UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Vibration-Controlled Transient Elastography Scores to Predict Liver-Related Events in Steatotic Liver Disease

VCTE-Prognosis Study Group; Lin, Huapeng; Lee, Hye Won; Yip, Terry Cheuk-Fung; Tsochatzis, Emmanuel; Petta, Salvatore; Bugianesi, Elisabetta; Yoneda, Masato; Zheng, Ming-Hua; Hagström, Hannes; Boursier, Jérôme; Calleja, José Luis; Goh, George Boon-Bee; Chan, Wah-Kheong; Gallego-Durán, Rocio; Sanyal, Arun J.; de Lédinghen, Victor; Newsome, Philip; Fan, Jian-Gao; Castéra, Laurent

DOI:
10.1001/jama.2024.1447

License:
None: All rights reserved

Document Version
Peer reviewed version

Citation for published version (Harvard):
VCTE-Prognosis Study Group, Lin, H, Lee, HW, Yip, TC-F, Tsochatzis, E, Petta, S, Bugianesi, E, Yoneda, M, Zheng, M-H, Hagström, H, Boursier, J, Calleja, JL, Goh, GB-B, Chan, W-K, Gallego-Durán, R, Sanyal, AJ, de Lédinghen, V, Newsome, P, Fan, J-G, Castéra, L, Lai, M, Harrison, SA, Fournier-Poizat, C, Wong, GL-H, Pennisi, G, Armandi, A, Nakajima, A, Liu, W-Y, Shang, Y, de Saint-Loup, M, Llop, E, Teh, KK-J, Lara-Romero, C, Asgharpour, A, Mahgoub, S, Chan, MS-W, Canivet, CM, Romero-Gomez, M, Kim, SU & Wong, VW-S 2024, 'Vibration-Controlled Transient Elastography Scores to Predict Liver-Related Events in Steatotic Liver Disease', *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2024.1447

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 15. May. 2024

Transient elastography scores to predict liver-related events in steatotic liver disease

Short title: Agile scores in MASLD

Huapeng Lin[1,2]*, PhD, Hye Won Lee[3]*, PhD, Terry Cheuk-Fung Yip[1,2]*, PhD, Prof Emmanuel Tsochatzis[4]*, PhD, Prof Salvatore Petta[5], PhD, Prof Elisabetta Bugianesi[6], MD, Prof Masato Yoneda[7], PhD, Prof Ming-Hua Zheng[8], MD, Prof Hannes Hagström[9,10], PhD, Prof Jérôme Boursier[11,12], MD, Prof José Luis Calleja[13], MD, George Boon-Bee Goh[14], MD, Prof Wah-Kheong Chan[15], MD, Rocio Gallego-Durán[16], PhD, Prof Arun J. Sanyal[17], MD, Prof Victor de Lédinghen[18], MD, Prof Philip N Newsome[19], MD, Prof Jian-Gao Fan[20], PhD, Prof Laurent Castéra[21], MD, Prof Michelle Lai[22], PhD, Prof Stephen Harrison[23, 24], MD, Céline Fournier-Poizat[25], PhD, Prof Grace Lai-Hung Wong[1,2], MD, Grazia Pennisi[5], MD, Angelo Armandi[6], MD, Prof Atsushi Nakajima[7], MD, Wen-Yue Liu[26], MD, Ying Shang[9], PhD, Marc de Saint-Loup[11], PhD, Elba Llop[13], PhD, Kevin Kim Jun Teh[14], MD, Carmen Lara-Romero[16], PhD, Amon Asgharpour[17], PhD, Sara Mahgoub[19], PhD, Mandy Sau-Wai Chan[25], BSc, Clemence M Canivet[11,12], PhD, Prof Manuel Romero-Gomez[16], MD, Prof Seung Up Kim[3], PhD, Prof Vincent Wai-Sun Wong[1,2], MD

Affiliations:

1. Medical Data Analytics Centre, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China
2. State Key Laboratory of Digestive Disease, Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong, China
3. Department of Internal Medicine, Institute of Gastroenterology, Yonsei University College of Medicine, Seoul, Republic of Korea
4. University College London Institute for Liver and Digestive Health, Royal Free Hospital and UCL, London, United Kingdom
5. Sezione di Gastroenterologia, Di.Bi.M.I.S., University of Palermo, Italy
6. Department of Medical Sciences, Division of Gastroenterology and Hepatology, A.O. Città della Salute e della Scienza di Torino, University of Turin, Turin, Italy
7. Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, Yokohama, Japan

8. MAFLD Research Center, Department of Hepatology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
9. Department of Medicine, Huddinge, Karolinska Institutet, Sweden
10. Division of Hepatology, Department of Upper GI Diseases, Karolinska University Hospital, Huddinge, Stockholm, Sweden
11. Hepato-Gastroenterology and Digestive Oncology Department, Angers University Hospital, Angers, France
12. HIFIH Laboratory, SFR ICAT 4208, Angers University, Angers, France
13. Department of Gastroenterology and Hepatology, Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain
14. Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore
15. Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Malaysia
16. Digestive Diseases Unit and CIBERehd, Virgen Del Rocío University Hospital, Seville, Spain
17. Division of Gastroenterology, Hepatology and Nutrition, Department of Internal Medicine, Virginia Commonwealth University School of Medicine, Richmond, VA, USA
18. Centre d'Investigation de la Fibrose Hépatique, Haut-Lévêque Hospital, University Hospital of Bordeaux, Pessac, France
19. National Institute for Health Research, Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, Birmingham, United Kingdom
20. Department of Gastroenterology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai Key Lab of Pediatric Gastroenterology and Nutrition, Shanghai, China
21. Université Paris Cité, UMR1149 (CRI), INSERM, Paris, France; Service d'Hépatologie, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris (AP-HP), Clichy, France
22. Division of Gastroenterology & Hepatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
23. Radcliffe Department of Medicine, University of Oxford Oxford, UK
24. Pinnacle Clinical Research, San Antonio, Texas, USA

25. Echosens, Paris, France
26. Department of Endocrinology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
*These authors contributed equally to this work.

Correspondence:
Prof Seung Up Kim
Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
Tel: 82-2-2228-1944
Email: KSUKOREA@yuhs.ac

Prof Vincent Wong
Department of Medicine and Therapeutics, Prince of Wales Hospital, Shatin, Hong Kong, China
Tel: 852-35054205
Email: wongv@cuhk.edu.hk

Word count: 2997
Number of tables: 2
Number of figures: 3

Author contributions:

HL and VW-SW performed literature search. VW-SW designed the study. HL, HWL, TC-FY, ET, SP, EB, MY, M-HZ, HH, JB, JLC, GB-BG, W-KC, RG-D, AJS, VdL, PNN, J-GF, GL-HW, GP, AA, AN, W-YL, YS, MdS-L, EL, KKJT, CL-R, AA, SM, CMC, MR-G, SUK and VW-SW collected data in this study. ET, SP, EB, M-HZ, HH, JB, JLC, GB-BG, W-KC, AJS, VdL, PNN, MR-G, SUK and VW-SW supervised the project. HL, TC-FY and VW-SW were responsible for data analysis and data interpretation, and drafted the manuscript. HL prepared the figures. All authors provided review and editing of the manuscript, and approved the final version of the manuscript.

Role of the funding source

The funder of the study did not have a role in study design, data collection, data analysis, data interpretation, or manuscript preparation. Echosens provided logistic support in contacting investigators and organizing investigator meetings but did not provide funding for this study.

Data sharing statement

Data are available upon reasonable request to corresponding authors.

Potential conflict-of-interest statements:

ET served as a consultant for Pfizer, NovoNordisk, Boehringer, and Siemens Healthineers; and a speaker for NovoNordisk, Echosens, and Dr Falk. SP served as a speaker or advisor for AbbVie, Echosens, MSD, Novo Nordisk, Pfizer, and Resalis. EB served as a consultant for Boehringer, MSD, Novo Nordisk, and Pfizer; and a speaker for MSD, Novo Nordisk, and Madrigal. She received research grants from Gilead Sciences. HH served as a consultant for AstraZeneca; and a hepatic events adjudication committee member for KOWA and GW Pharama. His institution has received research funding from AstraZeneca, Echosens, Gilead Sciences, Intercept, MSD, and Pfizer. JB served as a consultant for AstraZeneca, Echosens, Intercept, and Siemens; a speaker for AbbVie, Gilead Sciences, Intercept, and Siemens; and an

advisory board member for Bristol-Myers-Squibb, Intercept, Pfizer, MSD, and Novo Nordisk. His institution has received research funding from Diafir, Echosens, Intercept, Inventiva, and Siemens. JLC served as a consultant and speaker for Echosens, Gilead Sciences, and AbbVie. GB-BG served as a consultant for Roche and Ionis Pharmaceuticals; and a speaker for Echosens, Viatris, Abbott and Novo Nordisk. WKC served as a consultant for Roche, AbbVie, Novo Nordisk, and Boehringer Ingelheim; and a speaker for Echosens, Hisky Medical, and Viatris. AJS served as a consultant for 89Bio, Akero, Allergan, Alnylam Pharmaceuticals, Amgen Inc, AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, Genentech, Gilead Sciences, Histoindex, Intercept Pharmaceuticals, Inventiva, Madrigal, Merck, Novartis, Novo Nordisk, Pfizer, Poxel, Salix Pharmaceuticals, Siemens, Sun Pharmaceutical Industries Inc, Terns, and Valeant Pharmaceuticals; and a data safety monitoring board or advisory board member for Bard Peripheral Vascular Inc, NGM Biopharmaceuticals, and Sequana. He has received research funding from Albireo, Allergan, Echosens, Eli Lilly, Gilead Sciences, Intercept Pharmaceuticals, Mallinckrodt LLC, Merck, Novo Nordisk, Perspectum, Pfizer, Salix Pharmaceuticals, and Zydus; and holds the stocks of Durect, Exhalenz, Gen t, and Tiziana. MR-G served as a consultant for Siemens; and a speaker for Siemens and Echosens. He has received research funding from Siemens, Echosens, and Novo Nordisk. PN served as a consultant for Novo Nordisk, Boehringer Ingelheim, Gilead Sciences, Intercept, Poxel Pharmaceuticals, Pfizer, BMS, Eli Lilly, Madrigal, and GSK; and a speaker for Novo Nordisk and AiCME. He has received research funding from Novo Nordisk. LC served as a consultant for Echosens, Novo Nordisk, Madrigal, MSD, Pfizer, Sagimet, and Siemens Healthlineers; and a speaker for Echosens, Gilead, Inventiva, and Novo Nordisk. SH served as a consultant for Akero Therapeutics, Aligos, Altimmune, Arrowhead, Bluejay Therapeutics, Boxer Capital, Chronwell, Echosens, Enyo, Foresite Labs, Galectin, Galecto, Gilead, GSK, Hepagene, Hepion, Hepta Bio, HistoIndex, Humana, Intercept, Ionis, Madrigal, Medpace, NeuroBo Pharmaceuticals, Northsea, Novo Nordisk, Perspectum, Pfizer, Sonic Incytes, Sagimet, Terns, and Viking; an advisory board member for Akero, Altimmune, Arrowhead,BVF Partners, Chronwell, Echosens, Foresite Labs, Galectin, Genfit, Gilead, Hepion, Hightide, HistoIndex, Humana, Intercept, Madrigal, Medpace, Metacrine, NGM Bio, Northsea, Novo Nordisk, PathAI, Poxel, Sagimet Biosciences, Sonic Incytes Medical Corp, and Terns. He holds the stocks of Akero Therapeutics, Chronwell, Cirius, Galectin, Genfit,

Hepion, Hepta Bio, HistoIndex, Metacrine, NGM Bio, Northsea, and Sonic Incytes; and grants and contracts from Akero Therapeutics, Altimmune, Axcella, BMS, Corcept, Cymabay, Enyo, Galectin, Genentech, Genfit, Gilead, GSK, Hepion, Hightide, Immuron, Intercept, Inventiva, Ionis, Madrigal, NGM Bio, Novartis, Novo Nordisk, Northsea, Pfizer, Poxel, Sagimet, Terns, and Viking. CF-P is an employee of Echosens. GL-HW served as a consultant for AstraZeneca, Gilead Sciences, and Janssen; and a speaker for Abbott, Abbvie, Bristol-Myers Squibb, Echosens, Furui, Gilead Sciences, and Roche. She has received research funding from Gilead Sciences. MS-WC is an employee of Echosens. VW-SW served as a consultant or advisory board member for AbbVie, Boehringer Ingelheim, Echosens, Gilead Sciences, Intercept, Inventiva, Novo Nordisk, Pfizer, Sagimet Biosciences, TARGET PharmaSolutions, and Visirna; and a speaker for Abbott, AbbVie, Gilead Sciences, Novo Nordisk, and Unilab. He has received a research grant from Gilead Sciences, and is a co-founder of Illuminatio Medical Technology.

Key Points

Question

What are the clinical implications of single or serial measurements of vibration-controlled transient elastography (VCTE)-based Agile scores in metabolic dysfunction-associated steatotic liver disease?

Findings

This multi-center cohort study demonstrated the Agile scores outperformed most non-invasive tests and were at least similar if not better than histological fibrosis staging in predicting liver-related events. Importantly, on repeated testing, the Agile scores were largely stable, and patients with improvement in the Agile scores had substantial reduction in the risk of liver-related events.

Meaning

The VCTE based Agile scores are generally accurate for predicting liver-related events, making them suitable alternatives to liver biopsy in routine clinical practice and in phase 2b and 3 clinical trials for steatohepatitis treatment response.

Abstract

Importance: Metabolic dysfunction-associated steatotic liver disease (MASLD) is currently the most common chronic liver disease. It is important to develop non-invasive tests to assess the disease severity and prognosis.

Objective: We aimed to study the prognostic implications of baseline levels and dynamic changes of the vibration-controlled transient elastography (VCTE)-based Agile scores.

Design, Setting, and Participants: This cohort study included data of patients with MASLD who underwent VCTE examination at 16 centers in the United States, Europe, and Asia. The Agile scores were compared with histology and 8 other non-invasive tests.

Main Outcomes and Measures: The primary outcome was liver-related events (LREs), defined as hepatocellular carcinoma or hepatic decompensation (ascites, variceal hemorrhage, hepatic encephalopathy, or hepatorenal syndrome), liver transplantation, and liver-related deaths.

Results: 16 603 patients underwent VCTE examination at baseline. At a median follow-up of 51.7 months, 316 (1.9%) patients developed LREs. Both Agile 3+ and Agile 4 scores classified fewer patients between the low and high cutoffs than most fibrosis scores and achieved the highest discriminatory power in predicting LREs (integrated area under time-dependent receiver-operating characteristic curve 0.89). 10 920 patients had repeated VCTE at a median interval of 15 months and were included in the serial analysis. 81.9% and 92.1% of patients had stable Agile 3+ and Agile 4 scores (same risk categories at both assessments). The incidence of LREs was 0.6 and 30.1 per 1 000 person-years in patients with persistently low and high Agile 3+ scores, respectively. In patients with high Agile 3+ score at baseline, a decrease in the score by more than 20% was associated with substantial reduction in the risk of LREs. A similar trend was observed for the Agile 4 score, though it missed more LREs in the low-risk group.

Conclusions and Relevance: Single or serial Agile scores are highly accurate in predicting LREs in patients with MASLD.

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously known as non-alcoholic fatty liver disease (NAFLD), is currently the most common chronic liver disease that affects around 30% of the global adult population.[1] It has become one of the leading causes of cirrhosis and hepatocellular carcinoma (HCC) in middle- and high-income countries,[2] with an estimated annual direct medical costs of around US$103 billion in the United States and €35 billion in Europe.[3]

In patients with MASLD, there is a dose-response relationship between the severity of liver fibrosis and future risk of liver-related events (LREs).[4] In the past two decades, a number of non-invasive tests of fibrosis have been adopted for clinical use.[5] In particular, liver stiffness measurement (LSM) by vibration-controlled transient elastography (VCTE) not only reflects the degree of liver fibrosis but also predicts HCC, portal hypertension and varices.[6] Recently, by combining LSM and simple clinical parameters (platelet count, aminotransferases, diabetes, age and sex), we derived and validated the Agile 3+ and Agile 4 scores for the diagnosis of advanced fibrosis and cirrhosis in patients with MASLD with improved accuracy and reduced indeterminate zone compared with LSM alone.[7] Emerging data suggest that the Agile scores are also prognostic.[8] However, previous studies were limited by small sample sizes. Besides, the prognostic meaning of a change in non-invasive tests over time is unclear, especially as the tests are imperfect and may have false-positive and false-negative results.

With this background, we aimed to evaluate the prognostic implications of baseline and repeated Agile score and liver stiffness measurements in a large cohort of patients

with MASLD. We also compared the prognostic performance of the Agile score to that of other various non-invasive tests of hepatic fibrosis.

Methods

Study design and participants

This was a retrospective cohort study of patients with MASLD who had undergone VCTE examination at 16 centers from the United States, Europe, and Asia. Eligible patients were at least 18 years old with hepatic steatosis diagnosed by histology (steatosis in ≥5% of hepatocytes) or imaging studies (ultrasound, computed tomography or magnetic resonance imaging, or controlled attenuation parameter ≥248 dB/m by VCTE). Patients were excluded if they had other liver diseases such as chronic viral hepatitis, human immunodeficiency virus infection, excessive alcohol consumption (>30 g/day in men and >20 g/day in women), secondary causes of hepatic steatosis (e.g., use of systemic steroids), or a history of HCC, hepatic decompensation, liver resection, liver transplantation or other malignancies.

The study protocol was approved by the institutional review boards of the participating sites. The study was conducted in accordance with the principles of the Declaration of Helsinki. Informed written consent was waived because of the retrospective nature of this study.

Assessments

At each clinic visit, the medical history of a patient was recorded. Body mass index was calculated as body weight (kg) divided by body height (m) squared. A venous blood sample was taken after at least 8 hours of fasting for renal and liver

biochemistry and complete blood count. Controlled attenuation parameter and liver stiffness were assessed using the VCTE machine (FibroScan, Echosens, Paris, France) by trained operators as previously described, and patients needed to have at least 10 valid acquisitions (eMethods).[9]

Based on the above assessments, we calculated the VCTE-based scores including the Agile 3+, Agile 4 and FibroScan-aspartate aminotransferase (FAST) scores (supplement p 3).[7,10] For comparison, we also calculated simple fibrosis scores including the Fibrosis-4 index (FIB-4), NAFLD fibrosis score (NFS), AST-to-platelet ratio index (APRI), BARD score and AST-to-alanine aminotransferase ratio (AAR). All calculations and cut-offs were based on the existing literature.[11] Only parameters measured within 1 month of each other were used to calculate the scores. Otherwise, the particular noninvasive test was treated as missing.

Outcomes

The primary outcome was a composite endpoint of LREs including HCC, hepatic decompensation (ascites, variceal hemorrhage, hepatic encephalopathy or hepatorenal syndrome), liver transplantation and liver-related death. Secondary outcomes included HCC and hepatic decompensation, analyzed separately. The diagnosis of the events was based on prospective follow-up, chart review, or validated registries with positive predictive values of at least 90%.

Statistical analysis

All statistical analyses were performed using R software (version 4.2.2; R Core Team 2022). In the baseline model, the baseline date was defined as the date of the first non-

invasive test. For the Agile and FAST scores that included both VCTE and blood tests, the latter date was taken as baseline to avoid immortal time bias. Pairwise comparisons between the Agile scores and the other tests were performed by comparing the area under the receiver-operating characteristic curves (AUROC) using Z test for patients in whom the results of both tests were available.[12] We also calculated the integrated AUROC, area under the time-dependent precision-recall curves (AUPRC),[13] and integrated Brier score over time. The Agile scores and other tests were evaluated for continuous net reclassification improvement (NRI) with reference to LSM using the inverse probability weighting estimator.[14,15] All fibrosis scores classified patients into low-, intermediate-, and high-risk groups on the published low and high cut-offs. For histology, we stratified the three groups as F0-2, F3, and F4. The cumulative incidence of outcomes with adjustment of competing events was estimated by Gray's method and compared by Gray's test among different risk categories (eMethods). For both the primary outcome and HCC, non-liver-related death was treated as a competing event. For hepatic decompensation, both non-liver-related death and HCC were treated as competing events.

In the serial model, we considered patients with two or more VCTE examinations. For those with multiple examinations, we selected the first and last examinations, with a maximum five-year interval, and a minimum six-month separation. We assessed the incidence of the outcomes from the last VCTE examinations onwards. Patients developing LREs between these examinations were documented but not included in the serial prediction models. Transition among risk categories based on published cut-offs was depicted using Sankey diagrams. We also evaluated the prognostic

significance of serial non-invasive tests based on their relative change between the two examinations (eMethods).

Results

Participants

From February, 2004 to January, 2023, we identified 17 949 patients with one or more VCTE examinations. After excluding 1 346 patients according to the inclusion and exclusion criteria, 16 603 patients were included in the baseline model (Figure 1). Their mean age was 52.5 years, and 57.8% were men (Table 1). 34.7% and 34.8% had diabetes and hypertension, respectively. 3 030 (18.2%) patients were from the United States or Europe, and 13 573 (81.8%) patients were from Asia. Among 3 532 patients with liver biopsy, 33.5% had F3-4 fibrosis. The median interval (interquartile range [IQR]) between liver biopsy and VCTE examinations was 28 (0-214) days.

Baseline model

At a median follow-up of 51.7 months (IQR 25.2-85.2 months), 316 (1.9%) patients developed LREs, including 139 cases of HCC and 209 cases of hepatic decompensation (eTable 1). Both the Agile 3+ and Agile 4 scores demonstrated the highest AUROC and AUPRC for predicting LREs (Figure 2A and eFigure 1); they classified fewer patients (10.2% for Agile 3+ and 8.7% for Agile 4) in the intermediate-risk group than the other fibrosis scores. The Agile 3+ and Agile 4 scores also demonstrated the highest integrated AUROC and lowest integrated Brier score (eTable 2). Likewise, in the 10 678 patients with all studied fibrosis markers available, the Agile 3+ and Agile 4 scores demonstrated highest AUROC and lowest integrated Brier score (eFigure 1-2 and eTable 2).

By pairwise comparison, the AUROC for LREs of both Agile scores was significantly higher than histological fibrosis staging and other comparator fibrosis tests at 3 and 5 years, with the exception of a similar performance between the Agile scores and LSM at 3 years (eTable 3). The calibration was excellent for both Agile scores, but was generally unsatisfactory for the simple fibrosis scores (eFigure 3 and 4). The Agile scores better reclassified patients with and without LREs at 3 and 5 years according to their risk as compared to LSM, while other non-invasive tests generally had a similar or reduced correct reclassification as compared to LSM (Table 2). Analyzed separately, all the fibrosis tests were better at the prediction of hepatic decompensation than HCC (eFigure 5, eTable 4 and 5).

Among patients with baseline Agile 3+ score <0.451, 0.451-0.678, and ≥0.679, the incidence rates of LREs were 0.7, 3.3, and 24.9 per 1 000 person-years, respectively (*P*<.001) (Figure 2B, eTable 6). Among patients with baseline Agile 4 score <0.251, 0.251-0.842, and ≥0.843, the incidence rates of LREs were 1.2, 23.5, and 105.5 per 1 000 person-years, respectively (*P*<.001). Among the noninvasive tests, the Agile 4 score classified the highest proportion (89.8%) of patients in the low-risk group with a sensitivity of 0.74 and negative predictive value of 0.99 for 3-year LREs (eTable 7). In contrast, it classified the fewest patients (1.4%) in the high-risk group, compared with 14.3% for the Agile 3+ score. eFigure 6 shows the incidence of LREs in patients categorized by histology and other non-invasive tests. Similar to the ROC analysis, AAR, BARD and FAST were the least discriminatory.

The Agile scores consistently outperformed the other non-invasive tests in predicting LREs at 3 and 5 years in subgroups stratified by age, sex, presence of diabetes, body-mass index and reliability of LSM (eFigure 7). Both Agile scores had higher AUROC in patients older than 60 years than in younger patients. The prognostic performance of the fibrosis scores was largely similar across regions (eTable 8).

Serial model

Among 16 603 patients in the baseline model, 10 920 (65.8%) patients with repeated VCTE examinations at a median interval of 15 months (IQR 11.3-27.7 months) were included in the serial model (Figure 1). The clinical characteristics at the first examination of the patients in the serial model were similar to those of patients in the baseline model (Table 1). Between the first and last VCTE examinations, the proportion of patients with diabetes and hypertension increased by around 12%. Using published cut-offs, the risk classification by Agile scores was stable when either two or three examinations were considered (Figure 3A, eFigure 8-11). Patients with a longer time interval between two tests were more likely to have increased scores at the second assessment, suggesting genuine fibrosis progression instead of variability in scores on repeated testing (eFigure 10). In general, the Agile scores and LSM had a higher stability than the other non-invasive tests (eFigure 11).

eTable 9 and 10 show the incidence of LREs in patients with serial Agile scores. In patients with high Agile 3+ score at the first examination but intermediate score at the last examination, the incidence of LREs decreased markedly to 3.3 per 1 000 person-years. A similar trend was observed for the Agile 4 score (eTable 10) and LSM (eTable 11 and 12). In contrast, patients who had worsened Agile 3+ scores at the last

examination only had a mild increase in the risk of LREs over those who had stable scores (eTable 9). eTable 13-16 show consistent results in sensitivity analyses by including only patients who had two noninvasive tests performed within an interval of 3 years.

Apart from classifying patients into crude risk categories, another way to interpret serial test results is to determine their change over time. By restricted spline curve analysis, there was a positive non-linear relationship between changes in Agile scores/LSM and the risk of LREs (eFigure 12). Regardless of baseline Agile scores and LSM, a 10% or greater relative decrease in the test results was associated with a lower risk of LREs, whereas an increase in the test results was associated with increased risk of events (Figure 3B, eTable 17-19). As expected, the greater the change in Agile scores or LSM (e.g., 30% relative change), a greater change in the incidence of LREs was also observed. Compared with patients with stable Agile scores, those with a 30% or greater relative increase in the scores had significant changes in all the components of the scores (eTable S20).

Discussion

In this large multi-center study, we showed that the Agile scores had better performance in predicting LREs in patients with MASLD than commonly used simple fibrosis scores. Although the difference in prognostication between the Agile scores and LSM might be marginal, the Agile scores were stable over time, and changes in the scores over time provide insights that can impact clinical management.

In the baseline model, both the Agile 3+ and Agile 4 scores had the highest overall accuracy in predicting LREs. Although both Agile scores had identical integrated AUROC, it should be noted that the Agile 4 score classified around 90% of patients in the low-risk group and in turn missed twice as many patients who would develop LREs as the Agile 3+ score. The Agile 4 score mainly improved classification of patients without LREs, while the Agile 3+ score improved the classification of events. This is understandable as the Agile 3+ and Agile 4 scores were designed to detect advanced fibrosis and cirrhosis, respectively.[7] Therefore, the Agile 3+ score is preferred for prognostic purposes, whereas the main value of the Agile 4 score is for the diagnosis of MASLD-related cirrhosis. It is also worth noting that the superiority of the Agile scores over LSM alone was marginal. While the calculation of the Agile scores is based on routine parameters and thus does not cost extra, clinicians who prefer to use LSM alone for the sake of simplicity can also refer to the detailed analysis on the prognostication by LSM in this study.

Analyzed separately, all non-invasive tests of fibrosis were better at predicting hepatic decompensation than HCC (eFigure 5). This can be explained by the phenomenon of HCC arising in a non-cirrhotic liver. Although hepatic decompensation almost always develops in the background of cirrhosis, HCC appears to arise from a non-cirrhotic liver more often in MASLD (around 30%) than other chronic liver diseases.[16,17]

Compared with the existing literature,[8,18] our study assigns significance to not only baseline but also changes in LSM and Agile scores. Over 80% of patients, in two or three assessments, remained within the same risk categories based on published Agile score cut-offs (Figure 3A, eFigure 8-11). MASLD progression from no to minimal

fibrosis to cirrhosis or LREs typically spans 20 years.[19] Among patients with LSM and Agile score changes, reductions were more frequent than increases. Reduced LSM might reflect true fibrosis improvement due to lifestyle changes, but most likely resulted from initial false positives, potentially explaining why decreased LSM had a greater impact on LRE risk than increases (eFigure 12). False-positive LSM has been reported in patients with factors such as extreme body build, acute hepatitis, congestive heart failure, biliary obstruction, amyloidosis, and recent food intake.[20] In a previous study with a median 18-week interval between two VCTE examinations, 35% of patients with initially high LSM had normal LSM at the second assessment, with most showing no or mild fibrosis on subsequent liver biopsy.[21] Similarly, in our study, patients with reduced LSM or Agile scores over time had a lower LRE incidence compared to those with higher readings. Therefore, patients with abnormal LSM or Agile scores should consider repeat examinations before deciding on liver biopsy or treatment.

While customary, interpreting non-invasive tests based on published cut-offs can be crude and misleading. Some individuals do not cross these thresholds despite progression or regression, while minor fluctuations near cut-offs can lead to misinterpretation. To address this, we performed a restricted spline curve analysis (eFigure 12), which revealed that Agile score and LSM changes are positively associated with LRE risk. Prior studies recommended a 20% LSM relative change for prognostication.[22,23] Our study provides detailed data on the prognostic importance of varying Agile score/LSM changes.

In comparison, serial FIB-4 has also been shown to be prognostic in the general population and hospital settings.[24,25] However, FIB-4 is inferior to LSM and other specific fibrosis biomarkers in the diagnosis of advanced fibrosis.[26] FIB-4 also has suboptimal performance at extremes of age.[27,28]

According to the US Food and Drug Administration, to replace liver histology as a surrogate endpoint in clinical trials, a biomarker should demonstrate the ability to diagnose the fibrosis stage, predict prognosis, monitor disease progression, and reflect response to treatment.[5] Based on this and other studies, VCTE and the Agile scores have already fulfilled the first three requirements, but the latter requires correlation between histological response and changes in non-invasive tests in clinical trials involving an effective treatment. There have already been efforts to fill this knowledge gap using data from several clinical trials,[29,30] and we expect an acceleration in the validation of response biomarkers when some of the ongoing phase 3 trials show positive results. Meanwhile, the existing non-invasive tests can largely replace liver biopsies in routine practice.

Limitations

The study has several limitations. First, variable patient assessment intervals affect serial data interpretation, yet we analyzed non-invasive test changes and correlation with clinical outcomes after VCTE examinations interval stratification. Second, despite a sufficient sample size for clinical outcome evaluation, the 51.7-month median follow-up may be considered short, given chronic liver diseases' lengthy progression to cirrhosis and complications.[31] Third, this was a natural history cohort. When effective treatment for steatohepatitis becomes available, studies should be

conducted to identify suitable response biomarkers. Fourth, data of this study were from tertiary referral centers. The prognostic performance of VCTE and the Agile scores should be confirmed in a more general setting in the future. Although the Agile scores were compared with a number of simple fibrosis scores, future studies should compare the Agile scores with other specific biomarkers of fibrosis and/or steatohepatitis such as the enhanced liver fibrosis, NIS4 and NIS2+ scores.

Conclusions

The VCTE-based Agile scores are highly accurate in predicting LREs in patients with MASLD. In the short- to medium-term, the Agile scores have high stability on repeated testing. In the minority of patients with an early change in Agile scores, the lower score between two serial measurements more faithfully reflects the risk of LREs. In this situation, repeating Agile score measurements or testing another specific fibrosis biomarker should be contemplated before making decision on liver biopsy or treatment.

References

1. Wong VW, Ekstedt M, Wong GL, Hagstrom H. Changing epidemiology, global trends and implications for outcomes of NAFLD. *J Hepatol.* 2023;79(3):842-852.
2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet (London, England).* 2021;397(10290):2212-2224.
3. Younossi ZM, Henry L, Bush H, Mishra A. Clinical and Economic Burden of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Clin Liver Dis.* 2018;22(1):1-10.
4. Taylor RS, Taylor RJ, Bayliss S, et al. Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. *Gastroenterology.* 2020;158(6):1611-1625.e1612.
5. Sanyal AJ, Castera L, Wong VW. Noninvasive Assessment of Liver Fibrosis in NAFLD. *Clin Gastroenterol Hepatol.* 2023;21(8):2026-2039.
6. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno VIIF. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol.* 2022;76(4):959-974.
7. Sanyal AJ, Foucquier J, Younossi ZM, et al. Enhanced diagnosis of advanced fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile scores. *Journal of hepatology.* 2023;78(2):247-259.
8. 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. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* 2023;21(5):1293-1302.e1295.
9. Zhang X, Yip TC, Wong GL, et al. Clinical care pathway to detect advanced liver disease in patients with type 2 diabetes through automated fibrosis score calculation and electronic reminder messages: a randomised controlled trial. *Gut.* 2023.
10. Newsome PN, Sasso M, Deeks JJ, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *The lancet Gastroenterology & hepatology.* 2020;5(4):362-373.
11. Wong VW, Adams LA, de Ledinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. *Nat Rev Gastroenterol Hepatol.* 2018;15(8):461-478.
12. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in medicine.* 2013;32(30):5381-5397.
13. Yuan Y, Zhou QM, Li B, Cai H, Chow EJ, Armstrong GT. A threshold-free summary index of prediction accuracy for censored time to event data. *Stat Med.* 2018;37(10):1671-1681.
14. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in medicine.* 2011;30(1):11-21.

15. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Statistics in medicine.* 2013;32(14):2430-2442.
16. Mittal S, El-Serag HB, Sada YH, et al. Hepatocellular Carcinoma in the Absence of Cirrhosis in United States Veterans is Associated With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol.* 2016;14(1):124-131 e121.
17. Chan TT, Chan WK, Wong GL, et al. Positive Hepatitis B Core Antibody Is Associated With Cirrhosis and Hepatocellular Carcinoma in Nonalcoholic Fatty Liver Disease. *Am J Gastroenterol.* 2020;115(6):867-875.
18. Mozes FE, Lee JA, Vali Y, et al. 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. *Lancet Gastroenterol Hepatol.* 2023;8(8):704-713.
19. Hagstrom 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(6):1265-1273.
20. Wai JW, Fu C, Wong VW. Confounding factors of non-invasive tests for nonalcoholic fatty liver disease. *Journal of gastroenterology.* 2020;55(8):731-741.
21. Chow JC, Wong GL, Chan AW, et al. Repeating measurements by transient elastography in non-alcoholic fatty liver disease patients with high liver stiffness. *J Gastroenterol Hepatol.* 2019;34(1):241-248.
22. Loomba R, Huang DQ, Sanyal AJ, et al. Liver stiffness thresholds to predict disease progression and clinical outcomes in bridging fibrosis and cirrhosis. *Gut.* 2023;72(3):581-589.
23. Semmler G, Yang Z, Fritz L, et al. Dynamics in Liver Stiffness Measurements Predict Outcomes in Advanced Chronic Liver Disease. *Gastroenterology.* 2023.
24. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. *Journal of hepatology.* 2020;73(5):1023-1029.
25. Cholankeril G, Kramer JR, Chu J, et al. Longitudinal changes in fibrosis markers are associated with risk of cirrhosis and hepatocellular carcinoma in non-alcoholic fatty liver disease. *Journal of hepatology.* 2023;78(3):493-500.
26. Mozes 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(5):1006-1019.
27. McPherson S, Hardy T, Dufour JF, et al. Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. *The American journal of gastroenterology.* 2017;112(5):740-751.
28. Wong VW, Tak WY, Goh GBB, et al. Performance of Noninvasive Tests of Fibrosis Among Asians, Hispanic, and non-Hispanic Whites in the STELLAR Trials. *Clin Gastroenterol Hepatol.* 2023;21(1):90-102 e106.
29. Rinella ME, Dufour JF, Anstee QM, et al. Non-invasive evaluation of response to obeticholic acid in patients with NASH: Results from the REGENERATE study. *Journal of hepatology.* 2022;76(3):536-548.
30. Wai-Sun Wong V, Anstee QM, Nitze LM, et al. FibroScan-aspartate aminotransferase (FAST) score for monitoring histological improvement in non-alcoholic steatohepatitis activity during semaglutide treatment: post-hoc

analysis of a randomised, double-blind, placebo-controlled, phase 2b trial. *EClinicalMedicine.* 2023;66:102310.

31. Zhang X, Yip TC, Tse YK, et al. Duration of type 2 diabetes and liver-related events in nonalcoholic fatty liver disease: A landmark analysis. *Hepatology.* 2023.

Table 1: **Clinical characteristics of the cohorts in the baseline and serial models**

Characteristics	Baseline model	Serial model		*P* value[b]
		First test	Last test	
	N = 16 603	N = 10 920		
Age (years)	52.5 (13.7)	52.3 (13.5)	54.4 (13.6)	<.001
Female sex, n (%)	7 003 (42.2)	4 629 (42.4)	4 629 (42.4)	-
Male sex, n (%)	9 600 (57.8)	6 291 (57.6)	6 291 (57.6)	-
BMI (kg/m^2)	27.0 (24.5-30.0)	27.0 (24.5-30.0)	27.0 (24.6-30.1)	.09
Diabetes, n (%)	5 761 (34.7)	3 944 (36.1)	5 311 (48.6)	<.001
Hypertension, n (%)	5 769 (34.8)	3 925 (35.9)	5 291 (48.5)	<.001
ALT (IU/L)	37 (23-62)	36 (23-61)	30 (20-48)	<.001
AST (IU/L)	31 (23-47)	31 (22-46)	27 (21-38)	<.001
GGT (IU/L)	44 (27-79)	43 (26-76)	36 (23-63)	<.001
Albumin (g/L)	44.4 (3.9)	44.7 (3.5)	44.8 (3.6)	.02
Total bilirubin (µmol/L)	12.0 (8.6-15.4)	12.0 (8.6-15.4)	12.0 (10.0-17.1)	<.001
Platelet (×10^9/L)	237 (198-280)	238 (199-281)	235 (196-279)	<.001
Creatinine (µmol/L)	72 (60-83)	72 (60-83)	72 (61-84)	<.001
FibroScan				
Liver stiffness measurement (kPa)	6.0 (4.7-8.5)	6.0 (4.6-8.3)	5.5 (4.5-7.7)	<.001
Controlled attenuation parameter (dB/m)	303 (273-334)	302 (273-334)	295 (262-328)	<.001
Non-invasive tests[a]				
Agile 3+	0.16 (0.06-0.44)	0.17 (0.06-0.43)	0.21 (0.08-0.48)	<.001
Agile 4	0.01 (0.00-0.06)	0.01 (0.00-0.05)	0.01 (0.00-0.05)	.23
FibroScan-AST	0.28 (0.12-0.52)	0.27 (0.12-0.51)	0.19 (0.09-0.41)	<.001
Fibrosis-4 index	1.11 (0.74-1.71)	1.13 (0.76-1.71)	1.18 (0.81-1.75)	<.001
NAFLD fibrosis score	-1.99 (-3.03--0.78)	-1.98 (-3.00--0.83)	-1.62 (-2.67--0.49)	<.001
AST-to-platelets ratio index	0.33 (0.23-0.54)	0.33 (0.23-0.52)	0.30 (0.22-0.45)	<.001
AST-to-ALT ratio	0.83 (0.62-1.12)	0.84 (0.64-1.14)	0.90 (0.69-1.20)	<.001
BARD	2 (1-3)	2 (1-3)	2 (1-3)	<.001
Fibrosis stage[b] **N = 3 532**				
0	576 (16.3)	-	-	-
1	1 189 (33.7)	-	-	-
2	585 (16.6)	-	-	-
3	744 (21.1)	-	-	-
4	438 (12.4)	-	-	-
Median follow-up duration (months)	51.7 (25.2-85.2)		34.0 (12.4-55.9)	

Data are n (%), mean (standard deviation), or median (interquartile range).

[a] The formulas for the calculation of the non-invasive tests are presented in the Supplement page 3-4.

[b] Fibrosis stage (0-4) according to the NASH CRN system. Stage 0, no fibrosis; Stage 1, centrilobular pericellular fibrosis; Stage 2: centrilobular and periportal fibrosis; Stage 3: bridging fibrosis; Stage 4, cirrhosis.

[c] Paired samples tests between the first and last tests in the serial model.

Liver stiffness measurement is a non-invasive method to evaluate liver fibrosis, using transient elastography to measure liver stiffness, which helps in assessing the extent of fibrosis; Controlled attenuation parameter quantifies liver steatosis non-invasively, by measuring the attenuation of ultrasound waves through the liver, providing an indicator of fat levels.

Abbreviations: AST, aspartate aminotransferase. ALT, alanine aminotransferase. BMI, body-mass index. GGT, gamma-glutamyl transpeptidase. NAFLD, non-alcoholic fatty liver disease. VCTE, vibration-controlled transient elastography.

627 ***Table 2:*** **Paired comparisons of the Agile scores and other non-invasive tests versus liver stiffness measurement (LSM) on the net reclassification**
628 **improvement (NRI) for the prediction of 3-year and 5-year liver-related events in the baseline model**

	3-year liver-related events			5-year liver-related events		
Tests	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**	**Event NRI (95% CI)**	**Non-event NRI (95% CI)**	**Overall NRI (95% CI)**
Agile 3+ (N=12 948)	0.31 (0.14–0.49)	0.57 (0.53–0.61)	0.88 (0.68–1.08)	0.41 (0.27–0.54)	0.61 (0.57–0.65)	1.02 (0.86–1.18)
Agile 4 (N=12 948)	0.19 (0.02–0.36)	0.81 (0.79–0.83)	1.00 (0.82–1.18)	0.30 (0.14–0.43)	0.84 (0.82–0.85)	1.13 (0.96–1.28)
Liver stiffness measurement	Reference	Reference	Reference	Reference	Reference	Reference
Fibrosis-4 index (N=12 950)	-0.30 (-0.46–-0.04)	-0.78 (-0.81–-0.54)	-1.08 (-1.25–-0.63)	-0.31 (-0.46–-0.03)	-0.78 (-0.81–-0.51)	-1.09 (-1.24–-0.57)
NAFLD fibrosis score (N=12 064)	-0.18 (-0.37–0.04)	-0.57 (-0.69–0.12)	-0.75 (-0.98–0.06)	-0.16 (-0.28–0.10)	-0.52 (-0.64–0.12)	-0.68 (-0.87–0.20)
AST-to-platelets ratio index (N=12 975)	-0.40 (-0.56–-0.20)	-0.79 (-0.82–-0.75)	-1.19 (-1.35–-0.98)	-0.43 (-0.55–-0.26)	-0.80 (-0.83–-0.77)	-1.23 (-1.35–-1.05)
FibroScan-AST (N=11 541)	-0.16 (-0.37–0.05)	0.24 (0.14–0.31)	0.08 (-0.19–0.34)	-0.10 (-0.26–0.07)	0.26 (0.17–0.34)	0.17 (-0.07–0.38)
AST-to-platelets ratio index (N=13 159)	-0.37 (-0.53–-0.22)	-0.78 (-0.80–-0.70)	-1.15 (-1.31–-0.96)	-0.44 (-0.55–-0.30)	-0.78 (-0.81–-0.69)	-1.22 (-1.34–-1.02)
BARD (N=12 498)	-0.36 (-0.51–-0.18)	-0.14 (-0.32–-0.11)	-0.50 (-0.74–-0.31)	-0.32 (-0.46–-0.17)	-0.12 (-0.27–-0.09)	-0.44 (-0.68–-0.28)

629 Event NRI referred to the net proportion of LREs assigned a higher risk, which ranged from -1 to +1. Non-event NRI referred to the net proportion of non-LREs assigned a lower risk, which
630 ranged from -1 to +1. Overall NRI was the simple sum of event NRI and non-event NRI, which was a crude summary of event NRI and non-event NRI, ranged from -2 to +2. A positive NRI
631 referred to an improvement in correct reclassification, while a negative NRI referred to a reduction in correct reclassification. The 95% CI for NRI was estimated using 1,000 bootstrap samples.
632 Abbreviations: AST, aspartate aminotransferase. ALT, alanine aminotransferase. CI, confidence interval. NAFLD, non-alcoholic fatty liver disease. NRI, net reclassification improvement.
633

634

Figure legends

Figure 1: **Study participant flow**

Abbreviations: MASLD, metabolic dysfunction-associated steatotic liver disease. VCTE, vibration-controlled transient elastography.

Figure 2: **Prediction of liver-related events by non-invasive tests and liver histology**

A, AUROC and AUPRC for the prediction of liver-related events at 3 and 5 years. B, Cumulative incidence of liver-related events stratified by Agile 3+ score in the baseline model.

In panel B, the cut points for Agile 3+ score were based on the original publication. The low cut point (0.451) achieved sensitivity of ⩾85% to rule-out patients of fibrosis stage ⩾3, the high cut point (0.679) achieved specificity of ⩾90% to rule-in patients of fibrosis stage ⩾3. The median follow-up duration of each group was listed in the legend.

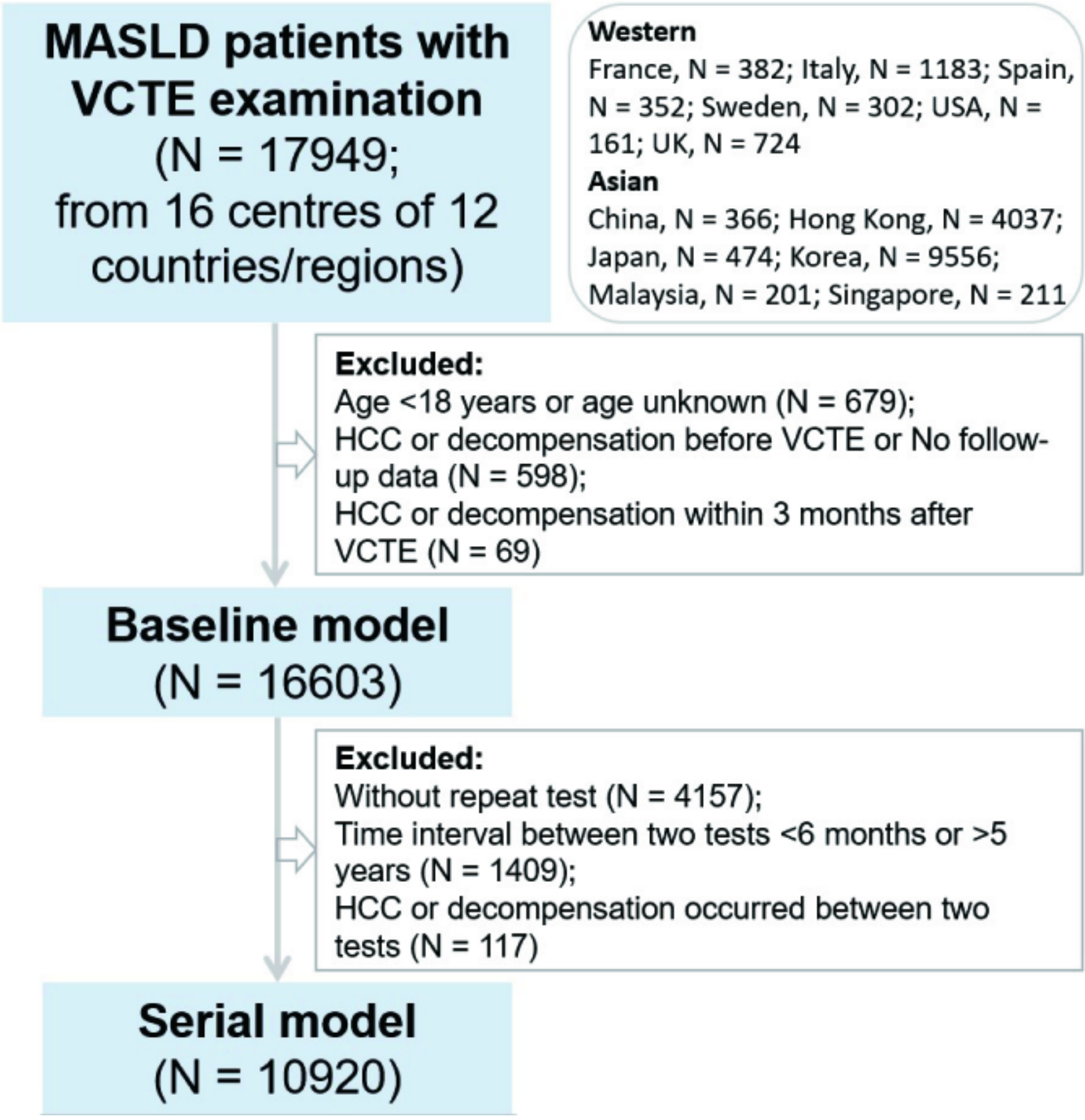
Abbreviations: AST, aspartate aminotransferase. ALT, alanine aminotransferase ratio. AUROC, area under the receiver-operating characteristic curve. AUPRC, area under the precision-recall curve. CI, confidence interval. LRE, liver-related event. NAFLD, non-alcoholic fatty liver disease

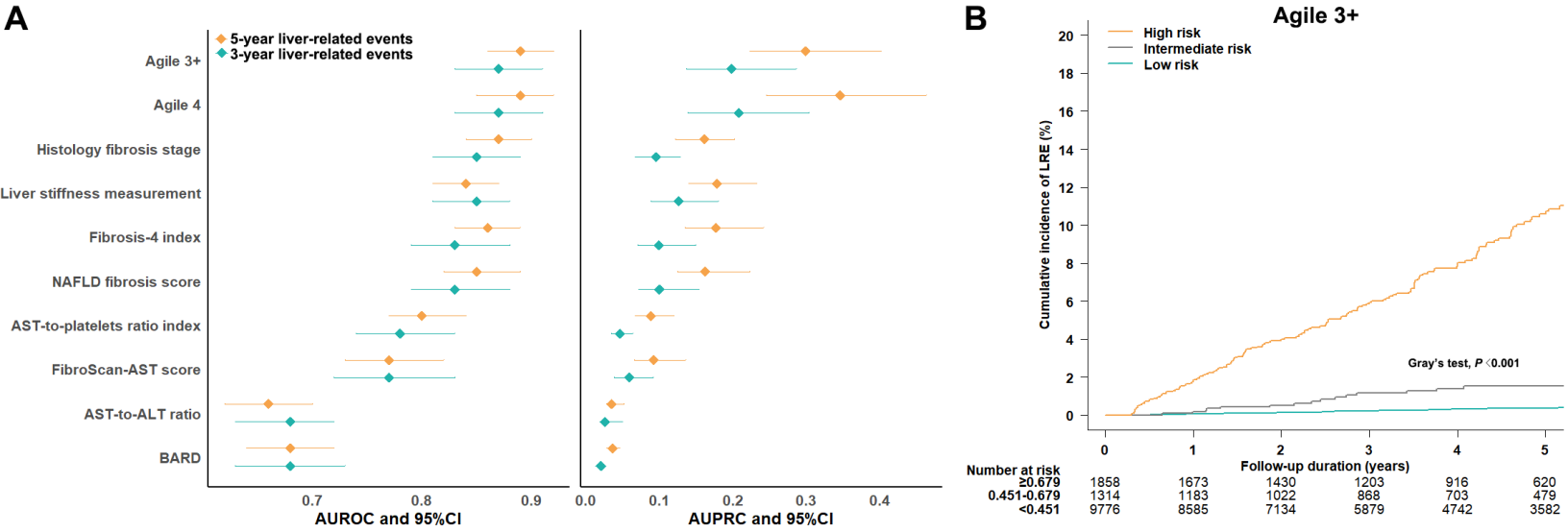
Figure 3: **Agile 3+ score in serial model.**

A, Change in the Agile 3+ between two vibration-controlled transient elastography examinations. B, Relative change in the Agile 3+ score and incident liver-related events after the last test.

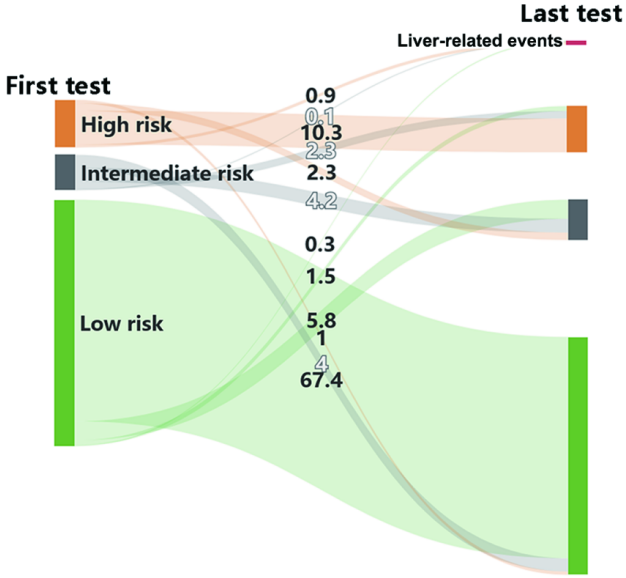
In panel A, the numbers in the middle represent the percentages of patients in each group. Patients who developed liver-related events before the last examination are shown in the top of the Sankey diagram.

Abbreviations: CI, confidence interval. LREs, liver-related events. PY, person-year.


MASLD patients with VCTE examination
(N = 17949;
from 16 centres of 12 countries/regions)
Western
France, N = 382; Italy, N = 1183; Spain, N = 352; Sweden, N = 302; USA, N = 161; UK, N = 724
Asian
China, N = 366; Hong Kong, N = 4037; Japan, N = 474; Korea, N = 9556; Malaysia, N = 201; Singapore, N = 211
Excluded:
Age <18 years or age unknown (N = 679);
HCC or decompensation before VCTE or No follow-up data (N = 598);
HCC or decompensation within 3 months after VCTE (N = 69)
Baseline model
(N = 16603)
Excluded:
Without repeat test (N = 4157);
Time interval between two tests <6 months or >5 years (N = 1409);
HCC or decompensation occurred between two tests (N = 117)
Serial model
(N = 10920)

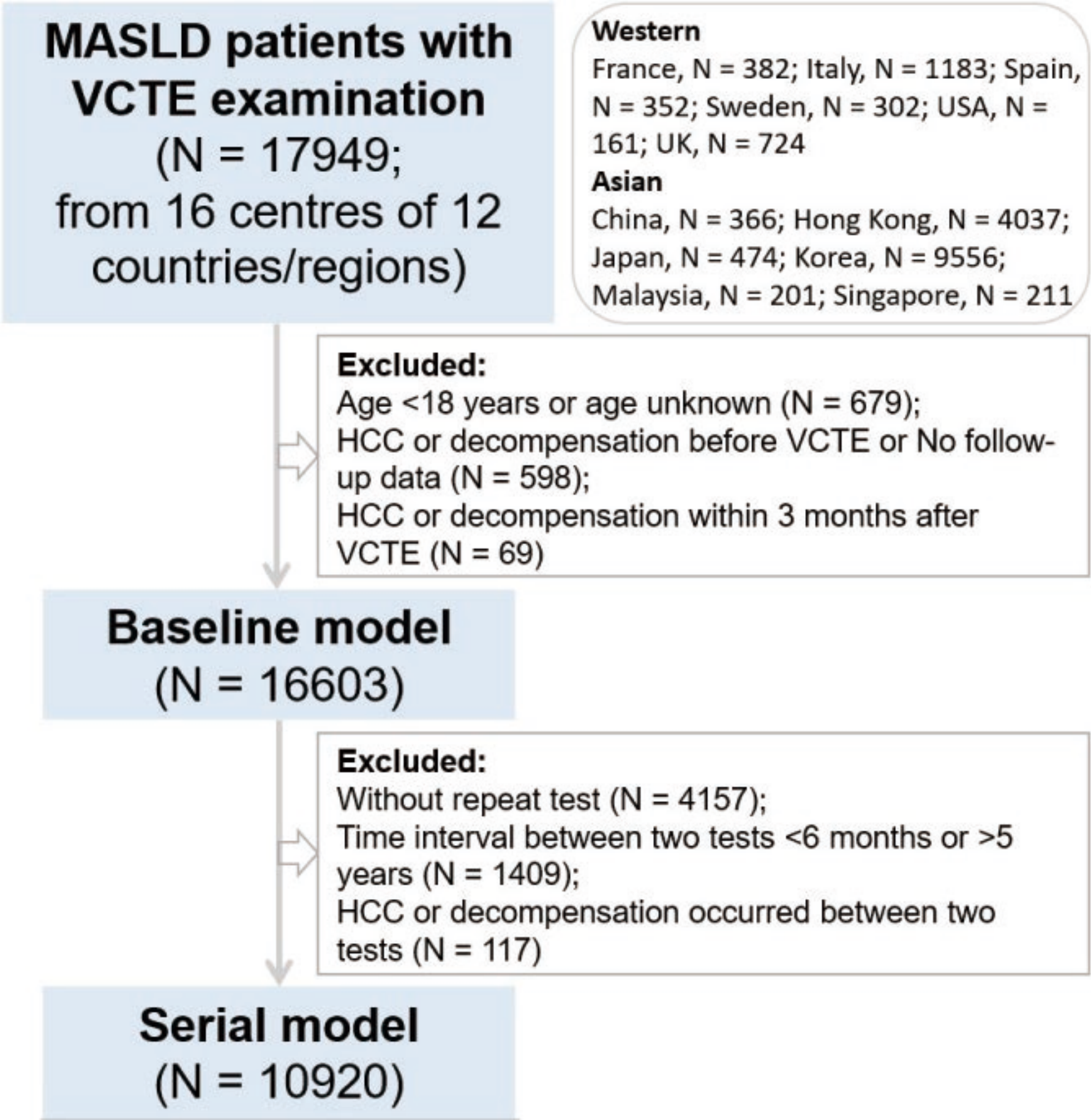

A
5-year liver-related events
3-year liver-related events
Agile 3+
Agile 4
Histology fibrosis stage
Liver stiffness measurement
Fibrosis-4 index
NAFLD fibrosis score
AST-to-platelets ratio index
FibroScan-AST score
AST-to-ALT ratio
BARD
0.7
0.8
0.9
AUROC and 95%CI
0.0
0.1
0.2
0.3
0.4
AUPRC and 95%CI
B
Agile 3+
High risk
Intermediate risk
Low risk
Cumulative incidence of LRE (%)
Gray's test, P <0.001
Follow-up duration (years)
Number at risk
≥0.679 1858 1673 1430 1203 916 620
0.451-0.679 1314 1183 1022 868 703 479
<0.451 9776 8585 7134 5879 4742 3582

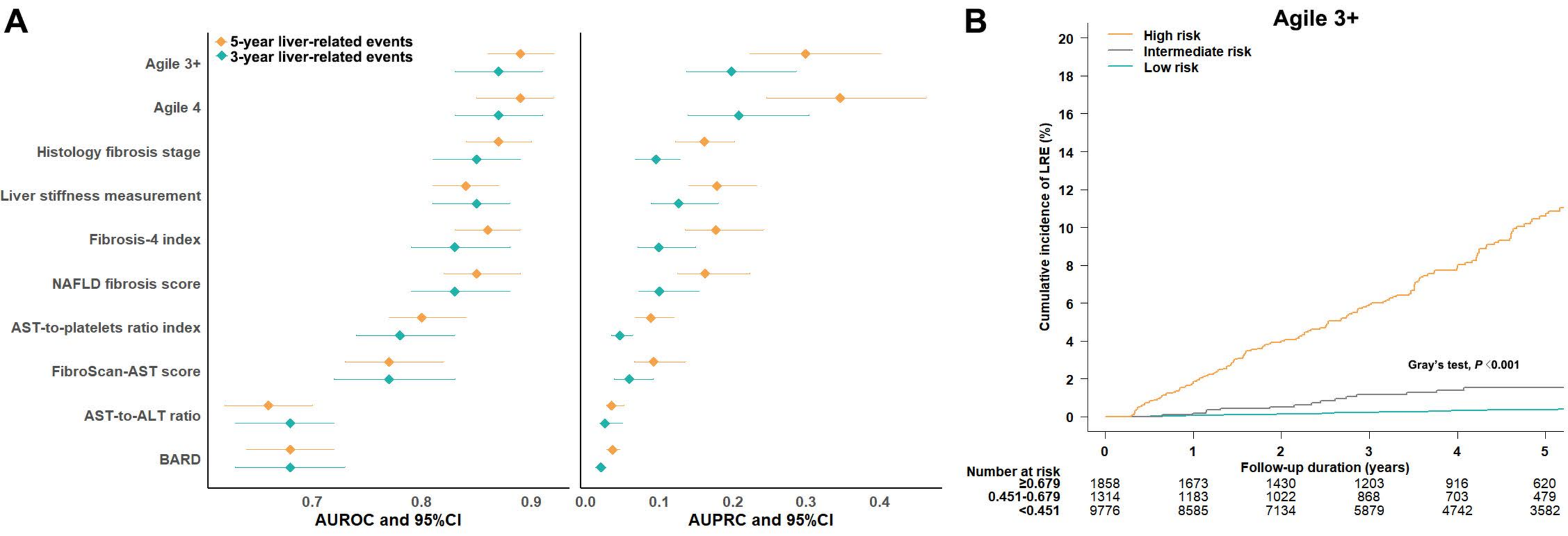
A



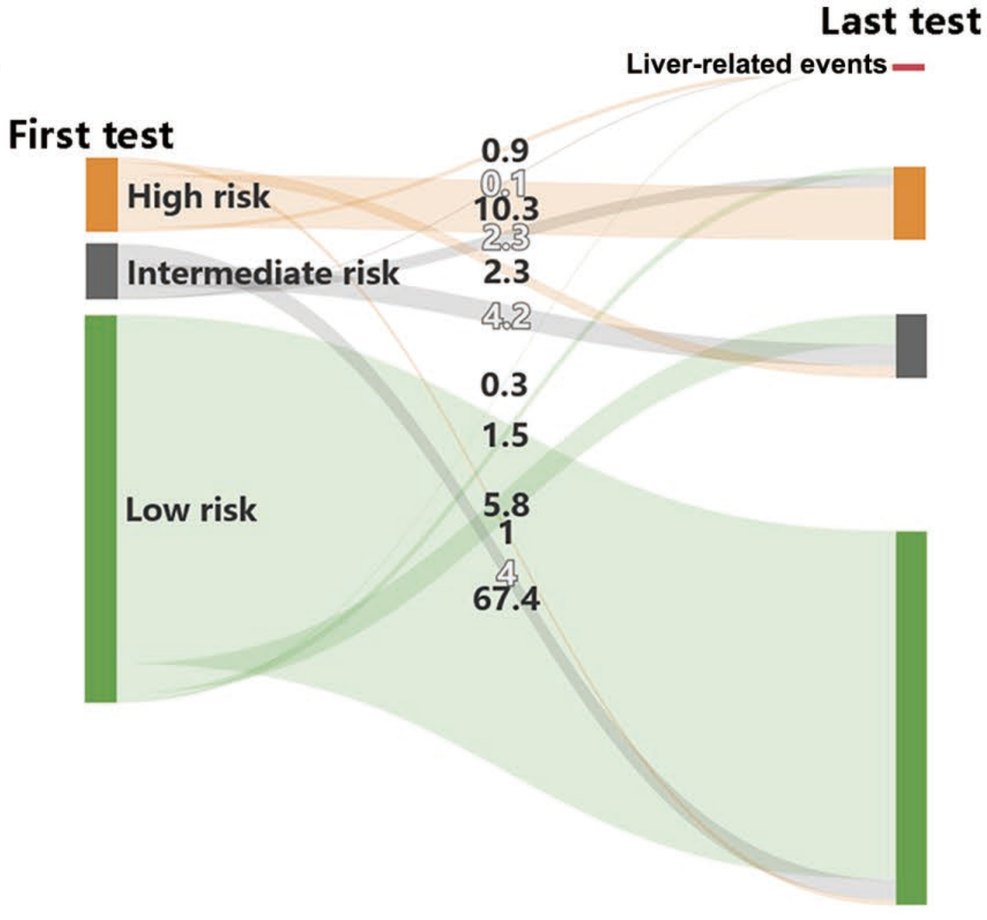
B

First test	Relative change 10%	N (%)	5-year LRE (%)	Relative change 20%	N (%)	5-year LRE (%)	Relative change 30%	N (%)	5-year LRE (%)
High risk	Decreasing >10%	403 (4.6)	1.8 (0.7-3.9)	Decreasing >20%	263 (3.0)	0.5 (0.1-2.7)	Decreasing >30%	162 (1.8)	-
	Stable	665 (7.6)	14.1 (10.2-18.6)	Stable	888 (10.2)	12.2 (9.0-15.9)	Stable	1 029 (11.8)	11.9 (8.9-15.4)
	Increasing ≥10%	129 (1.4)	19.7 (7.9-35.3)	Increasing ≥20%	46 (0.5)	31.0 (7.6-58.6)	Increasing ≥30%	6 (0.1)	33.3 (-)
Intermediate risk	Decreasing >10%	437 (5.0)	0.3 (0.1-1.5)	Decreasing >20%	341 (3.9)	0.4 (0.1-1.9)	Decreasing >30%	246 (2.8)	0.5 (0.1-2.7)
	Stable	195 (2.2)	0.6 (0.1-2.9)	Stable	366 (4.2)	0.6 (0.1-2.1)	Stable	528 (6.0)	0.4 (0.1-1.5)
	Increasing ≥10%	291 (3.3)	0.9 (0.2-3.9)	Increasing ≥20%	216 (2.4)	0.6 (0.1-3.0)	Increasing ≥30%	149 (1.7)	0.9 (0.1-4.3)
Low risk	Decreasing >10%	2 179 (25.0)	0.3 (0.1-1.0)	Decreasing >20%	1 763 (20.2)	0.4 (0.1-1.3)	Decreasing >30%	1 359 (15.6)	0.1 (0.1-0.5)
	Stable	698 (8.0)	-	Stable	1 463 (16.8)	0.3 (0.1-0.9)	Stable	2 194 (25.2)	0.4 (0.1-1.1)
	Increasing ≥10%	3 703 (42.5)	0.4 (0.2-0.8)	Increasing ≥20%	3 354 (38.5)	0.4 (0.1-0.8)	Increasing ≥30%	3 027 (34.7)	0.4 (0.2-0.9)


MASLD patients with VCTE examination
(N = 17949;
from 16 centres of 12 countries/regions)
Western
France, N = 382; Italy, N = 1183; Spain, N = 352; Sweden, N = 302; USA, N = 161; UK, N = 724
Asian
China, N = 366; Hong Kong, N = 4037; Japan, N = 474; Korea, N = 9556; Malaysia, N = 201; Singapore, N = 211
Excluded:
Age <18 years or age unknown (N = 679);
HCC or decompensation before VCTE or No follow-up data (N = 598);
HCC or decompensation within 3 months after VCTE (N = 69)
Baseline model
(N = 16603)
Excluded:
Without repeat test (N = 4157);
Time interval between two tests <6 months or >5 years (N = 1409);
HCC or decompensation occurred between two tests (N = 117)
Serial model
(N = 10920)


A
5-year liver-related events
3-year liver-related events
Agile 3+
Agile 4
Histology fibrosis stage
Liver stiffness measurement
Fibrosis-4 index
NAFLD fibrosis score
AST-to-platelets ratio index
FibroScan-AST score
AST-to-ALT ratio
BARD
0.7
0.8
0.9
AUROC and 95%CI
0.0
0.1
0.2
0.3
0.4
AUPRC and 95%CI
B
Agile 3+
High risk
Intermediate risk
Low risk
Cumulative incidence of LRE (%)
0
2
4
6
8
10
12
14
16
18
20
Gray's test, P <0.001
0
1
2
3
4
5
Follow-up duration (years)
Number at risk
≥0.679 1858 1673 1430 1203 916 620
0.451-0.679 1314 1183 1022 868 703 479
<0.451 9776 8585 7134 5879 4742 3582

A



B

First test	Relative change 10%	N (%)	5-year LRE (%)	Relative change 20%	N (%)	5-year LRE (%)	Relative change 30%	N (%)	5-year LRE (%)
High risk	Decreasing >10%	403 (4.6)	1.8 (0.7-3.9)	Decreasing >20%	263 (3.0)	0.5 (0.1-2.7)	Decreasing >30%	162 (1.8)	-
	Stable	665 (7.6)	14.1 (10.2-18.6)	Stable	888 (10.2)	12.2 (9.0-15.9)	Stable	1 029 (11.8)	11.9 (8.9-15.4)
	Increasing ≥10%	129 (1.4)	19.7 (7.9-35.3)	Increasing ≥20%	46 (0.5)	31.0 (7.6-58.6)	Increasing ≥30%	6 (0.1)	33.3 (-)
Intermediate risk	Decreasing >10%	437 (5.0)	0.3 (0.1-1.5)	Decreasing >20%	341 (3.9)	0.4 (0.1-1.9)	Decreasing >30%	246 (2.8)	0.5 (0.1-2.7)
	Stable	195 (2.2)	0.6 (0.1-2.9)	Stable	366 (4.2)	0.6 (0.1-2.1)	Stable	528 (6.0)	0.4 (0.1-1.5)
	Increasing ≥10%	291 (3.3)	0.9 (0.2-3.9)	Increasing ≥20%	216 (2.4)	0.6 (0.1-3.0)	Increasing ≥30%	149 (1.7)	0.9 (0.1-4.3)
Low risk	Decreasing >10%	2 179 (25.0)	0.3 (0.1-1.0)	Decreasing >20%	1 763 (20.2)	0.4 (0.1-1.3)	Decreasing >30%	1 359 (15.6)	0.1 (0.1-0.5)
	Stable	698 (8.0)	-	Stable	1 463 (16.8)	0.3 (0.1-0.9)	Stable	2 194 (25.2)	0.4 (0.1-1.1)
	Increasing ≥10%	3 703 (42.5)	0.4 (0.2-0.8)	Increasing ≥20%	3 354 (38.5)	0.4 (0.1-0.8)	Increasing ≥30%	3 027 (34.7)	0.4 (0.2-0.9)