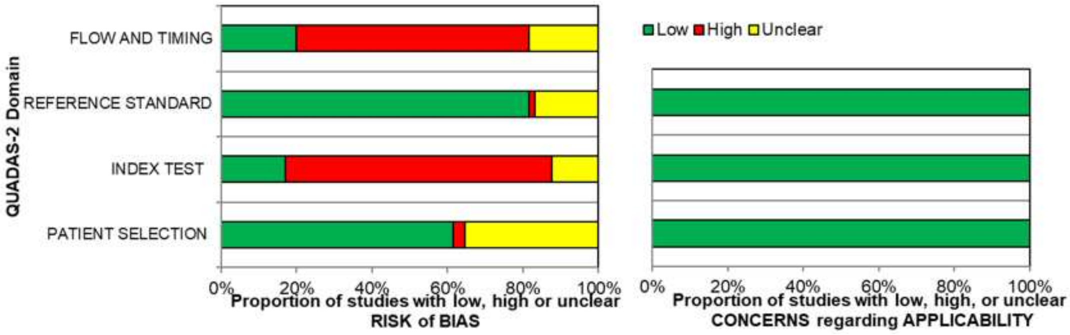

Figure S1 Methodological quality summary of VCTE studies. Red circles – high risk of bias, yellow circles – unclear risk of bias, green circles – low risk of bias

Study ID	Risk of bias: Patient selection	Risk of bias: Index test	Risk of bias: Reference test	Risk of bias: Flow and timing	Applicability concerns: Patient selection	Applicability concerns: Index test	Applicability concerns: Reference test
Chen 2011	?	−	+	+	+	+	+
Costa-Silva 2018	+	−	+	+	+	+	+
Cui 2015	+	+	+	+	+	+	+
Cui 2016	+	−	+	−	+	+	+
Imajo 2016	−	−	?	+	+	+	+
Kim 2013	−	−	?	−	+	+	+
Kim 2020	?	−	+	+	+	+	+
Lee 2020	?	−	+	+	+	+	+
Loomba 2013	?	−	?	?	+	+	+
Loomba 2014	+	−	+	?	+	+	+
Loomba 2016	+	−	+	+	+	+	+
Park 2017	+	−	+	+	+	+	+

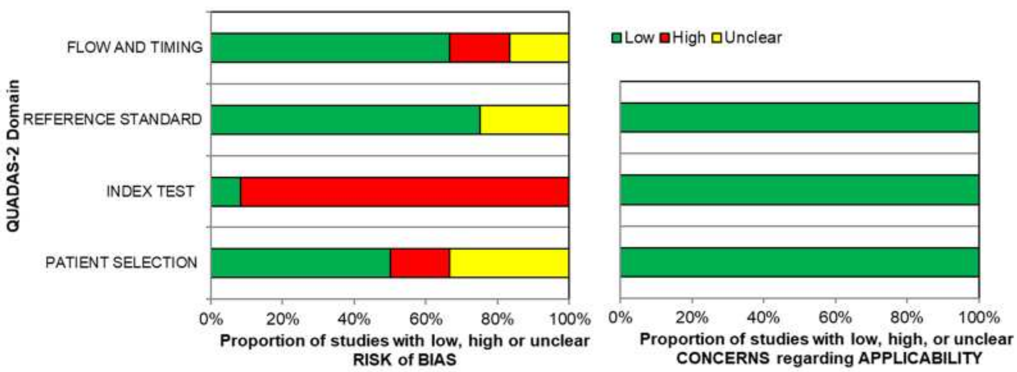

Figure S2 Methodological quality summary of MRE studies. Red circles – high risk of bias, yellow circles – unclear risk of bias, green circles – low risk of bias

Figure S3 Methodological quality summary of pSWE studies. Red circles – high risk of bias, yellow circles – unclear risk of bias, green circles – low risk of bias

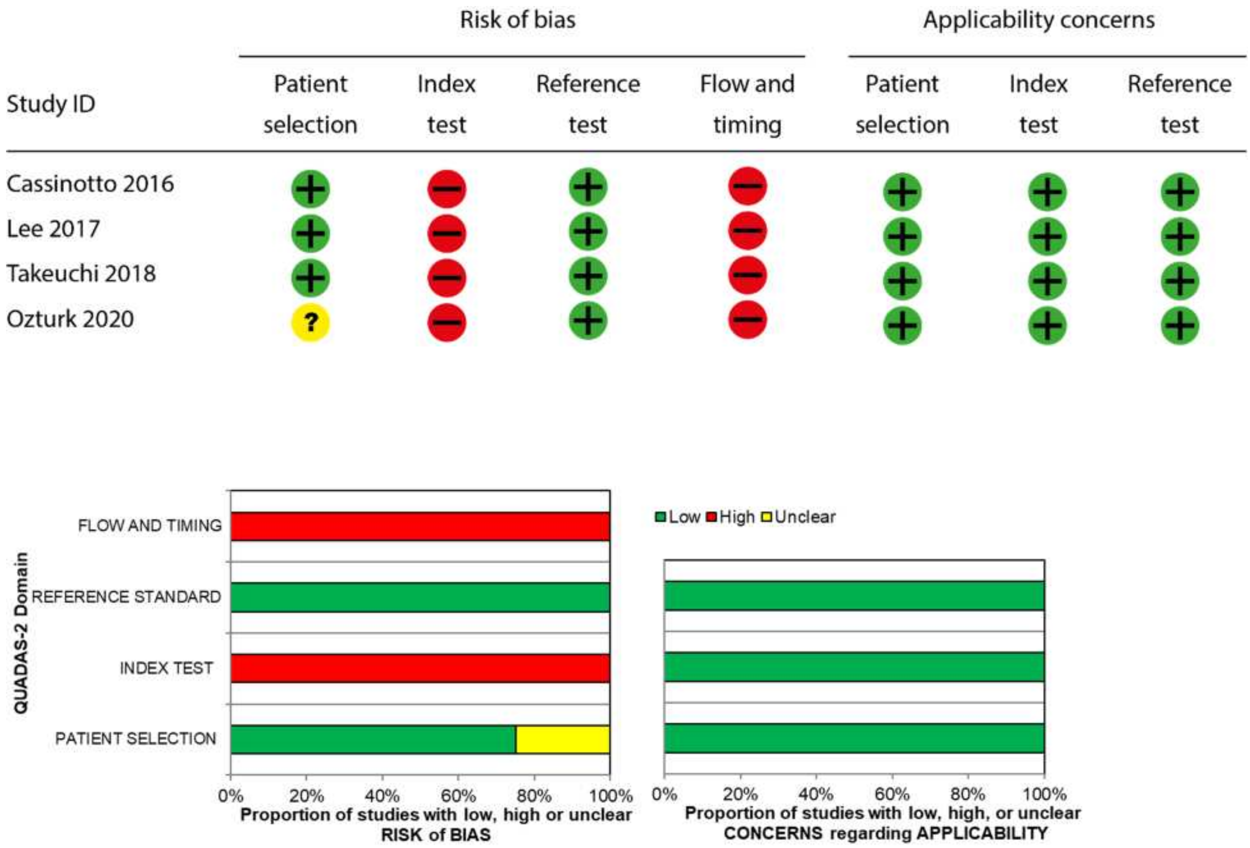

Figure S4 Methodological quality summary of 2DSWE studies. Red – high risk of bias, yellow – unclear risk of bias, green – low risk of bias

Study ID	Risk of bias: Patient selection	Risk of bias: Index test	Risk of bias: Reference test	Risk of bias: Flow and timing	Applicability concerns: Patient selection	Applicability concerns: Index test	Applicability concerns: Reference test
Eddowes 2018	?	+	+	−	+	+	+
Gallego-Durán 2016	?	−	+	+	+	+	+
Parente 2015	+	−	?	−	+	+	+
Pavlides 2017	+	−	+	−	+	+	+

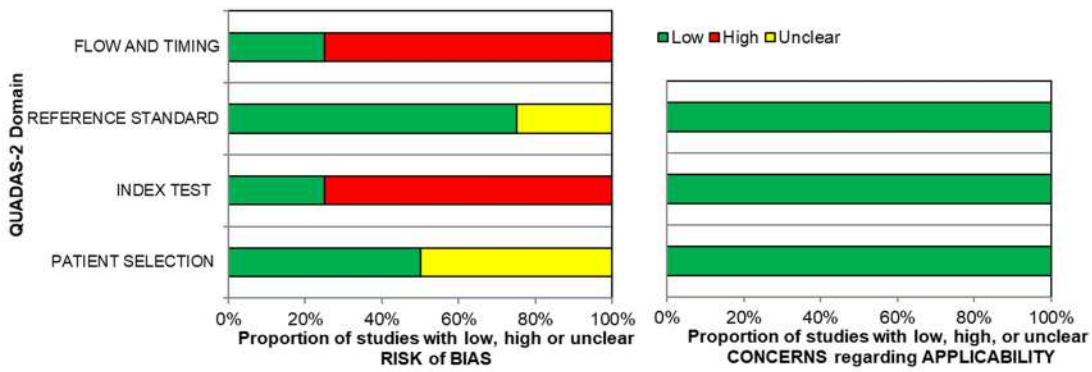

Figure S5 Methodological quality summary of MRI studies.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Cassinotto 2013 M	8.95	30	0	3	15	66.7 [51.0-80.0]	100.0 [31.0-100.0]
Cassinotto 2013 XL	7.00	28	0	3	18	60.9 [45.0-75.0]	100.0 [31.0-100.0]
Chan 2015	5.60	63	18	13	7	90.0 [79.0-95.0]	41.9 [25.0-61.0]
Gaia 2011	5.50	41	10	13	8	83.7 [70.0-92.0]	56.5 [35.0-76.0]
Garg 2018	6.00	48	7	9	12	80.0 [67.0-89.0]	56.3 [31.0-79.0]
Kumar 2013	6.10	69	10	22	19	78.0 [68.0-86.0]	68.0 [50.0-83.0]
Lee 2016	6.70	71	11	65	36	66.4 [57.5-75.2]	84.9 [77.9-92.0]
Leong 2020	7.68	70	3	13	14	83.3 [73.3-90.3]	81.3 [53.7-95.0]
Lupsor 2010	5.30	44	5	20	3	93.5 [82.1-98.6]	78.3 [56.3-92.5]
Okajima 2017	8.20	72	13	49	29	71.6 [61.2-79.6]	79.0 [66.5-87.9]
Ooi 2018	7.00	12	21	30	3	78.6 [51.4-94.7]	59.6 [44.2-72.1]
Park 2017	6.10	34	15	28	17	66.7 [52.0-78.9]	65.1 [49.0-78.5]
Seki 2017	7.20	86	13	49	23	78.8 [69.8-85.9]	78.3 [66.4-87.9]
Shima 2020	7.20	124	17	74	34	78.5 [71.1-84.5]	81.3 [71.5-88.4]
Yoneda 2008	5.90	68	2	16	11	86.1 [76.0-92.5]	88.9 [63.9-98.1]

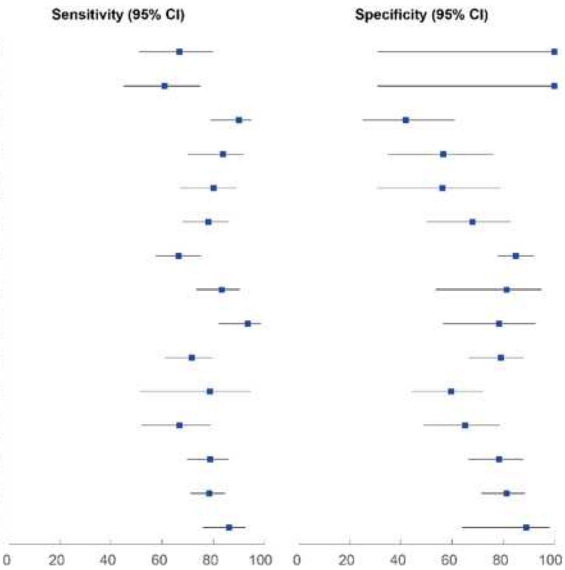

Figure S6 Forest plots of all included studies for the diagnosis of any fibrosis (≥F1) using VCTE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Costa-Silva 2018	3.14	22	2	6	19	78.6 [59.0-91.7]	90.5 [69.6-98.8]
Cui 2016	2.99	42	5	30	48	58.3 [46.0-70.0]	90.6 [79.0-96.0]
Imajo 2016	2.50	96	2	32	12	75.0 [66.0-82.0]	85.7 [56.0-97.0]
Kim 2020	2.58	32	0	1	14	97.0 [82.5-99.8]	100.0 [73.2-100.0]
Loomba 2014	3.02	41	4	33	39	55.4 [43.0-67.0]	90.7 [77.0-97.0]
Park 2017	2.65	43	10	13	37	76.5 [63.0-87.0]	79.1 [64.0-89.0]

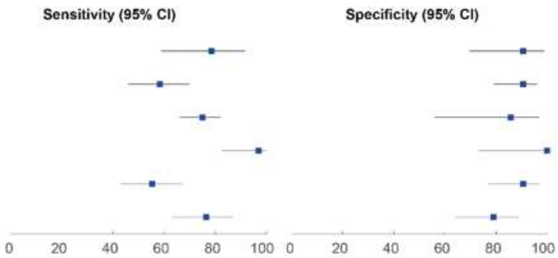

Figure S7 Forest plots of all included studies for the diagnosis of any fibrosis (≥F1) using MRE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Cassinotto 2013	1.81	28	0	3	29	49.1 [36.0-63.0]	100.0 [31.0-100.0]
Cui 2016	1.29	39	12	41	33	54.2 [42.1-65.8]	77.4 [63.5-87.3]
Fierbinteanu 2013	1.11	33	6	15	10	76.7 [61.0-87.7]	71.4 [47.7-87.8]
Medellin 2019	1.11	17	0	6	4	80.9 [57.4-93.7]	100.0 [51.7-100.0]

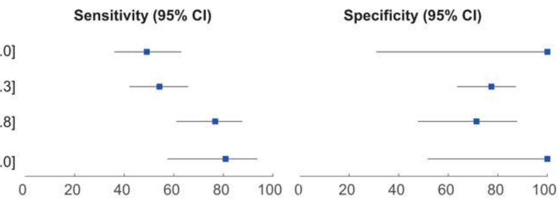

Figure S8 Forest plots of all included studies for the diagnosis of any fibrosis (≥F1) using pSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Attia 2016	6.70	10	3	2	11	87.0 [51.0-97.0]	76.0 [48.8-94.3]
Aykut 2014	7.90	33	3	11	41	75.0 [59.0-86.0]	93.1 [80.0-98.0]
Cardoso 2019 M	8.20	14	24	1	42	93.3 [68.1-99.8]	63.6 [50.9-75.1]
Cardoso 2019 XL	8.20	11	15	4	51	73.3 [44.9-92.2]	77.2 [65.3-86.7]
Cassinotto 2013 M	8.10	35	1	2	10	94.6 [80.0-99.0]	90.9 [57.0-100.0]
Cassinotto 2013 XL	7.20	25	1	10	13	71.4 [53.0-85.0]	92.9 [64.0-100.0]
Cassinotto 2016	9.80	93	7	63	60	59.6 [51.5-67.3]	89.6 [79.1-5.3]
Chan 2015	6.65	25	42	0	34	100.0 [83.0-100.0]	44.7 [33.0-57.0]
Eddowes 2019	8.20	159	45	66	103	70.7 [64.0-77.0]	69.6 [62.0-77.0]
Gaia 2011	7.00	25	8	8	31	75.8 [57.0-88.0]	79.5 [63.0-90.0]
Garg 2018	7.25	21	19	9	27	70.0 [50.0-85.0]	58.7 [43.0-73.0]
Inadomi 2020	8.95	91	24	34	75	72.8 [64.0-80.2]	75.8 [65.9-83.6]
Kao 2020	7.00	23	8	6	36	79.3 [59.7-91.3]	81.4 [66.8-91.3]
Karlas 2015	7.60	1	4	0	15	100.0 [5.0-100.0]	78.9 [54.0-93.0]
Kumar 2013	7.00	42	15	12	51	78.0 [64.0-88.0]	77.0 [65.0-86.0]
Lee 2017	7.40	29	4	17	44	62.5 [47.5-76.4]	91.7 [79.1-97.3]
Lee 2016	8.00	38	21	8	116	82.6 [71.7-93.6]	84.7 [78.3-91.0]
Lee 2019	8.95	58	36	22	68	72.5 [61.2-81.6]	65.4 [55.3-74.3]
Leong 2020	9.13	36	20	5	39	87.8 [73.0-95.4]	66.1 [52.5-77.6]
Loong 2017	9.00	45	18	24	128	65.2 [53.5-75.4]	87.7 [81.4-92.1]
Lupsor 2010	6.80	12	8	6	46	66.7 [41.1-86.8]	84.3 [71.4-93.0]
Mahadeva 2013	6.75	50	23	25	33	66.2 [54.7-76.9]	59.6 [45.0-71.6]
Naveau 2014	7.60	16	17	6	61	72.7 [49.6-88.3]	78.2 [67.1-86.4]
Oeda 2020 tr M	7.00	45	21	5	25	90.0 [77.4-96.3]	54.3 [39.2-68.8]
Oeda 2020 tr XL	6.70	41	14	9	32	82.0 [68.1-91.0]	69.6 [54.1-81.8]
Oeda 2020 val M	7.00	52	21	4	26	92.9 [81.9-97.7]	55.3 [40.2-69.5]
Oeda 2020 val XL	6.70	51	17	5	30	91.1 [79.6-96.7]	63.8 [48.5-76.9]
Okajima 2017	9.30	40	28	5	90	88.9 [75.2-95.8]	76.3 [67.4-83.4]
Ooi 2018	9.00	4	16	0	46	100.0 [39.6-100.0]	74.2 [61.3-84.1]
Park 2017	6.90	23	10	6	55	79.3 [59.7-91.3]	84.6 [73.1-92.0]
Pavlides 2017	3.80	24	10	2	2	92.3 [73.4-98.7]	16.7 [2.9-49.1]
Petta 2015 Hep	6.90	80	44	33	96	70.8 [61.4-78.8]	68.6 [60.1-76.0]
Petta 2011	7.25	47	23	21	55	69.0 [56.6-79.5]	70.0 [59.0-80.0]
Petta 2017 Hep	8.50	139	36	48	101	74.3 [67.3-80.3]	73.7 [65.4-80.7]
Rosso 2016	6.80	44	8	18	35	71.0 [57.9-81.4]	81.0 [66.1-91.1]
Seki 2017	8.60	51	11	13	106	79.7 [67.4-88.3]	90.6 [83.4-95.0]
Shima 2020	9.10	81	15	20	133	80.2 [70.8-87.2]	89.9 [83.6-94.0]
Wong 2018 M	5.00	222	121	4	44	98.2 [95.2-99.4]	26.7 [20.2-34.2]
Wong 2018 XL	5.00	234	107	23	69	91.1 [86.7-94.1]	39.2 [32.0-46.9]
Wong 2012 M	7.00	52	32	14	58	79.0 [69.0-89.0]	64.0 [54.0-74.0]
Wong 2012 XL	6.20	60	35	22	67	73.0 [64.0-83.0]	66.0 [57.0-76.0]
Wong 2010	7.00	80	35	21	110	79.2 [69.8-86.4]	75.9 [67.9-82.4]
Yoneda 2008	6.65	45	12	6	34	88.2 [75.4-95.1]	73.9 [58.6-85.2]
Ziol 2009	10.20	4	0	6	3	40.0 [14.0-73.0]	100.0 [31.0-100.0]

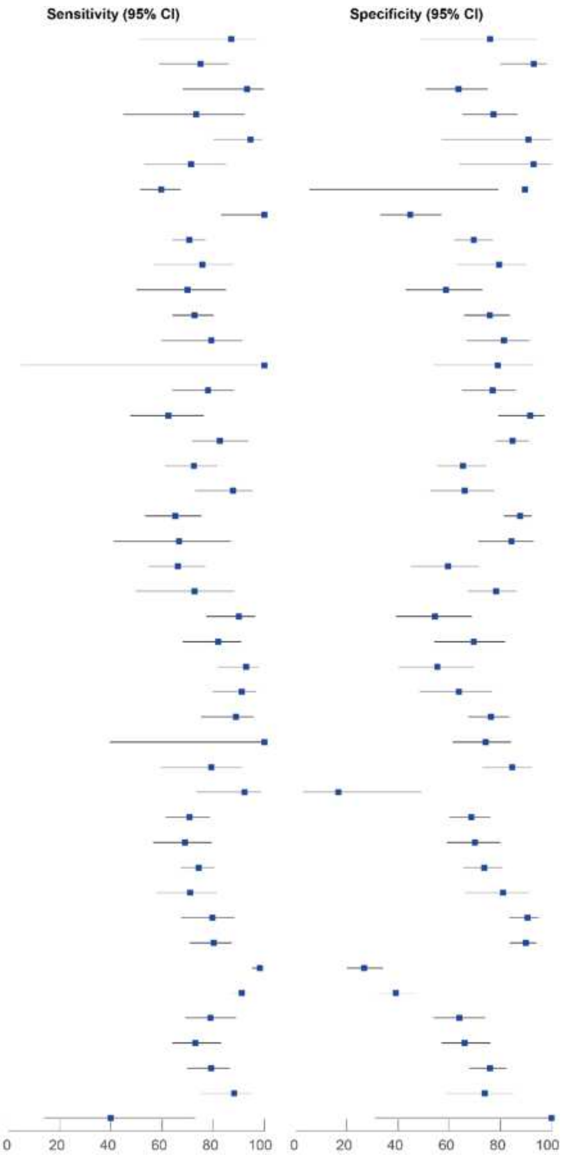

Figure S9 Forest plots of all included studies for the diagnosis of significant fibrosis (≥F2) using VCTE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Costa-Silva 2018	4.14	11	1	1	36	91.7 [61.5-99.8]	97.3 [85.8-99.9]
Cui 2016	3.62	22	4	11	88	66.7 [48.0-81.0]	95.7 [89.0-99.0]
Imajo 2016	3.40	67	10	10	55	87.3 [77.0-93.0]	85.0 [73.0-92.0]
Kim 2020	3.13	19	0	1	27	95.0 [73.1-99.7]	100.0 [84.5-100.0]
Loomba 2014	3.58	23	7	12	75	65.7 [48.0-80.0]	91.5 [83.0-96.0]
Park 2017	2.86	25	13	7	58	79.3 [60.0-90.0]	81.8 [70.0-90.0]

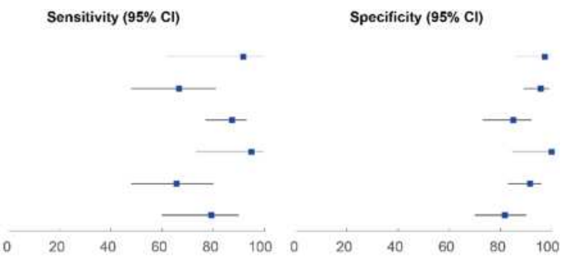

Figure S10 Forest plots of all included studies for the diagnosis of significant fibrosis (≥F2) using MRE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Attia 2016	1.18	35	6	10	46	78.0 [62.5-88.3]	88.0 [75.9-95.2]
Cassinotto 2016	1.32	90	7	70	69	56.3 [48.2-64.0]	90.8 [81.4-95.9]
Cassinotto 2013	1.81	27	1	17	15	61.4 [46.0-75.0]	93.8 [68.0-100.0]
Cui 2016	1.34	27	20	6	72	81.8 [63.9-92.4]	78.3 [68.2-85.9]
Fierbinteanu 2013	1.17	28	3	5	28	84.8 [67.3-97.5]	90.3 [73.1-97.5]
Karlas 2015	1.25	8	7	0	31	100.0 [59.8-100.0]	82.5 [65.1-91.7]
Lee 2017	1.29	18	5	19	41	48.7 [32.2-65.3]	90.0 [75.6-95.9]
Medellin 2019	1.34	12	2	4	9	75.0 [47.4-91.7]	81.8 [47.8-96.8]
Zhang 2013	1.31	19	7	12	29	61.3 [42.0-78.0]	80.6 [63.0-91.0]

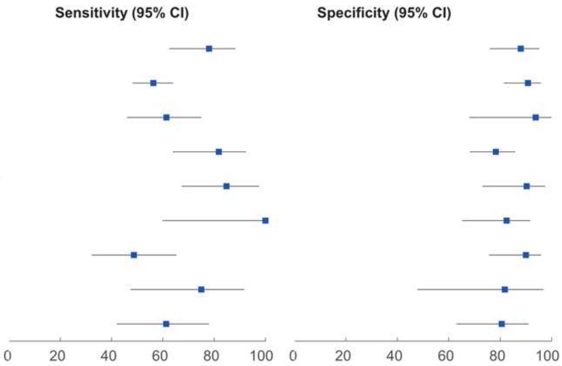

Figure S11 Forest plots of all included studies for the diagnosis of significant fibrosis (≥F2) using pSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Cassinotto 2016	8.70	116	7	48	61	70.7 [63.0-77.4]	89.7 [79.3-95.4]
Lee 2017	8.30	27	17	4	21	87.1 [69.2-95.8]	55.3 [38.5-71.0]
Ozturk 2020	8.40	23	29	7	57	76.7 [57.3-89.4]	66.3 [55.2-75.9]
Takeuchi 2018	11.57	24	14	22	11	52.2 [37.1-68.9]	44.0 [25.0-64.7]

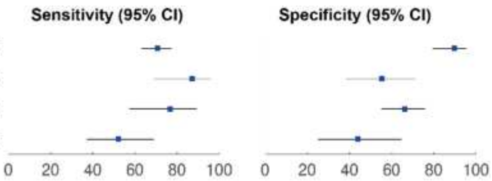

Figure S12 Forest plots of all included studies for the diagnosis of significant fibrosis (≥F2) using 2DSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Agrawal 2017	9.20	5	1	2	17	71.4 [30.0-95.0]	94.4 [71.0-100.0]
Anstee 2019	9.90	1099	108	224	176	83.1 [80.9-85.0]	62.0 [56.0-67.6]
Attia 2016	9.30	9	3	1	13	91.0 [54.1-99.5]	80.0 [53.7-95.0]
Boursier 2016	8.70	152	104	20	176	88.4 [82.0-93.0]	62.9 [57.0-68.0]
Boursier 2019 D	8.00	233	137	26	229	90.0 [85.5-93.2]	62.6 [57.4-67.5]
Boursier 2019 V	8.00	112	20	12	113	90.3 [83.4-94.7]	85.0 [77.5-90.4]
Cassinotto 2013 M	11.75	20	3	6	19	76.9 [56.0-90.0]	86.4 [64.0-96.0]
Cassinotto 2013 XL	10.15	16	4	6	23	72.7 [50.0-88.0]	85.2 [65.0-95.0]
Cassinotto 2016	8.20	83	51	9	80	90.2 [81.8-95.2]	61.1 [52.1-69.3]
Chan 2015	8.00	18	28	1	54	94.7 [72.0-100.0]	65.9 [54.0-76.0]
Eddowes 2019	9.70	99	59	41	174	70.7 [62.0-78.0]	74.7 [69.0-80.0]
Ergelen 2016	9.80	29	3	3	28	90.6 [74.0-98.0]	90.3 [73.0-97.0]
Gaia 2011	8.00	11	11	6	44	64.7 [39.0-85.0]	80.0 [67.0-89.0]
Garg 2018	12.50	7	8	4	57	63.6 [32.0-88.0]	87.7 [77.0-94.0]
Hee 2014	>9.6	17	40	0	41	100.0 [77.0-100.0]	50.6 [39.0-62.0]
Inadomi 2020	11.45	57	28	24	115	70.4 [59.0-79.7]	80.4 [72.8-86.4]
Kumar 2013	9.00	23	11	4	82	85.0 [65.0-95.0]	88.0 [79.0-93.0]
Kwok 2016	9.60	44	31	3	16	94.0 [81.0-98.0]	34.0 [21.0-49.0]
Labenz 2018	10.00	16	21	5	84	76.0 [52.0-91.0]	80.0 [71.0-87.0]
Lee 2016	9.00	27	22	1	133	96.4 [89.6-100.0]	85.8 [79.9-91.7]
Lee 2017	8.00	22	10	5	57	81.5 [61.3-93.0]	85.1 [77.0-93.3]
Leong 2020	9.28	30	24	3	43	90.9 [74.5-97.6]	64.2 [51.5-75.3]
Loong 2017	9.60	36	23	7	149	83.7 [70.0-91.9]	86.6 [80.7-90.9]
Lupsor 2010	10.40	5	2	0	65	100.0 [48.0-100.0]	96.9 [89.1-99.5]
Mahadeva 2013	7.10	20	34	9	68	70.4 [49.0-84.0]	66.6 [56.5-75.5]
Naveau 2014	7.60	9	24	0	67	100.0 [62.9-100.0]	73.6 [63.2-82.0]
Oeda 2020 tr M	10.80	22	14	7	53	75.9 [56.1-89.0]	79.1 [67.1-87.7]
Oeda 2020 tr XL	8.20	24	20	5	47	82.8 [63.5-93.4]	70.1 [57.6-80.4]
Oeda 2020 val M	10.80	20	23	13	47	60.6 [42.2-76.6]	67.1 [54.8-77.6]
Oeda 2020 val XL	8.20	27	26	6	44	81.8 [63.9-92.4]	62.9 [50.4-73.9]
Okajima 2017	10.00	21	39	1	102	95.5 [75.1-99.8]	72.3 [64.1-79.4]
Ooi 2018	10.30	2	12	1	51	66.7 [12.5-98.2]	81.0 [68.7-89.4]
Park 2017	7.30	14	17	4	59	77.8 [51.9-92.6]	77.6 [66.4-86.1]
Pavlides 2017	6.80	14	8	1	15	93.0 [66.0-99.7]	65.0 [42.8-82.8]
Petta 2011	8.75	25	25	8	88	76.0 [57.4-88.3]	78.0 [68.9-84.9]
Petta 2017 Hep	10.10	89	45	26	164	77.6 [68.5-84.4]	78.4 [72.2-83.7]
Petta 2015 Hep	8.40	41	48	12	152	77.4 [63.5-87.2]	76.0 [69.4-81.6]
Petta 2015 S	7.9-9.6	29	23	5	101	85.3 [68.2-94.5]	81.4 [73.3-87.6]
Petta 2015 T	7.9-9.6	17	13	8	83	68.0 [46.4-84.3]	86.4 [77.6-92.3]
Petta 2017 APT	9.60	174	100	61	426	74.0 [67.9-79.4]	81.0 [77.3-84.2]
Rosso 2016	6.80	32	24	6	43	84.2 [68.1-93.4]	64.0 [51.5-75.3]
Seki 2017	10.00	38	16	4	113	90.5 [76.5-96.9]	87.6 [80.4-92.5]
Shen 2015 SCD<25	8.70	5	14	2	46	71.0 [60.0-95.0]	77.0 [58.0-82.0]
Shen 2015 SCD>25	12.90	5	3	1	25	83.0 [37.0-99.0]	89.0 [71.0-97.0]
Shima 2020	9.90	56	31	9	153	86.2 [74.8-93.1]	83.2 [76.8-88.1]
Tapper 2016	9.90	20	23	1	76	95.0 [74.1-99.8]	77.0 [67.0-84.4]
Wong 2010	8.70	47	32	9	158	83.9 [71.1-91.9]	83.2 [76.9-88.0]
Wong 2012 M	8.70	35	25	7	89	83.0 [72.0-95.0]	78.0 [70.0-86.0]
Wong 2012 XL	7.20	49	60	5	70	91.0 [83.0-99.0]	54.0 [45.0-62.0]
Wong 2018 M	10.00	109	48	34	200	76.2 [68.2-82.8]	80.6 [75.1-85.3]
Wong 2018 XL	10.00	85	25	76	247	52.8 [44.8-60.7]	90.8 [86.6-93.8]
Yoneda 2008	9.80	23	13	4	57	85.2 [65.3-95.1]	81.4 [70.0-89.4]
Younes 2018	9.60	47	35	26	184	64.4 [52.2-75.0]	84.0 [78.3-88.5]
Ziol 2009	11.75	2	1	1	9	66.7 [13.0-98.0]	90.0 [54.0-99.0]

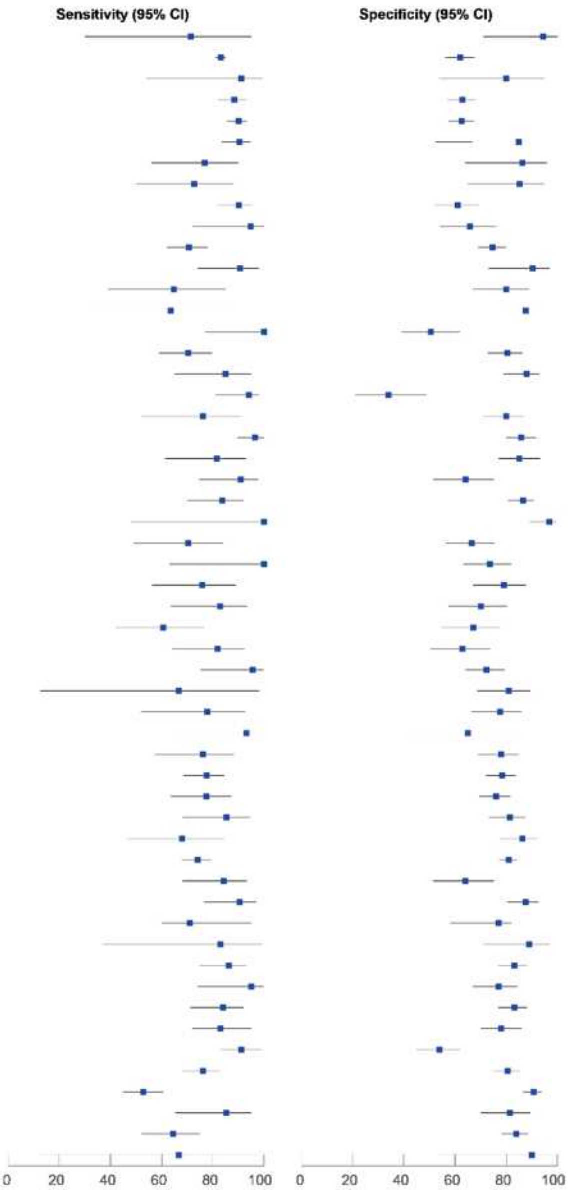

Figure S13 Forest plots of all included studies for the diagnosis of advanced fibrosis (≥F3) using VCTE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Costa-Silva 2018	4.39	10	1	1	37	90.9 [58.7-99.8]	97.3 [86.2-99.9]
Cui 2015	3.64	17	8	1	75	92.2 [71.0-100.0]	90.4 [81.0-95.0]
Cui 2016	3.62	19	7	2	97	90.5 [68.0-98.0]	93.3 [86.0-97.0]
Imajo 2016	4.80	34	13	11	84	74.5 [60.0-87.0]	86.9 [78.0-92.0]
Kim 2013	4.15	39	7	7	89	85.0 [71.0-93.0]	92.9 [85.0-97.0]
Kim 2020	4.34	8	3	0	36	100.0 [59.8-100.0]	92.3 [78.0-98.0]
Loomba 2014	3.64	19	9	3	86	86.4 [64.0-96.0]	90.5 [82.0-95.0]
Loomba 2013	3.20	6	8	0	38	100.0 [54.0-100.0]	83.0 [69.0-92.0]
Park 2017	2.99	16	16	5	66	77.8 [52.0-91.0]	80.3 [70.0-88.0]

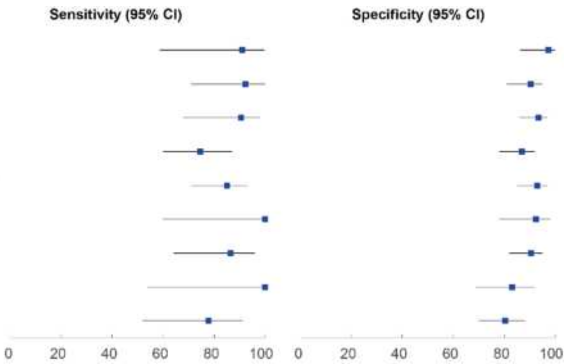

Figure S14 Forest plots of all included studies for the diagnosis of advanced fibrosis (≥F3) using MRE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Attia 2016	1.47	25	2	2	68	92.6 [74.2-98.7]	97.0 [89.1-99.5]
Cassinotto 2013	2.03	20	6	11	23	64.5 [45.0-80.0]	79.3 [60.0-91.0]
Cassinotto 2016	1.53	55	15	38	128	59.1 [48.4-69.1]	89.5 [83.0-93.8]
Cui 2016	1.34	20	27	1	77	95.2 [74.1-99.8]	74.0 [64.4-81.9]
Fierbinteanu 2013	1.48	19	2	3	40	86.4 [64.0-96.4]	95.2 [82.6-99.2]
Joo 2018	1.40	39	22	16	238	70.9 [56.9-82.0]	91.5 [87.3-94.5]
Lee 2017	1.36	18	6	2	57	90.5 [66.9-98.2]	90.0 [79.8-96.1]
Medellin 2019	1.55	9	4	1	13	90.0 [54.1-99.5]	76.5 [49.8-92.2]
Palmeri 2011	4.24	36	9	4	86	90.0 [75.4-96.7]	90.0 [82.3-95.3]
Zhang 2013	1.36	11	9	5	42	68.8 [41.0-88.0]	82.4 [69.0-91.0]

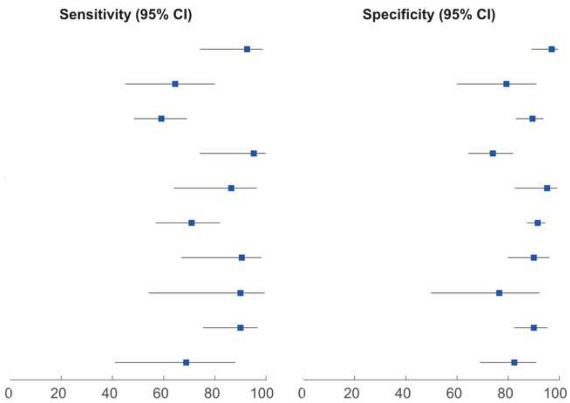

Figure S15 Forest plots of all included studies for the diagnosis of advanced fibrosis (≥F3) using pSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Cassinotto 2016	10.70	71	13	29	119	71.0 [60.9-79.4]	90.2 [83.4-94.4]
Lee 2017	10.70	18	19	2	30	90.0 [66.9-98.2]	61.2 [46.2-74.5]
Ozturk 2020	9.30	16	29	3	68	84.2 [59.5-95.8]	70.1 [59.8-78.8]
Takeuchi 2018	13.07	20	17	12	22	62.5 [43.7-78.3]	56.4 [39.8-71.8]

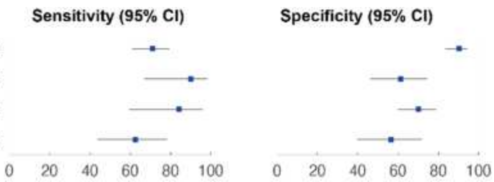

Figure S16 Forest plots of all included studies for the diagnosis of advanced fibrosis (≥F3) using 2DSWE.

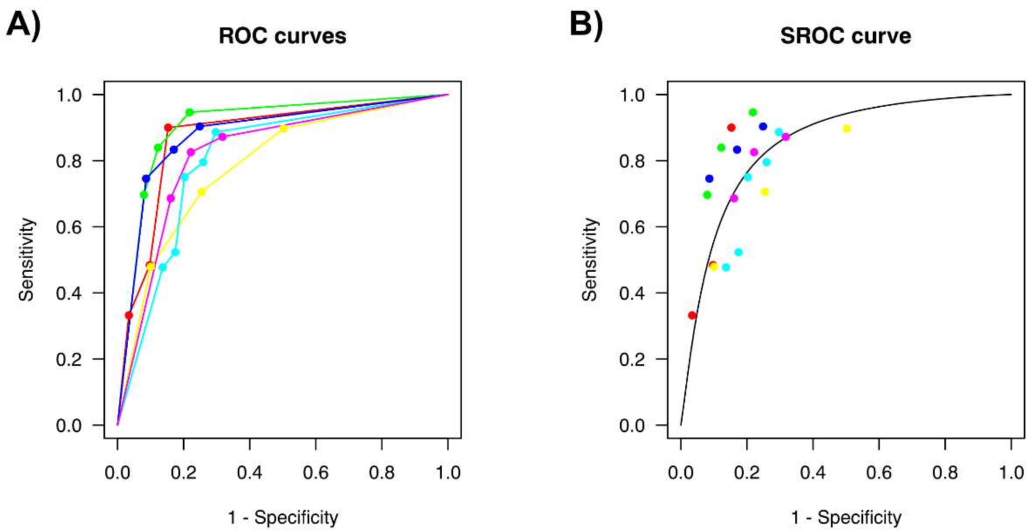

Figure S17 The performance of VCTE in detecting advanced fibrosis (≥F3) (A) Multiple-threshold ROC curves and (B) multiple-threshold sROC curve based on the multiple-thresholds model using homogenized thresholds. Circles represent information on sensitivity and specificity. AUC: 0.85 (0.80, 0.89). Max Youden-index results: cut-off: 8.7 kPa, sensitivity: 81% (73%, 87%), specificity: 76% (70%, 82%).

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Agrawal 2017	35.30	1	0	0	24	100.0 [5.0-100.0]	100.0 [83.0-100.0]
Attia 2016	11.70	8	3	0	15	100.0 [57.8-100.0]	83.0 [57.7-95.6]
Aykut 2014		9	19	0	60	100.0 [63.0-100.0]	76.3 [65.0-85.0]
Cassinotto 2013 M	18.60	12	3	2	31	85.7 [56.0-97.0]	91.2 [75.0-98.0]
Cassinotto 2013 XL	12.65	11	5	1	32	91.7 [60.0-100.0]	86.5 [70.0-95.0]
Cassinotto 2016	16.10	24	18	13	168	64.9 [47.4-79.3]	90.3 [84.9-94.0]
Chan 2015	17.00	3	6	0	92	100.0 [31.0-100.0]	93.9 [87.0-97.0]
Chan 2017	15.10	3	3	0	73	100.0 [31.0-100.0]	96.1 [88.0-99.0]
Eddowes 2019	13.60	29	72	5	267	85.3 [69.0-95.0]	78.8 [74.0-83.0]
Gaia 2011	10.50	7	3	2	60	77.8 [40.0-96.0]	95.2 [86.0-99.0]
Kumar 2013	19.40	8	2	2	108	80.0 [44.0-96.0]	98.0 [93.0-100.0]
Lee 2017	11.70	10	8	4	72	71.4 [42.0-90.4]	90.0 [80.7-95.3]
Lee 2016	11.00	17	17	0	149	100.0 [77.1-100.0]	89.8 [84.9-94.6]
Leong 2020	13.45	4	23	0	73	100.0 [39.6-100.0]	76.0 [66.0-83.9]
Mahadeva 2013	11.95	6	11	2	112	75.0 [35.6-95.5]	91.2 [84.2-95.2]
Oeda 2020 tr M	16.80	5	9	0	82	100.0 [46.3-100.0]	90.1 [81.6-95.1]
Oeda 2020 tr XL	14.30	5	8	0	83	100.0 [46.3-100.0]	91.2 [82.9-95.9]
Oeda 2020 val M	16.80	3	16	1	83	75.0 [21.9-98.7]	83.8 [74.8-90.2]
Oeda 2020 val XL	14.30	3	10	1	89	75.0 [21.9-98.7]	89.9 [81.8-94.8]
Park 2017	6.90	5	29	3	57	62.5 [25.9-89.8]	66.3 [55.2-75.9]
Pavlides 2017	14.70	4	3	0	31	100.0 [39.6-100]	91.0 [75.2-97.7]
Shima 2020	10.30	12	69	0	168	100.0 [69.9-100.0]	70.9 [64.6-76.5]
Wong 2018 M	15.00	32	25	27	307	54.2 [40.8-67.1]	92.5 [88.9-95.0]
Wong 2018 XL	15.00	33	17	37	346	47.1 [35.2-59.4]	95.3 [92.5-97.2]
Wong 2012 M	10.30	13	24	3	116	81.0 [62.0-100.0]	83.0 [77.0-89.0]
Wong 2012 XL	7.90	21	38	3	122	88.0 [74.0-100.0]	76.0 [69.0-82.0]
Wong 2010	11.50	19	20	6	201	76.0 [54.5-89.8]	91.0 [86.2-94.2]
Yoneda 2008	17.50	9	3	0	85	100.0 [62.9-100.0]	96.6 [90.0-99.1]

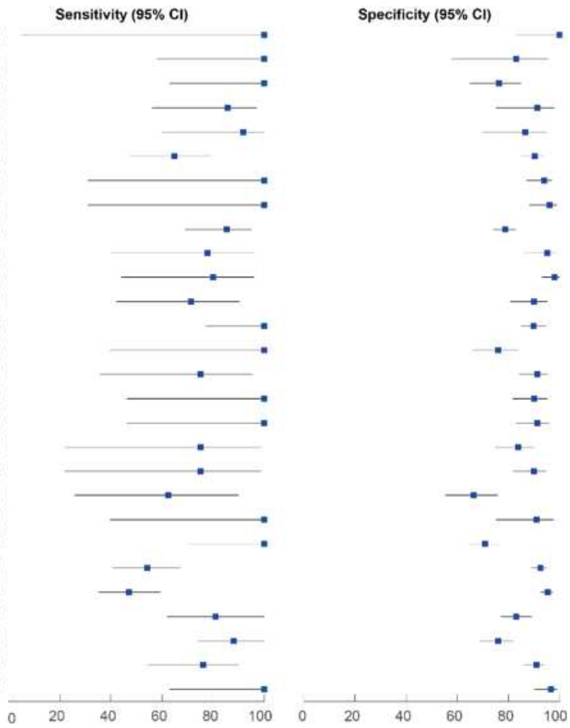

Figure S18 Forest plot of all included studies for the diagnosis of cirrhosis (F4) using VCTE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Costa-Silva 2018	5.37	3	5	0	41	100.0 [29.2-100.0]	89.1 [76.4-96.4]
Cui 2016	4.15	8	10	1	106	88.9 [57.0-99.0]	91.4 [86.0-97.0]
Imajo 2016	6.70	10	7	1	124	90.9 [57.0-100.0]	94.5 [89.0-98.0]
Loomba 2014	4.67	8	6	2	101	80.0 [44.0-96.0]	94.4 [88.0-98.0]
Park 2017	3.35	6	18	2	77	75.0 [36.0-96.0]	81.4 [71.0-88.0]

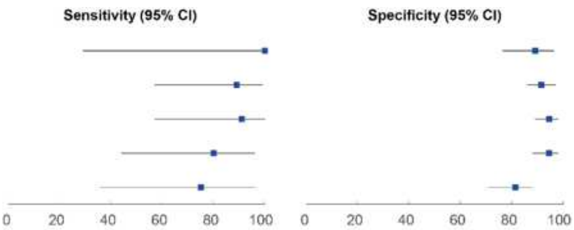

Figure S19 Forest plot of all included studies for the diagnosis of cirrhosis (F4) using MRE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Attia 2016	1.89	15	5	75	2	86.0 [62.3-97.9]	94.0 [85.4-97.7]
Cassinotto 2013	2.54	13	3	38	6	68.4 [43.0-86.0]	92.7 [79.0-98.0]
Cassinotto 2016	2.04	17	20	177	22	43.6 [28.2-60.2]	89.8 [84.5-93.5]
Cui 2016	2.48	7	8	108	2	77.8 [40.2-96.1]	93.1 [86.4-96.8]
Fierbinteanu 2013	1.64	11	4	48	1	91.7 [59.8-99.6]	92.3 [80.3-97.5]
Lee 2017	1.50	6	7	68	2	75.0 [35.6-95.5]	90.7 [81.1-95.8]
Medellin 2019	1.80	5	7	15	0	100.0 [46.3-100.0]	68.2 [45.1-85.3]
Zhang 2013	1.36	4	16	47	0	100.0 [40.0-100.0]	74.6 [62.0-84.0]

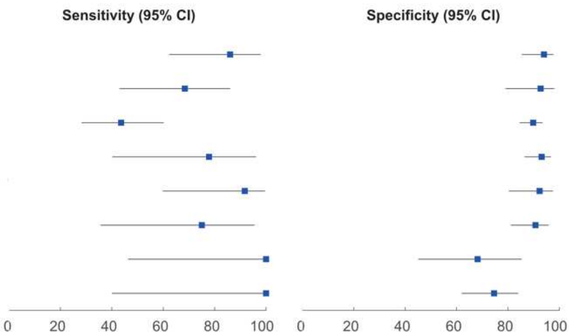

Figure S20 Forest plot of all included studies for the diagnosis of cirrhosis (F4) using pSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Cassinotto 2016	14.40	22	20	174	16	57.9 [40.9-73.3]	89.7 [84.3-93.4]
Lee 2017	15.10	9	13	46	1	90.0 [54.1-99.5]	78.0 [64.9-87.3]
Takeuchi 2018	15.70	5	12	54	0	100.0 [46.3-100.0]	81.8 [70.0-89.9]

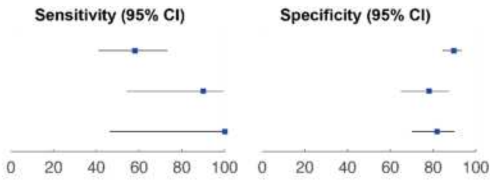

Figure S21 Forest plot of all included studies for the diagnosis of cirrhosis (F4) using 2DSWE.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Chen 2011	2.90	30	4	6	18	83.0 [66.0-93.0]	82.0 [59.0-94.0]
Costa-Silva 2018	3.24	18	3	7	21	72.0 [48.2-85.7]	87.5 [66.4-97.2]
Loomba 2014	3.26	38	2	53	24	42.0 [32.0-53.0]	92.3 [73.0-99.0]
Park 2017	2.53	46	7	26	15	63.9 [52.0-75.0]	68.2 [45.0-85.0]

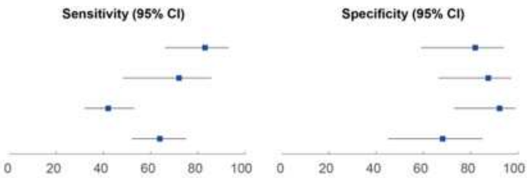

Figure S22 Forest plot of MRE studies for the diagnosis of NASH.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Parente 2015[1]	0.73	9	14	7	29	56.3 [30.6, 79.2]	67.4 [51.4, 80.5]
Parente 2015[2]	37.75	10	13	6	30	62.5 [35.9, 83.7]	69.8 [53.7, 82.3]
Parente 2015[3]	34.83	9	12	7	31	56.3 [30.6, 79.2]	72.1 [56.1, 84.2]

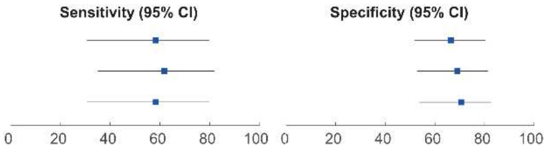

Figure S23 Forest plots of all included studies for the diagnosis of any fibrosis (≥F1) using MRI methods. ([1]pure molecular diffusion, [2]vascular fraction, [3]perfusion-related diffusion)

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Eddowes 2018[1]	822.00	33	10	1	6	97.1 [82.9, 99.8]	37.5 [16.3, 64.1]
Eddowes 2018[2]	875.00	33	8	1	8	97.1 [82.9, 99.8]	50.0 [25.5, 74.5]
Gallego-Duran 2016[3]	0.50	25	8	6	48	80.6 [61.9, 91.9]	85.7 [73.2, 93.2]
Gallego-Duran 2016[4]	0.50	24	11	7	45	77.4 [58.5, 89.7]	80.4 [67.2, 89.3]
Pavlides 2017	1.20	42	16	4	9	91.3 [78.3, 97.2]	36.0 [18.7, 57.4]

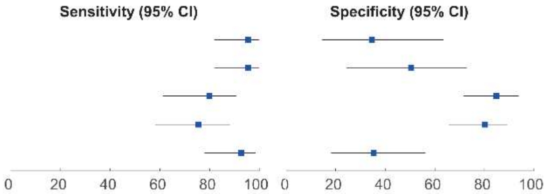

Figure S24 Forest plots of all included studies for the diagnosis of significant fibrosis (≥F2) using MRI methods. ([1, 2]different cut-offs from literature, [3]estimation cohort, [4]validation cohort)

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Pavlides 2017	1.70	24	27	2	18	92.3 [73.4, 98.7]	40.0 [26.1, 55.6]

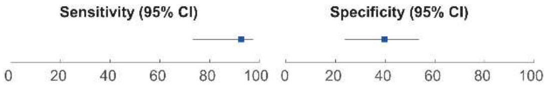

Figure S25 Forest plots of all included studies for the diagnosis of advanced fibrosis (≥F3) using MRI methods.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Pavlides 2017	3.00	10	16	1	44	90.1 [57.1, 99.5]	73.3 [60.1, 83.5]

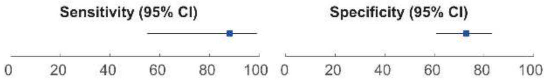

Figure S26 Forest plots of all included studies for the diagnosis of cirrhosis (F4) using MRI methods.

Study	Threshold	TP	FP	FN	TN	Se (95% CI)	Sp (95% CI)
Gallego-Duran 2016[1]	0.50	18	5	3	13	85.7 [62.6, 96.2]	72.2 [46.4, 89.3]
Gallego-Duran 2016[2]	0.50	38	17	6	26	86.4 [72.0, 94.3]	60.5 [44.5, 74.6]
Parente 2015[3]	0.76	15	13	7	24	68.2 [45.1, 85.3]	64.9 [47.4, 79.3]
Parente 2015[4]	41.45	15	11	7	26	68.2 [45.1, 85.3]	70.3 [52.8, 83.6]
Parente 2015[5]	34.23	11	11	11	26	50.0 [28.8, 71.2]	70.3 [52.8, 83.6]
Pavlides 2017	1.40	42	12	4	13	91.3 [78.3, 97.2]	52.0 [31.8, 71.7]

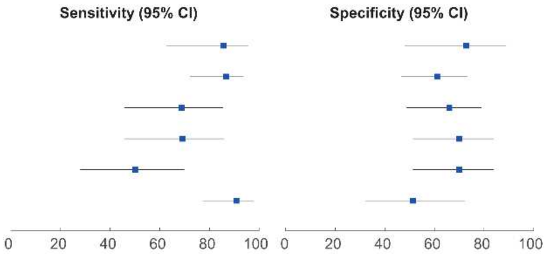

Figure S27 Forest plot of MRI studies for the diagnosis of NASH. ([1]estimation cohort, [2]validation cohort, [3]pure molecular diffusion, [4]vascular fraction, [5]perfusion-related diffusion)

Supplementary references

1. 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;37:1065-1073.
2. 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;47:631-644.
3. Gallego-Duran R, Cerro-Salido P, Gomez-Gonzalez E, et al. Imaging biomarkers for steatohepatitis and fibrosis detection in non-alcoholic fatty liver disease. Sci Rep 2016;6:31421.
4. Parente DB, Paiva FF, Oliveira Neto JA, et al. Intravoxel Incoherent Motion Diffusion Weighted MR Imaging at 3.0 T: Assessment of Steatohepatitis and Fibrosis Compared with Liver Biopsy in Type 2 Diabetic Patients. PLoS One 2015;10:e0125653.
5. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. J Hepatol 2008;48:835-47.
6. Shiina T, Nightingale KR, Palmeri ML, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology. Ultrasound Med Biol 2015;41:1126-47.
7. Venkatesh SK, Yin M, Ehman RL. Magnetic resonance elastography of liver: technique, analysis, and clinical applications. J Magn Reson Imaging 2013;37:544-55.
8. Piechnik SK, Ferreira VM, Dall'Armellina E, et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. J Cardiovasc Magn Reson 2010;12:69.
9. Nakamori S, Dohi K, Ishida M, et al. Native T1 Mapping and Extracellular Volume Mapping for the Assessment of Diffuse Myocardial Fibrosis in Dilated Cardiomyopathy. JACC Cardiovasc Imaging 2018;11:48-59.
10. Tunnicliffe EM, Banerjee R, Pavlides M, et al. A model for hepatic fibrosis: the competing effects of cell loss and iron on shortened modified Look-Locker inversion recovery T1 (shMOLLI-T1) in the liver. J Magn Reson Imaging 2017;45:450-462.
11. Bonekamp S, Torbenson MS, Kamel IR. Diffusion-weighted magnetic resonance imaging for the staging of liver fibrosis. J Clin Gastroenterol 2011;45:885-92.
12. Kwok R, Tse YK, Wong GL, et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease--the role of transient elastography and plasma cytokeratin-18 fragments. Aliment Pharmacol Ther 2014;39:254-69.
13. Wong VW, Vergniol J, Wong GL, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454-62.
14. 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.
15. de Franchis R, Baveno VIF. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. J Hepatol 2015;63:743-52.
16. Wong VW, Vergniol J, Wong GL, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862-71.
17. Vuppalanchi R, Siddiqui MS, Van Natta ML, et al. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. Hepatology 2018;67:134-144.
18. 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;17:156-163 e2.

19. Loomba R, Wolfson T, Ang B, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. Hepatology 2014;60:1920-8.
20. Barr RG, Ferraioli G, Palmeri ML, et al. Elastography Assessment of Liver Fibrosis: Society of Radiologists in Ultrasound Consensus Conference Statement. Radiology 2015;276:845-61.
21. Friedrich-Rust M, Nierhoff J, Lupsor M, et al. Performance of Acoustic Radiation Force Impulse imaging for the staging of liver fibrosis: a pooled meta-analysis. J Viral Hepat 2012;19:e212-9.
22. Banerjee R, Pavlides M, Tunnicliffe EM, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol 2014;60:69-77.
23. Pavlides M, Banerjee R, Sellwood J, et al. Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. J Hepatol 2016;64:308-315.
24. Hsu C, Caussy C, Imajo K, et al. Magnetic Resonance vs Transient Elastography Analysis of Patients With Non-alcoholic Fatty Liver Disease: a Systematic Review and Pooled Analysis of Individual Participants. Clin Gastroenterol Hepatol 2018.
25. Jiang W, Huang S, Teng H, et al. Diagnostic accuracy of point shear wave elastography and transient elastography for staging hepatic fibrosis in patients with non-alcoholic fatty liver disease: a meta-analysis. BMJ Open 2018;8:e021787.
26. Xiao G, Zhu S, Xiao X, et al. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Hepatology 2017;66:1486-1501.
27. Hashemi SA, Alavian SM, Gholami-Fesharaki M. Assessment of transient elastography (FibroScan) for diagnosis of fibrosis in non-alcoholic fatty liver disease: A systematic review and meta-analysis. Caspian J Intern Med 2016;7:242-252.
28. Liang Y, Li D. Magnetic resonance elastography in staging liver fibrosis in non-alcoholic fatty liver disease: a pooled analysis of the diagnostic accuracy. BMC Gastroenterol 2020;20:89.
29. Singh S, Venkatesh SK, Loomba R, et al. Magnetic resonance elastography for staging liver fibrosis in non-alcoholic fatty liver disease: a diagnostic accuracy systematic review and individual participant data pooled analysis. European Radiology 2016;26:1431-1440.
30. Liu H, Fu J, Hong R, et al. Acoustic Radiation Force Impulse Elastography for the Non-Invasive Evaluation of Hepatic Fibrosis in Non-Alcoholic Fatty Liver Disease Patients: A Systematic Review & Meta-Analysis. PLoS One 2015;10:e0127782.