Performance of non-invasive tests and histology for the prediction of clinical outcomes in patients with non-alcoholic fatty liver disease: an individual participant data meta-analysis

*Ferenc E Mózes, Jenny A Lee, Yasaman Vali, Osama Alzoubi, Katharina Staufer, Michael Trauner, Rafael Paternostro, Rudolf E Stauber, Adriaan G Holleboom, Anne-Marieke van Dijk, Anne Linde Mak, Jérôme Boursier, Marc de Saint Loup, Toshihide Shima, Elisabetta Bugianesi, Silvia Gaia, Angelo Armandi, Shalimar, Monica Lupșor-Platon, Vincent Wai-Sun Wong, Guanlin Li, Grace Lai-Hung Wong, Jeremy Cobbold, Thomas Karlas, Johannes Wiegand, Giada Sebastiani, Emmanuel Tsochatzis, Antonio Liguori, Masato Yoneda, Atsushi Nakajima, Hannes Hagström, Camilla Akbari, Masashi Hirooka, Wah-Kheong Chan, Sanjiv Mahadeva, Ruveena Rajaram, Ming-Hua Zheng, Jacob George, Mohammed Eslam, Salvatore Petta, Grazia Pennisi, Mauro Viganò, Sofia Ridolfo, Guruprasad Padur Aithal, Naaventhan Palaniyappan, Dae Ho Lee, Mattias Ekstedt, Patrik Nasr, Christophe Cassinotto, Victor de Lédinghen, Annalisa Berzigotti, Yuly P Mendoza, Mazen Noureddin, Emily Truong, Céline Fournier-Poizat, Andreas Geier, Miljen Martic, Theresa Tuthill, Quentin M Anstee, Stephen A Harrison, Patrick M Bossuyt, Michael Pavlides, on behalf of the LITMUS investigators**

Summary

Background Histologically assessed liver fibrosis stage has prognostic significance in patients with non-alcoholic fatty liver disease (NAFLD) and is accepted as a surrogate endpoint in clinical trials for non-cirrhotic NAFLD. Our aim was to compare the prognostic performance of non-invasive tests with liver histology in patients with NAFLD.

Methods This was an individual participant data meta-analysis of the prognostic performance of histologically assessed fibrosis stage (F0–4), liver stiffness measured by vibration-controlled transient elastography (LSM-VCTE), fibrosis-4 index (FIB-4), and NAFLD fibrosis score (NFS) in patients with NAFLD. The literature was searched for a previously published systematic review on the diagnostic accuracy of imaging and simple non-invasive tests and updated to Jan 12, 2022 for this study. Studies were identified through PubMed/MEDLINE, EMBASE, and CENTRAL, and authors were contacted for individual participant data, including outcome data, with a minimum of 12 months of follow-up. The primary outcome was a composite endpoint of all-cause mortality, hepatocellular carcinoma, liver transplantation, or cirrhosis complications (ie, ascites, variceal bleeding, hepatic encephalopathy, or progression to a MELD score ≥15). We calculated aggregated survival curves for trichotomised groups and compared them using stratified log-rank tests (histology: F0–2 *vs* F3 *vs* F4; LSM: <10 *vs* 10 to <20 *vs* ≥20 kPa; FIB-4: <1·3 *vs* 1·3 to ≤2·67 *vs* >2·67; NFS: <–1·455 *vs* –1·455 to ≤0·676 *vs* >0·676), calculated areas under the time-dependent receiver operating characteristic curves (tAUC), and performed Cox proportional-hazards regression to adjust for confounding. This study was registered with PROSPERO, CRD42022312226.

Findings Of 65 eligible studies, we included data on 2518 patients with biopsy-proven NAFLD from 25 studies (1126 [44·7%] were female, median age was 54 years [IQR 44–63), and 1161 [46·1%] had type 2 diabetes). After a median follow-up of 57 months [IQR 33–91], the composite endpoint was observed in 145 (5·8%) patients. Stratified log-rank tests showed significant differences between the trichotomised patient groups (p<0·0001 for all comparisons). The tAUC at 5 years were 0·72 (95% CI 0·62–0·81) for histology, 0·76 (0·70–0·83) for LSM-VCTE, 0·74 (0·64–0·82) for FIB-4, and 0·70 (0·63–0·80) for NFS. All index tests were significant predictors of the primary outcome after adjustment for confounders in the Cox regression.

Interpretation Simple non-invasive tests performed as well as histologically assessed fibrosis in predicting clinical outcomes in patients with NAFLD and could be considered as alternatives to liver biopsy in some cases.

Funding Innovative Medicines Initiative 2.

Lancet Gastroenterol Hepatol 2023

Published Online June 5, 2023 https://doi.org/10.1016/S2468-1253(23)00141-3

See Online/Comment https://doi.org/10.1016/S2468-1253(23)00149-8

*The LITMUS investigators are listed in the appendix (pp 2–5)

Oxford Centre for Clinical Magnetic Resonance Research, Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK (F E Mózes DPhil, Prof S A Harrison MD, M Pavlides DPhil)**; Department of Epidemiology and Data Science, Amsterdam Public Health** (J A Lee PhD, Y Vali PhD, Prof P M Bossuyt PhD) **and Department of Vascular Medicine** (A G Holleboom PhD, A-M van Dijk MD, A L Mak MD)**, Amsterdam UMC, University of Amsterdam, Netherlands; School of Medicine, The University of Jordan, Amman, Jordan** (O Alzoubi MD)**; Division of Gastroenterology and Hepatology, Department of Medicine III** (K Staufer MD, Prof M Trauner PhD, R Paternostro MD) **and Division of Transplantation, Department of Surgery** (K Staufer)**, Medical University of Vienna, Vienna, Austria; Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Graz, Austria** (Prof R E Stauber MD);

Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease globally, affecting 25–30% of the general population and up to 70% of those with obesity and type 2 diabetes.[1] NAFLD includes a range of histologically defined pathology, from accumulation of fat only (isolated steatosis), to accumulation of fat with associated inflammation and liver cell damage (hepatocyte ballooning; collectively termed non-alcoholic steatohepatitis [NASH]), and

Laboratoire HIFIH, UPRES EA 3859, SFR ICAT 4208, Université d'Angers, Angers, France (Prof J Boursier PhD); **Service d'Hépato-Gastroentérologie et Oncologie Digestive, Centre Hospitalier Universitaire d'Angers, Angers, France** (Prof J Boursier, M de Saint Loup MD); **Hepatology Center, Saiseikai Suita Hospital, Osaka, Japan** (T Shima MD); **Department of Medical Sciences, University of Turin, Torino, Italy** (Prof E Bugianesi MD, S Gaia MD, A Armandi MD); **All India Institute of Medical Sciences, New Delhi, India** (Prof Shalimar MD); **Department of Medical Imaging, Iuliu Haţieganu University of Medicine and Pharmacy, Regional Institute of Gastroenterology and Hepatology "Prof. Dr. Octavian Fodor", Cluj-Napoca, Romania** (Prof M Lupșor-Platon PhD); **Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China** (Prof V W-S Wong MD, G Li MD, Prof G L-H Wong MD); **Translational Gastroenterology Unit, University of Oxford, Oxford, UK** (J Cobbold PhD, M Pavlides); **Oxford National Institute for Health and Care Research (NIHR) Biomedical Research Centre, Oxford University Hospitals National Health Service (NHS) Foundation Trust, University of Oxford, Oxford, UK** (J Cobbold, M Pavlides); **Department of Oncology, Gastroenterology, Hepatology, Pulmonology and Infectious Diseases, University Hospital Leipzig, Leipzig, Germany** (T Karlas MD, J Wiegand MD); **Division of Gastroenterology and Hepatology and Division of Infectious Diseases, McGill University Health Centre, Montreal, QC, Canada** (G Sebastiani MD); **Sheila Sherlock Liver Unit, Royal Free London NHS Foundation Trust, London, UK** (Prof E Tsochatzis PhD, A Liguori MD); **Institute for Liver and Digestive Health, University College London, London, UK** (Prof E Tsochatzis, A Liguori); **Department of Translational Medicine and**

Research in context

Evidence before this study

Development of non-invasive biomarkers that can replace biopsy for the assessment of non-alcoholic fatty liver disease (NAFLD) remains an unmet clinical need. Liver stiffness measurement by vibration-controlled transient elastography (LSM-VCTE) is one of the more widely studied biomarkers, with data primarily originating from cross-sectional studies examining its diagnostic accuracy against liver histology. Some prospective studies have also shown that LSM-VCTE can provide prognostic information, but its performance has not been directly compared with that of biopsy. To identify patients that have been assessed with both LSM-VCTE and biopsy at baseline, we previously did a systematic review of the literature for cross-sectional studies comparing vibration-controlled transient elastography to liver histology. We searched PubMed/MEDLINE, EMBASE, and CENTRAL (Cochrane Library) from inception of the database until Jan 12, 2022. We searched for the MESH term of "elasticity imaging techniques" OR title and abstracts for the term "transient elastography" or variations of this, AND the MESH term "non-alcoholic fatty liver disease" OR titles and abstracts for the terms "NAFLD" OR NASH" OR variations of them. We did not apply any language restrictions. Authors were contacted for the baseline data included in the cross-sectional studies and, where possible, published or unpublished follow-up and outcome data. From the available data, the fibrosis-4 index (FIB-4) and NAFLD fibrosis score (NFS) were computed where possible. To the best of our knowledge, the data we have collected comprise the largest dataset directly comparing the prognostic performance of liver histology with that of LSM-VCTE, FIB-4, and NFS.

Added value of this study

In this study, we provide the first substantial evidence that non-invasive tests have similar prognostic performance to histologically assessed liver fibrosis. Histologically assessed liver fibrosis, LSM-VCTE, FIB-4, and NFS all provide useful prognostic information. There was no significant difference between the prognostic performance of the three biomarkers and histological fibrosis staging on time-dependent receiver operating characteristic curve analysis. Both histology and all three biomarkers were significant predictors of clinical outcomes when adjusting for potential confounders.

Implications of all the available evidence

The prognostic significance of histologically assessed liver fibrosis has supported the use of liver biopsy for the evaluation of patients with NAFLD in many contexts. In clinical practice, liver biopsy might be needed for the diagnosis of cirrhosis, which remains an important endpoint and affects clinical management. In clinical trials, eligibility criteria and surrogate endpoints are also based on liver histology. Our data showing similar prognostic performance of biomarkers and histology would support a change in practice, such that non-invasive tests replace biopsy in these contexts.

increasing degrees of fibrosis up to cirrhosis (F0–4).[2] Worsening stages of the disease, from isolated steatosis to NASH with fibrosis to cirrhosis, are associated with a progressively increased risk of adverse clinical outcomes.[3]

NAFLD is a major public health problem, yet there are currently no approved treatments for this condition. This is an area of intense activity, with multiple clinical trials assessing the efficacy of numerous pharmacotherapies. Although regulatory approval is generally granted for therapies that lead to improvement in how patients survive, feel, and function, these outcomes are not feasible endpoints for trials in people with non-cirrhotic NASH due to the long natural history of the disease. As there is an association between histologically assessed liver fibrosis and inflammation (ie, NASH) and adverse clinical outcomes,[3,4] improvements in these parameters are accepted as surrogate endpoints for the purposes of regulatory approval. However, this approach introduces the need for liver biopsy, which is a major limiting factor in the conduct of clinical trials due to patient reluctance to undergo invasive procedures and sampling and observer-dependent variability,[5,6] which means that increased numbers of patients are needed to achieve the required statistical power. Therefore, there is a great need to validate biomarkers other than histology as alternative surrogate endpoints, something that is likely to accelerate NAFLD drug development.

There is now a wealth of data and experience on the use of non-invasive biomarkers from numerous studies that examined the diagnostic accuracy of non-invasive tests against liver histology[7,8] in cross-sectional comparative studies. However, there is a relative scarcity of studies examining the prognostic value of non-invasive tests, and direct comparisons with liver histology have only been done in relatively few patients. Despite recent advances, no non-invasive test has yet been accepted as a surrogate endpoint and this clinical and drug development need remains unfulfilled. This gap is recognised by the regulators who encourage study sponsors to include evaluations of biomarkers in their studies.[9]

The aim of our study was to compare the prognostic performance of liver histology, liver stiffness measurements by vibration-controlled transient elastography (LSM-VCTE), the fibrosis-4 index (FIB-4),[10] and the NAFLD fibrosis score (NFS)[11] for adverse clinical outcomes in adult patients with NAFLD.

Methods

Search strategy and selection criteria

This study is an individual participant data meta-analysis. As part of the evidence synthesis efforts of the LITMUS

project, we had already conducted a systematic review and individual participant data meta-analysis of cross-sectional studies evaluating the diagnostic accuracy of imaging and simple non-invasive tests, including LSM-VCTE, FIB-4, and NFS, against liver histology as the reference standard.[7,12] The individual participant data meta-analysis reported in this Article is an extension of the aforementioned systematic review, now focusing on the prognostic performance of non-invasive tests.

Our initial searches were conducted from inception up to July 27, 2020. For this study, we updated the literature searches on PubMed/MEDLINE, EMBASE, and CENTRAL (Cochrane Library; for search terms, see the appendix [p 6]) up to Jan 12, 2022; no language restrictions were set. Both the initial and subsequent literature searches aimed to identify patients that had already been assessed in cross-sectional studies for the diagnostic accuracy of LSM-VCTE against liver histology that could then be followed up for the development of clinical outcomes. Studies in adult patients (aged >18 years) with NAFLD were eligible for inclusion if participants had undergone liver biopsy that was assessed for fibrosis and steatohepatitis, had LSM-VCTE within 6 months, and had been followed up for at least 12 months from the time of biopsy. Patients with chronic liver conditions other than NAFLD, such as alcohol-related liver disease or chronic viral hepatitis, were excluded.

Authors of eligible studies were contacted by email and the study protocol was shared with them. As our literature search identified cross-sectional studies, we asked authors if their local ethical frameworks and approvals would allow them to collect and share outcome data. Other than existing local approvals, no additional ethical approval was sought for this meta-analysis, as only anonymised data were collected centrally. A reminder message was sent after 2 weeks to authors who did not respond. The studies of authors who did not respond, or

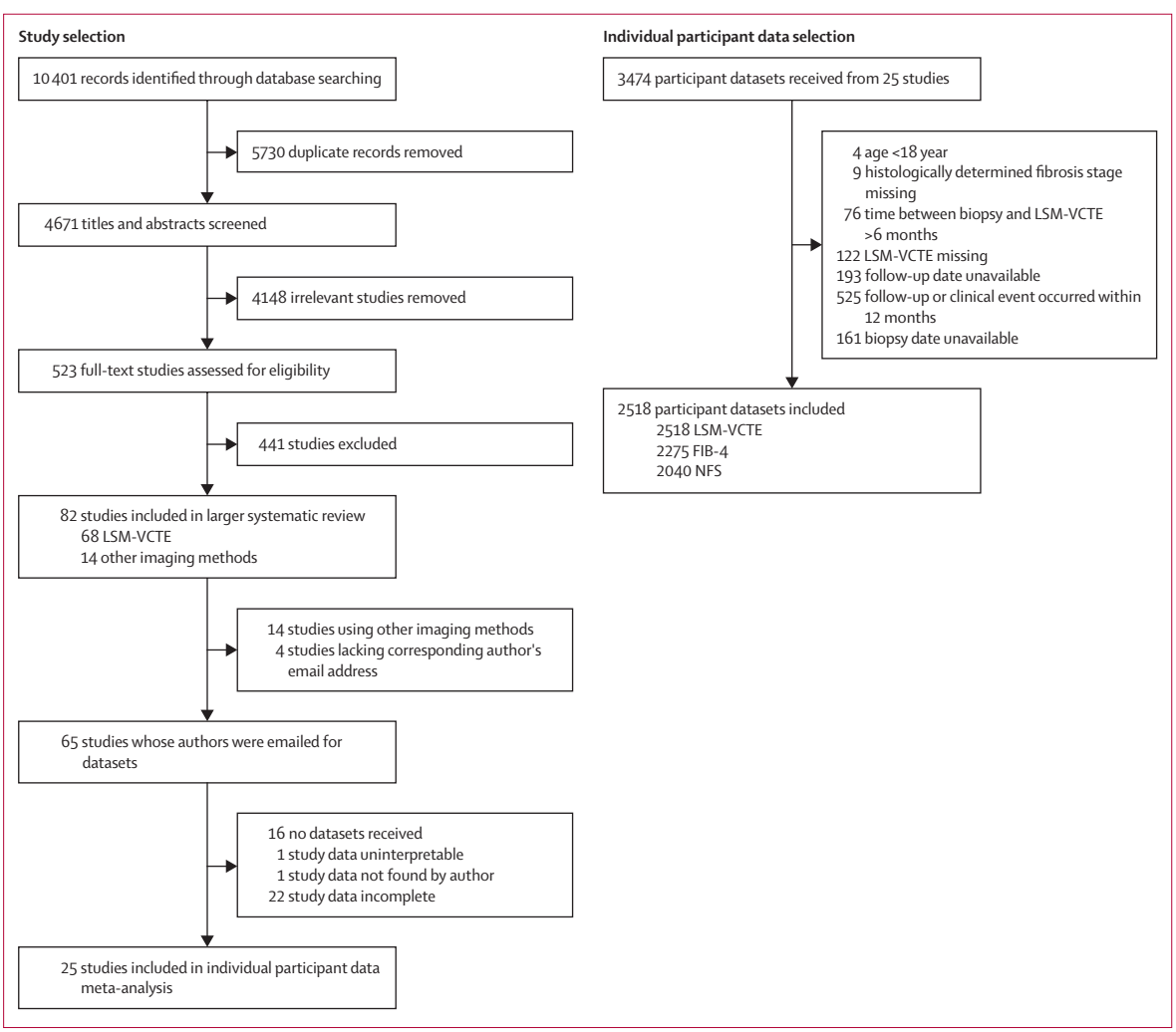

Figure 1: **Study profile**
LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. FIB-4=fibrosis-4 index. NFS=non-alcoholic fatty liver disease fibrosis score.

Surgery, Università Cattolica Del Sacro Cuore, Rome, Italy (A Liguori)**; Department of Gastroenterology and Hepatology, School of Medicine, Yokohama City University, Yokohama, Japan** (M Yoneda MD, Prof A Nakajima MD)**; Division of Liver and Pancreatic diseases, Department of Upper GI, Karolinska University Hospital, Stockholm, Sweden** (H Hagström PhD)**; Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden** (H Hagström, C Akbari MD)**; Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Touon, Japan** (M Hirooka MD)**; Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia** (Prof W-K Chan PhD, Prof S Mahadeva MD, R Rajaram MD)**; MAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China** (Prof M-H Zheng MD)**; Wenzhou Key Laboratory of Hepatology, Wenzhou, China** (Prof M-H Zheng)**; Institute of Hepatology, Wenzhou Medical University, Wenzhou, China** (Prof M-H Zheng)**; Key Laboratory of Diagnosis and Treatment for The Development of Chronic Liver Disease in Zhejiang Province, Wenzhou, China** (Prof M-H Zheng)**; Storr Liver Centre, Westmead Institute for Medical Research, Westmead Hospital, University of Sydney, Sydney, NSW, Australia** (Prof J George PhD, Prof M Eslam PhD)**; Section of Gastroenterology and Hepatology, PROMISE, Palermo, Italy** (S Petta MD, G Pennisi MD)**; Gastroenterology Hepatology and Transplantation Unit, ASST Papa Giovanni XXIII, Bergamo, Italy** (M Viganò MD)**; Hepatology Unit, Ospedale San Giuseppe, University of Milan, Milan, Italy** (S Ridolfo MD)**; NIHR Nottingham Biomedical Research Centre and Nottingham Digestive Diseases Centre, Translational Medical Sciences, School of Medicine,**

University of Nottingham, Nottingham, UK (Prof G P Aithal PhD, N Palaniyappan PhD); **Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea** (Prof D H Lee PhD); **Department of Health, Medical and Caring Sciences, Linköping University, Linköping, Sweden** (M Ekstedt PhD, P Nasr MD); **Department of Diagnostic and Interventional Radiology, Saint-Eloi Hospital and Institut Desbrest d'Epidémiologie et de Santé Publique, IDESP UMR UA11 INSERM, University Hospital of Montpellier, Montpellier, France** (C Cassinotto MD); **Centre d'Investigation de la Fibrose Hépatique, Hôpital Haut-Lévêque, Bordeaux University**

where they could not collect or share outcome data, were excluded without any attempts to extract aggregate data. Data were curated centrally by FEM and MP, and any queries were resolved directly with the authors of the primary studies. Attempts were made to recover missing data either until data were received or confirmed as unavailable. The study protocol was registered with PROSPERO.

Data analysis

We provided a template spreadsheet with the required fields, which was completed by the authors and returned to the analysis team by email. We collected data on baseline demographics (ie, age, sex, weight, BMI, presence of type 2 diabetes), histology (ie, fibrosis and NAS score), blood tests (ie, γ-Glutamyl transferase, alanine aminotransferase, aspartate aminotransferase, platelet count, and albumin), LSM-VCTE, and the number of days between liver biopsy and LSM-VCTE. We collected data on mortality (all-cause and liver-related), liver transplantation, hepatocellular carcinoma, and cirrhosis complications (ie, presence of ascites, hepatic encephalopathy, variceal bleeding, and increase in MELD score to ≥15), cardiovascular events (ie, myocardial infarction and stroke), and non-primary liver cancer. Dates of clinical event and dates of last follow-up were recorded.

The primary outcome was a composite endpoint that included all-cause mortality, liver transplantation, hepatocellular carcinoma, cirrhosis decompensation (variceal bleeding, ascites, hepatic encephalopathy), and increase in MELD score to 15 or higher. In patients who developed more than one event, only the first one was considered in the analysis. Cardiovascular events and non-primary liver cancer (ie, all extrahepatic primary cancers) were secondary outcomes.

The Quality Assessment of Prognostic Accuracy Studies (QUAPAS) tool was used to assess the methodological quality and risk of bias in the included studies.[13] In accordance with QUAPAS, we developed a tailored assessment framework that did not include signalling questions relating to analysis as the statistical analysis was defined as part of this study and not published previously.

The main index tests were histologically quantified liver fibrosis assessed using the NASH Clinical Research Network scoring system[2] and LSM-VCTE performed with FibroScan (Echosens; Paris, France). The FIB-4 and NFS scores were considered where data were available. For each index test, we considered the prognostic performance of previously established thresholds. For LSM-VCTE, we chose the upper cutoff of 20 kPa, which had a 95% sensitivity in our cross-sectional individual participant data meta-analysis, and a lower cutoff of 10 kPa, which we also previously validated as a so-called rule-in cutoff for advanced fibrosis (appendix p 6).[7]

Baseline characteristics were summarised as median (IQR). Survival analysis was done on all the available data for each outcome with each non-invasive test and histology scores. The time to event was defined as the time from the day of the biopsy or the day when non-invasive tests were performed to the day of the outcome. The time to event was considered censored at the time of last visit for those not reaching the primary outcome.

For the survival curve analysis, data were trichotomised on the basis of cutoffs from the literature (fibrosis stage: F0–2 *vs* F3 *vs* F4[4]; LSM: <10 kPa *vs* 10 to <20 kPa *vs* ≥20 kPa[7]; FIB-4: <1·3 *vs* 1·3 to 2·67 *vs* >2·67[10]; NFS: <–1·455 *vs* –1·455 to 0·676 *vs* >0·676[11]). Study-specific cumulative hazard functions and derived aggregated survival curves were calculated. Aggregated survival curves were obtained by averaging over study-specific survival curves. Differences between groups were evaluated using the log-rank test statistic.

Standardised and weighted time-dependent receiver operating characteristic (tROC) curve analysis was performed to assess the prognostic performance of the index tests; study source was considered as a fixed-effect

	Total cohort (n=2518)	FIB-4 subgroup (n=2275)	NFS subgroup (n=2040)
Demographics			
Female	1126 (44·7%)	1018 (44·8%)	927 (45·4%)
Male	1392 (55·3%)	1257 (55·2%)	1113 (54·6%)
BMI ≥30 kg/m²	993 (39·4%)	900 (39·6%)	814 (39·9%)
Type 2 diabetes	1161 (46·1%)	1044 (45·9%)	968 (47·5%)
Age, years	54 (44–63)	54 (43–63)	54 (43–63)
BMI, kg/m²	29 (26–32)	29 (26–32)	29 (26–32)
Biopsy data			
Fibrosis stage			
F0	441 (17·5%)	423 (18·6%)	377 (18·5%)
F1	745 (29·6%)	666 (29·3%)	606 (29·7%)
F2	511 (20·3%)	443 (19·5%)	390 (19·1%)
F3	511 (20·3%)	462 (20·3%)	415 (20·3%)
F4	310 (12·3%)	281 (12·3%)	252 (12·4%)
NAFLD activity score	4 (3–5)	4 (3–5)	4 (3–5)
NASH	1219 (48·4%)	1107 (48·7%)	1040 (51·0%)
Blood tests			
Alanine aminotransferase, IU/L	58 (37–89)	59 (37–89)	58 (37–89)
Aspartate aminotransferase, IU/L	41 (29–61)	42 (30–62)	42 (29–62)
Platelets, ×10⁹/L	220 (175–267)	220 (175–267)	221 (175–269)
Albumin, g/L	42 (37–45)	42 (37–46)	42 (37–46)
γ-Glutamyl transferase, IU/L	62 (35–113)	63 (36–113)	62 (36–113)
Non-invasive tests			
LSM-VCTE, kPa	8·6 (6·1–12·8)	8·6 (6·1–12·9)	8·6 (6·1–12·8)
FIB-4 index	1·33 (0·86–2·11)	1·33 (0·86–2·12)	1·32 (0·85–2·10)
NFS	–0·953 (–2·390 to 0·423)	–0·953 (–2·390 to 0·423)	–0·953 (–2·390 to 0·423)

Data are presented as median (IQR) or n (%). BMI=body-mass index. FIB-4=fibrosis-4 index. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. NAFLD=non-alcoholic fatty liver disease. NFS=NAFLD fibrosis score.

Table 1: **Baseline characteristics of the total cohort and subgroups with fibrosis-4 index and NFS**

covariate.[14] Area under the tROC curve (tAUC) and cumulative sensitivities and dynamic specificities are reported at 3, 5, and 10 years. Cumulative sensitivity was defined as the probability of positive tests among patients reaching the outcome before the specified timepoint, and dynamic specificity was defined as the probability of a negative test among patients not reaching the outcome up to the specified timepoint.

Multivariable Cox proportional-hazards modelling was performed to calculate hazard ratios (HRs) for the primary and secondary outcomes, adjusted for potentially confounding variables. To evaluate whether histology or non-invasive tests were prognostic, two models were compared using a generalised log-likelihood ratio test: one containing potential confounders and another containing the markers of interest as well as potential confounders. Fibrosis stage and LSM-VCTE were adjusted for age, sex, BMI, and presence of type 2 diabetes; FIB-4 was adjusted for sex, BMI, and presence of type 2 diabetes; NFS was adjusted for sex only. Cox regression models were stratified by study, as fixed effects, to account for study-level clustering. Because fibrosis stage has five categories, we did not include non-invasive test results as continuous variables but relied on literature-based cutoffs.

p values less than 0·05 were used to indicate significant differences. Analyses were performed using the R statistical software, version 4.2.3, using the survival, survminer, and RISCA packages. This report was prepared in accordance with the recommendations of the PRISMA-IPD statement. The study was registered with PROSPERO, CRD42022312226.

Role of the funding source

The funder of the study and the authors' institutions had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

After duplicate removal and screening, we emailed the authors of 65 studies, of whom 25 provided individual participant datasets (appendix p 14); of the 3474 participants in these studies, 2518 were eligible to

Hospital, Pessac, France (Prof V de Lédinghen PhD); **INSERM1312, Bordeaux University, Bordeaux, France** (Prof V de Lédinghen); **Department for Visceral Medicine and Surgery, Inselspital, Bern University Hospital** (Prof A Berzigotti PhD), **Department of Biomedical Research** (Prof A Berzigotti, Y P Mendoza PhD), **and Graduate School for Health Sciences** (Y P Mendoza), **University of Bern, Bern, Switzerland; Houston Research Institute, Houston Methodist Hospital, Houston, TX, USA** (M Noureddin MD); **Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA** (E Truong MD); **Echosens, Paris, France** (C Fournier-Poizat MD); **Division of Hepatology, University Hospital Würzburg,**

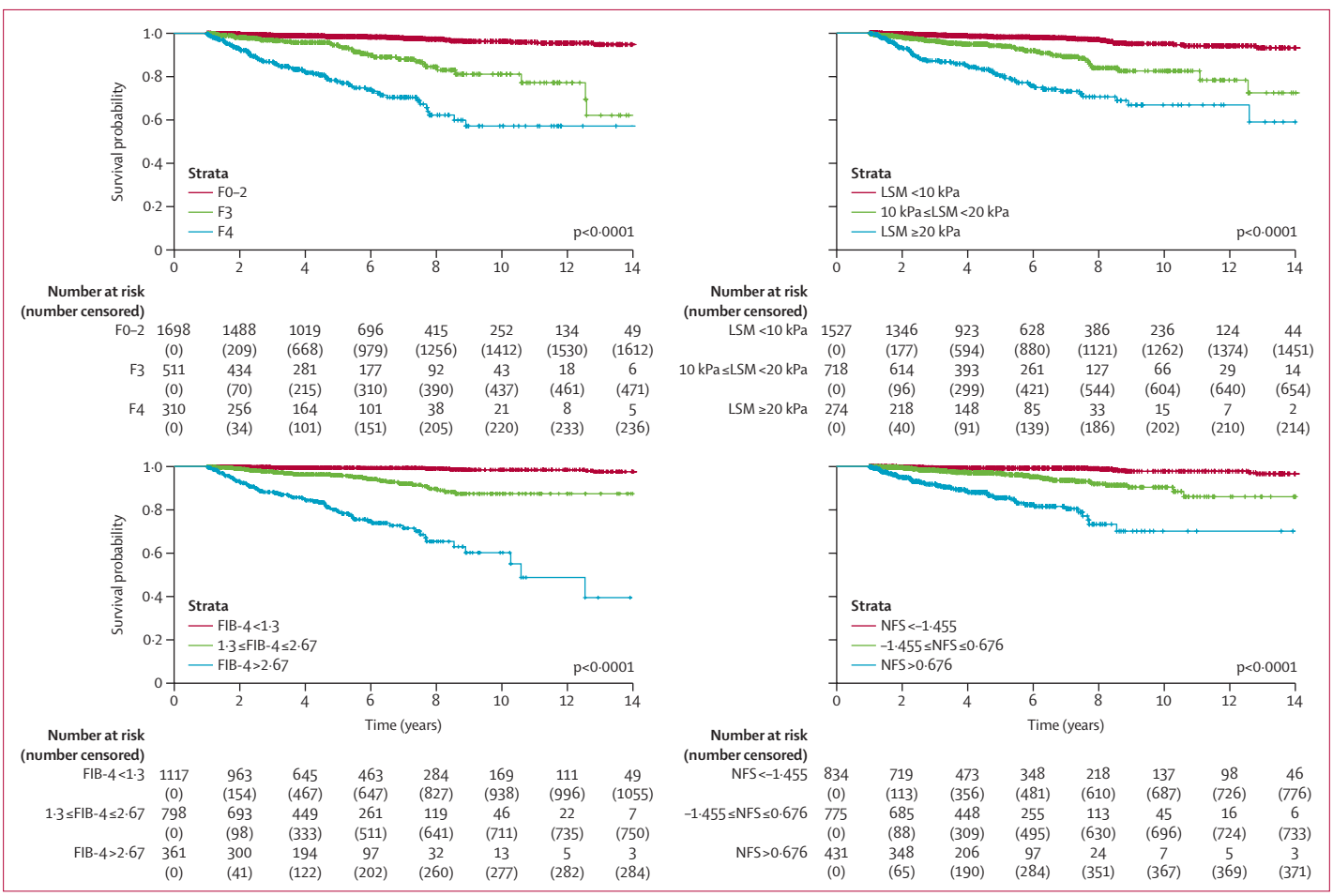

***Figure 2*: Estimates of aggregated survival probabilities for participant groups stratified by cutoffs extracted from the literature**

(A) Histologically assessed fibrosis. (B) LSM-VCTE. (C) FIB-4. (D) NFS. p values were calculated using stratified log-rank tests. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. FIB-4=fibrosis-4 index. NFS=non-alcoholic fatty liver disease fibrosis score.

Würzburg, Germany (Prof A Geier PhD); **Novartis Institutes for BioMedical Research, Basel, Switzerland** (M Martic PhD); **Digital Sciences and Translational Imaging, Pfizer, Cambridge, MA, USA** (T Tuthill PhD); **Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK** (Prof Q M Anstee PhD); **Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK** (Prof Q M Anstee)

Correspondence to: Dr Michael Pavlides, Oxford Centre for Clinical Magnetic Resonance Research (OCMR), John Radcliffe Hospital, Oxford OX3 9DU, UK
michael.pavlides@cardiov.ox.ac.uk

See Online for appendix

be included in the meta-analysis (figure 1). The median age of participants was 54 years (IQR 44–63), 1126 (44·7%) of 2518 participants were female, 1161 (46·1%) participants had type 2 diabetes, and 993 (39·4%) participants had a BMI of 30 kg/m² or more (table 1). The prevalence of fibrosis stages was 441 (17·5%) for F0, 745 (29·6%) for F1, 511 (20·3%) for F2, 511 (20·3%) for F3, and 310 (12·3%) for F4. All 2518 participants had histologically assessed fibrosis stage and LSM-VCTE available. FIB-4 was available in 2275 (90·3%) participants, and NFS was available in 2040 (81·0%) participants.

Participants were recruited between 2003 and 2021. Median follow-up was 57 months (IQR 33–91), and 145 (5·8%) of 2518 participants had an event included in our primary outcome (appendix p 15). Liver-related events developed in 114 (4·5%) participants, and 39 (1·5%) died after developing a liver-related event. There were 31 (1·2%) deaths from non-liver-related causes (appendix p 15). Baseline characteristics for participants followed up for 3, 5, and 10 years, broken down in event-free and with-event groups are presented in the appendix (pp 16–18).

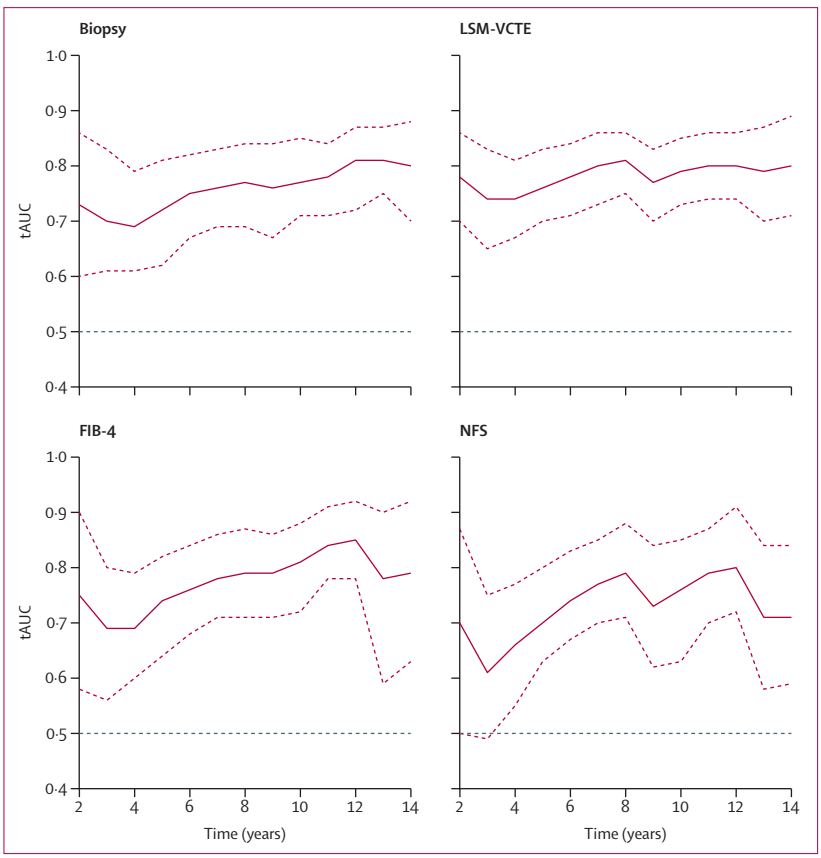

Figure 3: **tAUCs for histologically assessed fibrosis stage, LSM-VCTE, FIB-4, and NFS in predicting liver-related events and all-cause mortality**
Dashed lines represent 95% CIs. FIB-4=fibrosis-4 index. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. NFS=non-alcoholic fatty liver disease fibrosis score. tAUC=time-dependent area under the receiver operating characteristic curve.

In the participant selection domain, 14 (56%) of 25 studies had an unclear risk of bias due to unclear exclusion criteria and four (16%) had a high risk of bias for excluding participants without a valid biopsy or LSM-VCTE (appendix p 7). In the outcome domain, 20 (80%) of 25 studies had an unclear or high risk of bias due to no investigator blinding to index tests upon follow-up or collection of outcomes through medical notes review.

There were significant differences for all index tests when the total study group was trichotomised according to previously published cutoffs (p<0·0001 for all comparisons; figure 2). Survival curves for pairwise comparisons between histological fibrosis stage and individual non-invasive tests are shown in the appendix (p 8).

In general, LSM-VCTE had numerically higher tAUCs than histology, FIB-4 had similar tAUCs to histology, and NFS had numerically lower tAUCs than histology (figure 3), but pairwise differences were not significant (appendix p 9). Pairwise comparisons of tAUCs at 3-year, 5-year, and 10-year timepoints between histology and non-invasive tests are presented in table 2.

For the development of the primary outcome after 5 years, histologically diagnosed cirrhosis had a cumulative sensitivity of 33·3% (17/51) and dynamic specificity of 90·5% (1033/1141), which was similar to the performance for LSM-VCTE of 20 kPa or higher (cumulative sensitivity 29·4% [15/51] and dynamic specificity 92·0% [1050/1141]) and NFS of 0·676 or higher (cumulative sensitivity 31·6% [12/38] and dynamic specificity 84·6% [721/852]; table 3).

All index tests were significant predictors of the primary outcome on univariable Cox regression (appendix p 19). All index tests stratified by literature-based cutoffs were significant predictors of the primary outcome after adjusting for confounders (table 4; p<0·0001 when comparing log-likelihoods of Cox models with and without histology or non-invasive tests).

	3 years	5 years	10 years
Histology	0·70 (0·61–0·83; n=1816)	0·72 (0·62–0·81; n=1193)	0·77 (0·71–0·85; n=316)
LSM-VCTE	0·74 (0·65–0·83; n=1816)	0·76 (0·70–0·83; n=1193)	0·79 (0·73–0·85; n=316)
Histology	0·72 (0·61–0·83; n=1622)	0·74 (0·65–0·82; n=1032)	0·80 (0·68–0·86; n=227)
Fibrosis-4 index	0·69 (0·56–0·80; n=1622)	0·74 (0·64–0·82; n=1032)	0·81 (0·72–0·88; n=227)
Histology	0·71 (0·62–0·84; n=1440)	0·73 (0·65–0·82; n=891)	0·81 (0·72–0·88; n=188)
NFS	0·61 (0·49–0·75; n=1440)	0·70 (0·63–0·80; n=891)	0·76 (0·63–0·85; n=188)

Estimates of time-dependent AUCs are shown with 95% CIs. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. AUC=area under the receiver operating characteristic curve. NFS=non-alcoholic fatty liver disease fibrosis score.

Table 2: **Pairwise comparisons of prognostic performance of non-invasive tests versus histology, assessed with time-dependent AUC**

	Cumulative sensitivity	Dynamic specificity
Histology, F3–4 (*vs* F0–2)	66·7% (57–75)	72·0% (70–75)
Histology, F4 (*vs* F0–3)	33·3% (23–43)	90·5% (89–93)
LSM-VCTE, ≥10·0 kPa (*vs* <10kPa)	70·6% (62–79)	66·0% (64–69)
LSM-VCTE, ≥20·0 kPa (*vs* <20kPa)	29·4% (19–40)	92·0% (90–93)
FIB-4, ≥1·30 (*vs* <1·3)	82·6% (77–88)	54·5% (52–58)
FIB-4, >2·67 (*vs* ≤2·67)	41·3% (32–51)	87·7% (86–90)
NFS, ≥–1·455 (*vs* <–1·455)	78·9% (72–84)	46·5% (44–51)
NFS, >0·676 (*vs* ≤0·676)	31·6% (22–43)	84·6% (82–87)

Estimates are shown with 95% CIs. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. FIB-4=fibrosis-4 index. NFS=non-alcoholic fatty liver disease fibrosis score.

Table 3: **Prognostic performance of published non-invasive test cutoffs and histology for the development of the primary outcome at 5 years**

	Adjusted HR (95% CI)	Participants who reached the primary endpoint
Histologically assessed fibrosis stage		
F0	Reference	8/441 (1·8%)
F1	1·83 (0·75–4·47)	17/745 (2·3%)
F2	2·05 (0·79–5·27)	14/511 (2·7%)
F3	3·91 (1·64–9·34)	35/511 (6·8%)
F4	17·34 (7·57–39·70)	71/310 (22·9%)
LSM-VCTE stratified by literature cutoffs		
LSM-VCTE <10 kPa	Reference	34/1526 (2·2%)
10 kPa ≤LSM <20 kPa	3·12 (1·94–5·02)	51/718 (7·1%)
LSM-VCTE ≥20 kPa	10·65 (6·53–17·35)	60/274 (21·9%)
FIB-4 stratified by literature cutoffs		
FIB-4 <1·30	Reference	15/1116 (1·3%)
1·3 ≤FIB-4 ≤2·67	4·29 (2·30–8·00)	43/798 (5·4%)
FIB-4 >2·67	18·76 (10·17–34·60)	75/361 (20·8%)
NFS stratified by literature cutoffs		
NFS <–1·455	Reference	14/834 (1·7%)
–1·455 ≤NFS ≤0·676	3·83 (1·97–7·42)	38/775 (4·9%)
NFS >0·676	18·51 (9·38–36·50)	58/431 (13·5%)

Histology and LSM-VCTE were adjusted for age, sex, BMI, and presence of type 2 diabetes; FIB-4 was adjusted for sex, BMI, and presence of type 2 diabetes; NFS was adjusted for sex only. All analysed were stratified by study type. HR=hazard ratio. BMI=body-mass index. LSM-VCTE=liver stiffness measurement by vibration-controlled transient elastography. FIB-4=fibrosis-4 index. NFS=non-alcoholic fatty liver disease fibrosis score.

Table 4: **Multivariable Cox regression stratified by study and adjusted for confounders**

For more on the **LITMUS project** see litmus-project.eu

For the **protocol** see https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022312226

Cardiovascular events occurred in 78 (3·1%) of 2518 participants. Log-rank tests comparing event-free survival based on histology and NFS suggested significant differences between patient subgroups stratified using literature cutoffs (p=0·03 for histology; p=0·01 for NFS; appendix p 10). tAUCs were similar across histology and non-invasive tests and had poor predictive performance in the short and medium term. Predictive performance was higher in the long term, with NFS reaching tAUCs greater than 0·70 (appendix p 11). The non-invasive tests were also prognostic in Cox proportional-hazards regression after adjusting for confounders (p=0·01 when comparing log-likelihoods of Cox models with and without histology or non-invasive tests; appendix p 20).

Non-primary liver cancer occurred in 49 (1·9%) of 2518 participants. Log-rank tests comparing event-free survival based on histology and LSM-VCTE suggested significant differences between patient subgroups stratified using literature cutoffs (p<0·0001 for histology; p<0·0001 for LSM-VCTE; appendix p 12). Histologically assessed fibrosis did not appear prognostic for non-primary liver cancer, in contrast with some non-invasive tests, at a time horizon of 10 years with a tAUC of 0·44 (95% CI 0·24–0·63) for histologically assessed fibrosis, 0·64 (0·46–0·81) for LSM-VCTE, 0·67 (0·55–0·82) for FIB-4, and 0·65 (0·43–0·91) for NFS (appendix p 13). None of the non-invasive tests were significant predictors of non-primary liver cancer in confounder-adjusted Cox proportional-hazards regression (p>0·05 when comparing log-likelihoods of Cox models with and without histology or non-invasive tests; appendix p 21).

Discussion

This study compared the prognostic performance of well known non-invasive tests (LSM-VCTE, FIB-4, and NFS) to that of histologically determined fibrosis stage in patients with NAFLD. Our main findings were that non-invasive tests provided useful prognostic information, as shown by the event-free survival based on previously defined cutoffs. There were no significant differences between non-invasive tests and histological fibrosis staging in terms of performance for predicting future clinical adverse events on tROC analysis. Similar to histology, non-invasive tests were significant predictors of clinical outcomes after adjusting for confounders.

When considering the effects of treatment in patients with NASH cirrhosis, the regulators recommend a composite endpoint that includes ascites, variceal haemorrhage, hepatic encephalopathy, progression to a MELD score of 15 or higher, liver transplantation, and mortality from any cause.[15] In our study, as well as in the recent study by Sanyal and colleagues, the outcomes reflected this regulatory endpoint with the addition of hepatocellular carcinoma as part of the composite outcome. Although development of hepatocellular carcinoma might not be an important consideration in the context of clinical trials, we believe that its inclusion in prognostic studies is necessary. There is now an increasing number of studies showing that non-invasive tests have good prognostic performance in patients with NAFLD.[16–18] However, the results of these studies are not directly comparable with our study or with each other, as each evaluated a slightly different endpoint, some of which include development of varices[19] or jaundice[17] and exclude progression to a MELD score of 15 or higher, or

all-cause mortality. Although each of these composite endpoints has its own merits, harmonisation of future studies to the regulatory endpoint plus hepatocellular carcinoma would help with the interpretation of the literature. A MELD score of 15 or higher signifies the point at which the risk of mortality related to liver transplantation is lower than the mortality risk of not having a transplant and therefore indicates the presence of advanced liver disease. NAFLD is usually associated with diabetes and cardiac comorbidities, and this makes all-cause mortality very relevant in these patients.

Despite the growing evidence for the good prognostic performance of LSM-VCTE in patients with NAFLD, comparisons with the prognostic performance of liver histology are fairly scarce. We are only aware of one recent study of 594 participants, in which baseline histology and LSM-VCTE had similar performance, and both were superior to FIB-4.[19] Our results would suggest that FIB-4, LSM-VCTE, and histologically assessed fibrosis perform equally well, perhaps due to greater patient numbers and a longer follow-up period than the aforementioned study. An earlier study in patients with chronic hepatitis C also reported similar results, with AUCs of 0·76 (95% CI 0·64–0·84) for histology, 0·82 (0·68–0·90) for LSM-VCTE, and 0·75 (0·63–0·83) for FIB-4 after 5 years of follow-up.[20]

The prognostic performance of categorical histology scores is well established,[4,21,22] and these data form the basis of the use of liver histology in clinical practice and as a surrogate marker in trials. A plethora of experience also exists with non-invasive test cutoffs derived from the association of non-invasive tests with histology with some of these entering clinical guidelines.[23] For the first time, in our study, we show that the prognostic performance of previously described non-invasive test cutoffs (eg, LSM-VCTE ≥20 kPa) and histologically diagnosed cirrhosis is very similar (table 3). As the tROC curves are similar for all index tests, additional non-invasive test cutoffs with corresponding performance to fibrosis classes can be identified.

Our analysis for cardiovascular events was notable for the poor performance of non-invasive tests except for the prediction of long-term events with NFS. This is most probably because this composite score captures information about age, obesity, and diabetes, which are themselves risk factors for cardiovascular outcomes.

Prognostic biomarkers have been examined in other fields of medicine, particularly in oncology. To put our study in the context of the wider prognostic literature, we have identified some studies that report tAUCs at similar timepoints as our study. In a study of 203 patients with subjective cognitive decline at baseline, 23 progressed to mild cognitive impairment or Alzheimer's disease during a 6-year period. Models using baseline measurements of amyloid β in serum achieved a tAUC of 0·86 (95% CI 0·77–0·86) to 0·91 (0·84–0·97) at 3 years and 0·89 (0·80–0·98) to 0·91 (0·83–1·00) at 5 years for the prognosis of progression to mild cognitive impairment or Alzheimer's disease.[24] In hepatocellular carcinoma, pyroptosis-related genes in the tumour could predict 3-year survival with a tAUC of 0·68,[25] whereas a gene signature in patients with chronic hepatitis B and hepatocellular carcinoma achieved a tAUC of 0·78 at 3 years and 0·74 at 5 years.[26]

The main strength of our study was that it allowed the direct comparison of non-invasive tests to histology in the same participants over a long follow-up period (median follow-up duration was 57 months [IQR 33–91]). Furthermore, by virtue of our inclusion criteria, there was no risk of bias in the index test and flow and timing domains of the QUAPAS assessment, as we only included cases with both LSM-VCTE and histology within 6 months and the necessary minimum follow-up period of 12 months was defined a priori. In the absence of approved pharmacotherapies for NAFLD, treatment between baseline and follow-up is unlikely to have been a confounder, although we do not have data on whether lifestyle interventions led to weight loss, which might have been achieved in some patients.

However, there is potential bias from the way the follow-up data were collected as, in most cases, this was done through case note reviews and not through systematic, periodic clinical evaluation in the context of prospective cohort studies. This approach could explain the difference in the overall incidence of outcomes in our study (5·8% over 4·8 years) compared with the prospective study by Sanyal and colleagues (8·8% over 4 years).[4] This difference is mainly explained by the incidence of progression to a MELD score of 15 or higher (3·6% *vs* 1% [21 of 2518] in our analysis; appendix p 15), indicating that this outcome might have been under-reported in our analysis due to the method of outcome data collection in the primary studies. The difference in outcomes might also be due to other discrepancies in the patient population, such as sex (66% were female in the study by Sanyal and colleagues *vs* 45% in our study) or obesity (median BMI was 33 kg/m² in Sanyal and colleagues *vs* 29 kg/m² in our study). One other possible source of bias in our study is the absence of information on failed index tests, as we do not have data on the cases that were excluded from the primary datasets due to inadequate biopsy size and failed or unreliable LSM-VCTE examinations. The true performance of LSM-VCTE and histologically assessed fibrosis in clinical practice might therefore be over-estimated by our analysis. The performance of FIB-4 and NFS can be assumed to be correct, as these index tests are based on blood tests and clinical data that are more universally available.

Most of our data came from Europe and Asia with very few cases from Australia and North America, limiting the geographical generalisability of our findings. Additionally, biopsies were not centrally read by virtue of our study protocol. Finally, the changing practice and use of non-invasive test screening strategies in the community or before selecting patients for biopsy could

not be accounted for in this retrospective study with data collected over more than 10 years.

Despite the limitations of our study, we believe that it represents a major advance in the field as the first study to show, in a large patient group, that non-invasive tests provide similar prognostic information to histologically assessed liver fibrosis. This finding could have implications in clinical practice as it will allow clinicians to assess the risk of clinical outcomes without the need for liver biopsy. Furthermore, our results could have implications for the conduct of clinical trials, in which patients might be chosen for inclusion on the basis of non-invasive test classification, and efficacy of treatment might be assessed using non-invasive test surrogate endpoints.

Liver biopsy has been fundamental to our understanding of liver disease since its development in the 1940s and 1950s,[27] and underpinned the emergence of hepatology as a medical specialty. However, the invasive nature of this technique and its other limitations have driven advances that have made biopsy redundant in many contexts. For example, the advent of direct-acting antivirals for hepatitis C means that biopsy is no longer needed before treatment,[28] and improvement in immunoserology[29] means that liver biopsy is not needed in the diagnosis of primary biliary cholangitis. NAFLD remains one of the last areas of hepatology for which liver biopsy is used for disease-staging purposes. We hope that our data provide the first validation step towards the use of non-invasive tests instead of histology by clinicians and regulators for the ultimate benefit of patients.

Contributors

FEM, CF-P, AG, MM, TT, QMA, SAH, PMB, and MP contributed to the planning and design of the study. FEM and MP both directly accessed and verified the data reported in the Article. OA, FEM, and MP performed the literature review and screened manuscripts for inclusion. KS, MT, RP, RES, AGH, A-MvD, ALM, JB, MdSL, TS, EB, SG, AA, S, ML-P, VW-SW, GL, GL-HW, JC, TK, JW, GS, ETs, AL, MY, AN, HH, CA, MH, W-KC, SM, RR, M-HZ, JG, MEs, SP, GP, MV, SR, GPA, NP, DHL, MEk, PN, CC, VdL, AB, YPM, MN, and ET collected and provided individual patient data. FEM, YV, JAL, PMB, and MP performed statistical analyses and data interpretation. FEM and MP wrote the first draft of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

KS is an employee of Versantis AG. MT received speaker fees from BMS, Falk Foundation, Gilead, Intercept, Janssen, MSD, and Roche; advised for AbbVie, Albireo, BiomX, Boehringer Ingelheim, Falk Pharma, Genfit, Gilead, Hightide, Intercept, Janssen, MSD, Novartis, Phenex, Pliant, Regulus, Siemens, and Shire; received travel grants from AbbVie, Falk, Gilead, Intercept, and Janssen; and received research grants from Albireo, Alnylam, Cymabay, Falk, Gilead, Intercept, MSD, Takeda, and UltraGenyx. MT is also co-inventor of patents on the medical use of norUDCA filed by the Medical Universities of Graz and Vienna. AGH reports consultancy for Julius Clinical, Novo Nordisk, Echosens, Inventiva, and has received research grants from Novo Nordisk and Gilead. EB has served as a consultant or advisory board member for Boehringer Ingelheim, Gilead Sciences, Intercept, Merck, Novo Nordisk, Pfizer, ProSciento, and as a speaker for Gilead Sciences, Intercept, Merck, Novo Nordisk, and Pfizer. EB has also received a research grant from Gilead Sciences for fatty liver research. VW-SW has served as a consultant or advisory board member for 3V-BIO, AbbVie, Allergan, Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens, Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk, Perspectum, Pfizer, ProSciento, Sagimet Biosciences, TARGET PharmaSolutions, and Terns; and as a speaker for AbbVie, Bristol-Myers Squibb, Echosens, and Gilead Sciences. VW-SW has also received a research grant from Gilead Sciences for fatty liver research. TK and JW received unrestricted research grants from Echosens. TK has served as a speaker for Echosens. GPA has served as a consultant and an advisory board member for Pfizer, Inventiva Pharma, GlaxoSmithKline, and KaNDy Therapeutics; has been a consultant to BerGenBio, Median Technologies, FRACTYL, Amryt Pharmaceuticals, and AstraZeneca; and has given presentations on behalf of Roche Diagnostics and Medscape all through the University of Nottingham contract. GS has acted as speaker for Merck, Gilead, AbbVie, Novo Nordisk, and Pfizer; served as an advisory board member for Pfizer, Merck, Novo Nordisk, Gilead, and Intercept; and has received unrestricted research funding from Theratechnologies. ETs has served on the advisory boards for Boehringer, Pfizer, NovoNordisk, Orphalan, Univar, and Alexion; and has been a speaker for NovoNordisk and Dr Falk. MY received research support from Kowa Co. HH's institutions have received research grants from Astra Zeneca, EchoSens, Gilead, Intercept, MSD, and Pfizer. W-KC has served as a consultant or advisory board member for AbbVie, Boehringer Ingelheim, and Novo Nordisk; and as a speaker for Hisky Medical and Viatris. SM received honorarium fees from Echosens. MV has served as a consultant or a speaker for Gilead Sciences and Intercept Pharmaceuticals. VdL reports consultancy fees for AbbVie, BMS, Echosens, Gilead Sciences, Intercept Pharmaceuticals, MSD, Myr-Pharma, Pfizer, Supersonic Imagine, and Tillotts. MN has been on the advisory board or has been a consultant for 89BIO, Altimmune, BI, Gilead, cohBar, Cytodyn, Pfizer, GSK, Novo Nordisk, EchoSens, Madrigal, NorthSea, Perspectum, Terns, Takeda, Sami-Sabina group, Siemens, and Roche diagnostic; has received research support from Allergan, Akero, BMS, Gilead, Galmed, Galectin, Genfit, Conatus, Corcept, Enanta, Madrigal, Novartis, Pfizer, Shire, TERNS, Viking, and Zydus; and is a shareholder or has stocks in Anaetos, Chrownwell, Cytodyn, Ciema, Rivus Pharma, and Viking. CF-P is employed by Echosens. AG has served as a speaker and consultant for AbbVie, Alexion, AstraZeneca, Bayer, BMS, CSL Behring, Eisai, Falk, Gilead, Heel, Intercept, Ipsen, Merz, MSD, Novartis, Pfizer, Roche, Sanofi-Aventis, Sequana; and received research funding from Intercept, Falk, and Novartis. MM is employed by Novartis. TT is employed by Pfizer. QMA is coordinator of the IMI2 LITMUS consortium; has received research grant funding from AbbVie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer, Vertex; has received consultancy fees on behalf of Newcastle University for Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly, Galmed, Genfit, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe, Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk, Pfizer, Poxel, ProSciento, Raptor Pharma, Servier, Viking Therapeutics; and has received speaker fees from Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit, Gilead, Integritas Communications, MedScape. SAH has received research grants from Akero, Altimmune, Axcella-Cirius, CiVi Biopharma, Cymabay, Galectin, Genfit, Gilead Sciences, Hepion Pharmaceuticals, Hightide Therapeutics, Intercept, Madrigal, Metacrine, NGM Bio, Northsea Therapeutics, Novartis, Novo Nordisk, Poxel, Sagimet, and Viking; and has received consulting fees from Akero, Altimmune, Alentis, Arrowhead, Axcella, Echosens, Enyo, Foresite Labs, Galectin, Genfit, Gilead Sciences, Hepion, Hightide, HistoIndex, Intercept, Kowa, Madrigal, Metacrine, NeuroBo, NGM, Northsea, Novartis, Novo Nordisk, Poxel, Perspectum, Sagimet, Terns, and Viking. MP is a shareholder of Perspectum. All other authors declared no competing interests.

Data sharing

We do not plan to make the individual participant data publicly available. Requests for access to the full dataset can be made by sending an email with a research plan to the corresponding author who will then gain authorisation from the primary investigators regarding the sharing of their data. The decision to share data will be made by each individual primary investigator after review of the submitted research plan. Alternatively, requests for data from individual studies can be made directly to the primary investigators.

Acknowledgments

This individual participant data meta-analysis has been conducted as part of the imaging study in the Liver Investigation: Testing Marker Utility in Steatohepatitis (LITMUS) study. The LITMUS study is a large multicentre study aiming to evaluate biomarkers on non-alcoholic fatty liver disease. The LITMUS study is funded by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under Grant Agreement 777377. This Joint Undertaking receives support from the EU's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the view of the LITMUS consortium and neither IMI nor the EU and EFPIA are liable for any use that may be made of the information contained herein. GS is supported by a Senior Salary Award from Fonds de Recherche du Quebec–Sante (#296306). FEM was partially funded by a Sir Henry Dale Fellowships awarded jointly by the Wellcome Trust and the Royal Society [221805/Z/20/Z]. DHL received a grant from KHIDI (HI14C1135), funded by the Ministry of Health & Welfare, Korea. QMA is an NIHR Senior Investigator and is supported by the Newcastle NIHR Biomedical Research Centre.

References

1 Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018; **15:** 11–20.

2 Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; **41:** 1313–21.

3 Simon TG, Roelstraete B, Khalili H, Hagström H, Ludvigsson JF. Mortality in biopsy-confirmed nonalcoholic fatty liver disease: results from a nationwide cohort. *Gut* 2021; **70:** 1375–82.

4 Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. *N Engl J Med* 2021; **385:** 1559–69.

5 Pavlides M, Birks J, Fryer E, et al. Interobserver variability in histologic evaluation of liver fibrosis using categorical and quantitative scores. *Am J Clin Pathol* 2017; **147:** 364–69.

6 Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol* 2020; **73:** 1322–32.

7 Mózes FE, Lee JA, Selvaraj EA, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. *Gut* 2022; **71:** 1006–19.

8 Vali Y, Lee J, Boursier J, et al. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. *Lancet Gastroenterol Hepatol* 2023; published online March 20. https://doi.org/10.1016/s2468-1253(23)00017-1.

9 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: developing drugs for treatment. Nov 29, 2018. https://www.fda.gov/media/119044/download (accessed Sept 15, 2021).

10 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:** 1104–12.

11 Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007; **45:** 846–54.

12 Selvaraj EA, Mózes FE, Jayaswal ANA, et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis. *J Hepatol* 2021; **75:** 770–85.

13 Lee J, Mulder F, Leeflang M, Wolff R, Whiting P, Bossuyt PM. QUAPAS: an adaptation of the QUADAS-2 tool to assess prognostic accuracy studies. *Ann Intern Med* 2022; **175:** 1010–18.

14 Le Borgne F, Combescure C, Gillaizeau F, et al. Standardized and weighted time-dependent receiver operating characteristic curves to evaluate the intrinsic prognostic capacities of a marker by taking into account confounding factors. *Stat Methods Med Res* 2018; **27:** 3397–410.

15 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Nonalcoholic steatohepatitis with compensated cirrhosis: developing drugs for treatment. June 3, 2019. https://www.fda.gov/media/127738/download (accessed May 15, 2023).

16 Younes R, Caviglia GP, Govaere O, et al. Long-term outcomes and predictive ability of non-invasive scoring systems in patients with non-alcoholic fatty liver disease. *J Hepatol* 2021; **75:** 786–94.

17 Pennisi G, Enea M, Pandolfo A, et al. AGILE 3+ score for the diagnosis of advanced fibrosis and for predicting liver-related events in NAFLD. *Clin Gastroenterol Hepatol* 2022; published online July 13. https://doi.org/10.1016/j.cgh.2022.06.013.

18 Lee J, Vali Y, Boursier J, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: a systematic review. *Liver Int* 2021; **41:** 261–70.

19 Boursier J, Hagström H, Ekstedt M, et al. Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events. *J Hepatol* 2022; **76:** 1013–20.

20 Vergniol J, Foucher J, Terrebonne E, et al. Noninvasive tests for fibrosis and liver stiffness predict 5-year outcomes of patients with chronic hepatitis C. *Gastroenterology* 2011; **140:** 1970–79.

21 Hagström H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. *J Hepatol* 2017; **67:** 1265–73.

22 Ekstedt M, Hagström H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015; **61:** 1547–54.

23 Berzigotti A, Tsochatzis E, Boursier J, et al. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis—2021 update. *J Hepatol* 2021; **75:** 659–89.

24 Stockmann J, Verberk IMW, Timmesfeld N, et al. Amyloid-β misfolding as a plasma biomarker indicates risk for future clinical Alzheimer's disease in individuals with subjective cognitive decline. *Alzheimers Res Ther* 2020; **12:** 169.

25 Duan S, Gao J, Lou W, et al. Prognostic signature for hepatocellular carcinoma based on 4 pyroptosis-related genes. *BMC Med Genomics* 2022; **15:** 166.

26 Huang Y, Tu WL, Yao YQ, Cai YL, Ma LP. Construction of a novel gene-based model for survival prediction of hepatitis B virus carriers with HCC development. *Front Genet* 2021; **12:** 720888.

27 Menghini G. One-second needle biopsy of the liver. *Gastroenterology* 1958; **35:** 190–99.

28 Pawlotsky JM, Negro F, Aghemo A, et al. EASL recommendations on treatment of hepatitis C: final update of the series. *J Hepatol* 2020; **73:** 1170–218.

29 Hirschfield GM, Beuers U, Corpechot C, et al. EASL Clinical Practice Guidelines: the diagnosis and management of patients with primary biliary cholangitis. *J Hepatol* 2017; **67:** 145–72.