

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](https://doi.org/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.

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

67 25. Echosens, Paris, France

68 26. Department of Endocrinology, First Affiliated Hospital of Wenzhou Medical
69 University, Wenzhou, China

70 *These authors contributed equally to this work.

71

72 **Correspondence:**

73 Prof Seung Up Kim

74 Department of Internal Medicine, Yonsei University College of Medicine, Seoul,
75 Republic of Korea

76 Tel: 82-2-2228-1944

77 Email: KSUKOREA@yuhs.ac

78

79 Prof Vincent Wong

80 Department of Medicine and Therapeutics, Prince of Wales Hospital, Shatin, Hong
81 Kong, China

82 Tel: 852-35054205

83 Email: wongv@cuhk.edu.hk

84

85 **Word count:** 2997

86 **Number of tables:** 2

87 **Number of figures:** 3

88

89 Author contributions:

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

99

100 Role of the funding source

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

105

106 Data sharing statement

107 Data are available upon reasonable request to corresponding authors.

108

109 Potential conflict-of-interest statements:

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

120 advisory board member for Bristol-Myers-Squibb, Intercept, Pfizer, MSD, and Novo
121 Nordisk. His institution has received research funding from Diafir, Echosens,
122 Intercept, Inventiva, and Siemens. JLC served as a consultant and speaker for
123 Echosens, Gilead Sciences, and AbbVie. GB-BG served as a consultant for Roche and
124 Ionis Pharmaceuticals; and a speaker for Echosens, Viatrix, Abbott and Novo Nordisk.
125 WKC served as a consultant for Roche, AbbVie, Novo Nordisk, and Boehringer
126 Ingelheim; and a speaker for Echosens, Hisky Medical, and Viatrix. AJS served as a
127 consultant for 89Bio, Akero, Allergan, Alnylam Pharmaceuticals, Amgen Inc,
128 AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, Genentech, Gilead
129 Sciences, Histoindex, Intercept Pharmaceuticals, Inventiva, Madrigal, Merck,
130 Novartis, Novo Nordisk, Pfizer, Poxel, Salix Pharmaceuticals, Siemens, Sun
131 Pharmaceutical Industries Inc, Terns, and Valeant Pharmaceuticals; and a data safety
132 monitoring board or advisory board member for Bard Peripheral Vascular Inc, NGM
133 Biopharmaceuticals, and Sequana. He has received research funding from Albireo,
134 Allergan, Echosens, Eli Lilly, Gilead Sciences, Intercept Pharmaceuticals,
135 Mallinckrodt LLC, Merck, Novo Nordisk, Perspectum, Pfizer, Salix Pharmaceuticals,
136 and Zydus; and holds the stocks of Durect, Exhalenz, Gen t, and Tiziana. MR-G
137 served as a consultant for Siemens; and a speaker for Siemens and Echosens. He has
138 received research funding from Siemens, Echosens, and Novo Nordisk. PN served as
139 a consultant for Novo Nordisk, Boehringer Ingelheim, Gilead Sciences, Intercept,
140 Poxel Pharmaceuticals, Pfizer, BMS, Eli Lilly, Madrigal, and GSK; and a speaker for
141 Novo Nordisk and AiCME. He has received research funding from Novo Nordisk. LC
142 served as a consultant for Echosens, Novo Nordisk, Madrigal, MSD, Pfizer, Sagimet,
143 and Siemens Healthlineers; and a speaker for Echosens, Gilead, Inventiva, and Novo
144 Nordisk. SH served as a consultant for Akero Therapeutics, Aligos, Altimmune,
145 Arrowhead, Bluejay Therapeutics, Boxer Capital, Chronwell, Echosens, Enyo,
146 Foresite Labs, Galectin, Galecto, Gilead, GSK, Hepagene, Hepion, Hepta Bio,
147 HistoIndex, Humana, Intercept, Ionis, Madrigal, Medpace, NeuroBo Pharmaceuticals,
148 Northsea, Novo Nordisk, Perspectum, Pfizer, Sonic Incytes, Sagimet, Terns, and
149 Viking; an advisory board member for Akero, Altimmune, Arrowhead, BVF Partners,
150 Chronwell, Echosens, Foresite Labs, Galectin, Genfit, Gilead, Hepion, Hightide,
151 HistoIndex, Humana, Intercept, Madrigal, Medpace, Metacrine, NGM Bio, Northsea,
152 Novo Nordisk, PathAI, Poxel, Sagimet Biosciences, Sonic Incytes Medical Corp, and
153 Terns. He holds the stocks of Akero Therapeutics, Chronwell, Cirius, Galectin, Genfit,

154 Hepion, Hepta Bio, HistoIndex, Metacrine, NGM Bio, Northsea, and Sonic Incytes;
155 and grants and contracts from Akero Therapeutics, Altimune, Axcella, BMS,
156 Corcept, Cymabay, Enyo, Galectin, Genentech, Genfit, Gilead, GSK, Hepion,
157 Hightide, Immuron, Intercept, Inventiva, Ionis, Madrigal, NGM Bio, Novartis, Novo
158 Nordisk, Northsea, Pfizer, Poxel, Sagimet, Terns, and Viking. CF-P is an employee of
159 Echosens. GL-HW served as a consultant for AstraZeneca, Gilead Sciences, and
160 Janssen; and a speaker for Abbott, Abbvie, Bristol-Myers Squibb, Echosens, Furui,
161 Gilead Sciences, and Roche. She has received research funding from Gilead Sciences.
162 MS-WC is an employee of Echosens. VW-SW served as a consultant or advisory
163 board member for AbbVie, Boehringer Ingelheim, Echosens, Gilead Sciences,
164 Intercept, Inventiva, Novo Nordisk, Pfizer, Sagimet Biosciences, TARGET
165 PharmaSolutions, and Visirna; and a speaker for Abbott, AbbVie, Gilead Sciences,
166 Novo Nordisk, and Unilab. He has received a research grant from Gilead Sciences,
167 and is a co-founder of Illuminatio Medical Technology.
168

169 **Key Points**

170 **Question**

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

174

175 **Findings**

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

181

182 **Meaning**

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

186

187

188 **Abstract**

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

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

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

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

203 **Results:** 16 603 patients underwent VCTE examination at baseline. At a median
204 follow-up of 51.7 months, 316 (1.9%) patients developed LREs. Both Agile 3+ and
205 Agile 4 scores classified fewer patients between the low and high cutoffs than most
206 fibrosis scores and achieved the highest discriminatory power in predicting LREs
207 (integrated area under time-dependent receiver-operating characteristic curve 0.89).
208 10 920 patients had repeated VCTE at a median interval of 15 months and were
209 included in the serial analysis. 81.9% and 92.1% of patients had stable Agile 3+ and
210 Agile 4 scores (same risk categories at both assessments). The incidence of LREs was
211 0.6 and 30.1 per 1 000 person-years in patients with persistently low and high Agile
212 3+ scores, respectively. In patients with high Agile 3+ score at baseline, a decrease in
213 the score by more than 20% was associated with substantial reduction in the risk of
214 LREs. A similar trend was observed for the Agile 4 score, though it missed more
215 LREs in the low-risk group.

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

218

219 **Introduction**

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

226

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

241

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

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

246

247 **Methods**

248 **Study design and participants**

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

259

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

264

265 **Assessments**

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

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

273

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

282

283 **Outcomes**

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

290

291 **Statistical analysis**

292 All statistical analyses were performed using R software (version 4.2.2; R Core Team
293 2022). In the baseline model, the baseline date was defined as the date of the first non-

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

310

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

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

320

321 **Results**

322 **Participants**

323 From February, 2004 to January, 2023, we identified 17 949 patients with one or more
324 VCTE examinations. After excluding 1 346 patients according to the inclusion and
325 exclusion criteria, 16 603 patients were included in the baseline model (Figure 1).

326 Their mean age was 52.5 years, and 57.8% were men (Table 1). 34.7% and 34.8% had
327 diabetes and hypertension, respectively. 3 030 (18.2%) patients were from the United
328 States or Europe, and 13 573 (81.8%) patients were from Asia. Among 3 532 patients
329 with liver biopsy, 33.5% had F3-4 fibrosis. The median interval (interquartile range
330 [IQR]) between liver biopsy and VCTE examinations was 28 (0-214) days.

331

332 **Baseline model**

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

343

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

354

355 Among patients with baseline Agile 3+ score <0.451 , $0.451-0.678$, and ≥ 0.679 , the
356 incidence rates of LREs were 0.7, 3.3, and 24.9 per 1 000 person-years, respectively
357 ($P<.001$) (Figure 2B, eTable 6). Among patients with baseline Agile 4 score <0.251 ,
358 $0.251-0.842$, and ≥ 0.843 , the incidence rates of LREs were 1.2, 23.5, and 105.5 per 1
359 000 person-years, respectively ($P<.001$). Among the noninvasive tests, the Agile 4
360 score classified the highest proportion (89.8%) of patients in the low-risk group with a
361 sensitivity of 0.74 and negative predictive value of 0.99 for 3-year LREs (eTable 7).
362 In contrast, it classified the fewest patients (1.4%) in the high-risk group, compared
363 with 14.3% for the Agile 3+ score. eFigure 6 shows the incidence of LREs in patients
364 categorized by histology and other non-invasive tests. Similar to the ROC analysis,
365 AAR, BARD and FAST were the least discriminatory.

366

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

372

373 **Serial model**

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

386

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

392 examination only had a mild increase in the risk of LREs over those who had stable
393 scores (eTable 9). eTable 13-16 show consistent results in sensitivity analyses by
394 including only patients who had two noninvasive tests performed within an interval of
395 3 years.

396

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

408

409 **Discussion**

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

415

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

430

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

436

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

441 fibrosis to cirrhosis or LREs typically spans 20 years.¹⁹ Among patients with LSM
442 and Agile score changes, reductions were more frequent than increases. Reduced
443 LSM might reflect true fibrosis improvement due to lifestyle changes, but most likely
444 resulted from initial false positives, potentially explaining why decreased LSM had a
445 greater impact on LRE risk than increases (eFigure 12). False-positive LSM has been
446 reported in patients with factors such as extreme body build, acute hepatitis,
447 congestive heart failure, biliary obstruction, amyloidosis, and recent food intake.²⁰ In
448 a previous study with a median 18-week interval between two VCTE examinations,
449 35% of patients with initially high LSM had normal LSM at the second assessment,
450 with most showing no or mild fibrosis on subsequent liver biopsy.²¹ Similarly, in our
451 study, patients with reduced LSM or Agile scores over time had a lower LRE
452 incidence compared to those with higher readings. Therefore, patients with abnormal
453 LSM or Agile scores should consider repeat examinations before deciding on liver
454 biopsy or treatment.

455

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

464

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

469

470 According to the US Food and Drug Administration, to replace liver histology as a
471 surrogate endpoint in clinical trials, a biomarker should demonstrate the ability to
472 diagnose the fibrosis stage, predict prognosis, monitor disease progression, and reflect
473 response to treatment.⁵ Based on this and other studies, VCTE and the Agile scores
474 have already fulfilled the first three requirements, but the latter requires correlation
475 between histological response and changes in non-invasive tests in clinical trials
476 involving an effective treatment. There have already been efforts to fill this
477 knowledge gap using data from several clinical trials,^{29,30} and we expect an
478 acceleration in the validation of response biomarkers when some of the ongoing phase
479 3 trials show positive results. Meanwhile, the existing non-invasive tests can largely
480 replace liver biopsies in routine practice.

481

482 **Limitations**

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

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

496

497 **Conclusions**

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

505

506

507 **References**

- 508 1. Wong VW, Ekstedt M, Wong GL, Hagstrom H. Changing epidemiology,
509 global trends and implications for outcomes of NAFLD. *J Hepatol.*
510 2023;79(3):842-852.
- 511 2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet*
512 (*London, England*). 2021;397(10290):2212-2224.
- 513 3. Younossi ZM, Henry L, Bush H, Mishra A. Clinical and Economic Burden of
514 Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Clin Liver*
515 *Dis.* 2018;22(1):1-10.
- 516 4. Taylor RS, Taylor RJ, Bayliss S, et al. Association Between Fibrosis Stage
517 and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A
518 Systematic Review and Meta-Analysis. *Gastroenterology.* 2020;158(6):1611-
519 1625.e1612.
- 520 5. Sanyal AJ, Castera L, Wong VW. Noninvasive Assessment of Liver Fibrosis
521 in NAFLD. *Clin Gastroenterol Hepatol.* 2023;21(8):2026-2039.
- 522 6. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno VIII.
523 Baveno VII - Renewing consensus in portal hypertension. *J Hepatol.*
524 2022;76(4):959-974.
- 525 7. Sanyal AJ, Foucquier J, Younossi ZM, et al. Enhanced diagnosis of advanced
526 fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile
527 scores. *Journal of hepatology.* 2023;78(2):247-259.
- 528 8. Pennisi G, Enea M, Pandolfo A, et al. AGILE 3+ Score for the Diagnosis of
529 Advanced Fibrosis and for Predicting Liver-related Events in NAFLD.
530 *Clinical gastroenterology and hepatology : the official clinical practice*
531 *journal of the American Gastroenterological Association.* 2023;21(5):1293-
532 1302.e1295.
- 533 9. Zhang X, Yip TC, Wong GL, et al. Clinical care pathway to detect advanced
534 liver disease in patients with type 2 diabetes through automated fibrosis score
535 calculation and electronic reminder messages: a randomised controlled trial.
536 *Gut.* 2023.
- 537 10. Newsome PN, Sasso M, Deeks JJ, et al. FibroScan-AST (FAST) score for the
538 non-invasive identification of patients with non-alcoholic steatohepatitis with
539 significant activity and fibrosis: a prospective derivation and global validation
540 study. *The lancet Gastroenterology & hepatology.* 2020;5(4):362-373.
- 541 11. Wong VW, Adams LA, de Ledinghen V, Wong GL, Sookoian S. Noninvasive
542 biomarkers in NAFLD and NASH - current progress and future promise. *Nat*
543 *Rev Gastroenterol Hepatol.* 2018;15(8):461-478.
- 544 12. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-
545 dependent areas under receiver operating characteristic curves for censored
546 event times with competing risks. *Statistics in medicine.* 2013;32(30):5381-
547 5397.
- 548 13. Yuan Y, Zhou QM, Li B, Cai H, Chow EJ, Armstrong GT. A threshold-free
549 summary index of prediction accuracy for censored time to event data. *Stat*
550 *Med.* 2018;37(10):1671-1681.
- 551 14. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net
552 reclassification improvement calculations to measure usefulness of new
553 biomarkers. *Statistics in medicine.* 2011;30(1):11-21.

- 554 15. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a
555 class of measures to assess improvement in risk prediction systems with
556 survival data. *Statistics in medicine*. 2013;32(14):2430-2442.
- 557 16. Mittal S, El-Serag HB, Sada YH, et al. Hepatocellular Carcinoma in the
558 Absence of Cirrhosis in United States Veterans is Associated With
559 Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol*.
560 2016;14(1):124-131 e121.
- 561 17. Chan TT, Chan WK, Wong GL, et al. Positive Hepatitis B Core Antibody Is
562 Associated With Cirrhosis and Hepatocellular Carcinoma in Nonalcoholic
563 Fatty Liver Disease. *Am J Gastroenterol*. 2020;115(6):867-875.
- 564 18. Mozes FE, Lee JA, Vali Y, et al. Performance of non-invasive tests and
565 histology for the prediction of clinical outcomes in patients with non-alcoholic
566 fatty liver disease: an individual participant data meta-analysis. *Lancet*
567 *Gastroenterol Hepatol*. 2023;8(8):704-713.
- 568 19. Hagstrom H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts
569 mortality and time to development of severe liver disease in biopsy-proven
570 NAFLD. *J Hepatol*. 2017;67(6):1265-1273.
- 571 20. Wai JW, Fu C, Wong VW. Confounding factors of non-invasive tests for
572 nonalcoholic fatty liver disease. *Journal of gastroenterology*. 2020;55(8):731-
573 741.
- 574 21. Chow JC, Wong GL, Chan AW, et al. Repeating measurements by transient
575 elastography in non-alcoholic fatty liver disease patients with high liver
576 stiffness. *J Gastroenterol Hepatol*. 2019;34(1):241-248.
- 577 22. Loomba R, Huang DQ, Sanyal AJ, et al. Liver stiffness thresholds to predict
578 disease progression and clinical outcomes in bridging fibrosis and cirrhosis.
579 *Gut*. 2023;72(3):581-589.
- 580 23. Semmler G, Yang Z, Fritz L, et al. Dynamics in Liver Stiffness Measurements
581 Predict Outcomes in Advanced Chronic Liver Disease. *Gastroenterology*.
582 2023.
- 583 24. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated
584 FIB-4 measurements can help identify individuals at risk of severe liver
585 disease. *Journal of hepatology*. 2020;73(5):1023-1029.
- 586 25. Cholankeril G, Kramer JR, Chu J, et al. Longitudinal changes in fibrosis
587 markers are associated with risk of cirrhosis and hepatocellular carcinoma in
588 non-alcoholic fatty liver disease. *Journal of hepatology*. 2023;78(3):493-500.
- 589 26. Mozes FE, Lee JA, Selvaraj EA, et al. Diagnostic accuracy of non-invasive
590 tests for advanced fibrosis in patients with NAFLD: an individual patient data
591 meta-analysis. *Gut*. 2022;71(5):1006-1019.
- 592 27. McPherson S, Hardy T, Dufour JF, et al. Age as a Confounding Factor for the
593 Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. *The*
594 *American journal of gastroenterology*. 2017;112(5):740-751.
- 595 28. Wong VW, Tak WY, Goh GBB, et al. Performance of Noninvasive Tests of
596 Fibrosis Among Asians, Hispanic, and non-Hispanic Whites in the STELLAR
597 Trials. *Clin Gastroenterol Hepatol*. 2023;21(1):90-102 e106.
- 598 29. Rinella ME, Dufour JF, Anstee QM, et al. Non-invasive evaluation of
599 response to obeticholic acid in patients with NASH: Results from the
600 REGENERATE study. *Journal of hepatology*. 2022;76(3):536-548.
- 601 30. Wai-Sun Wong V, Anstee QM, Nitze LM, et al. FibroScan-aspartate
602 aminotransferase (FAST) score for monitoring histological improvement in
603 non-alcoholic steatohepatitis activity during semaglutide treatment: post-hoc

- 604 analysis of a randomised, double-blind, placebo-controlled, phase 2b trial.
605 *EClinicalMedicine*. 2023;66:102310.
- 606 31. Zhang X, Yip TC, Tse YK, et al. Duration of type 2 diabetes and liver-related
607 events in nonalcoholic fatty liver disease: A landmark analysis. *Hepatology*.
608 2023.
609
- 610

611 **Table 1: Clinical characteristics of the cohorts in the baseline and serial models**
 612

Characteristics	Baseline model N = 16 603	Serial model		P value ^b
		First test N = 10 920	Last test	
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)	

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

614 ^aThe formulas for the calculation of the non-invasive tests are presented in the Supplement page 3-4.

615 ^bFibrosis stage (0-4) according to the NASH CRN system. Stage 0, no fibrosis; Stage 1, centrilobular pericellular
 616 fibrosis; Stage 2: centrilobular and periportal fibrosis; Stage 3: bridging fibrosis; Stage 4, cirrhosis.

617 ^cPaired samples tests between the first and last tests in the serial model.

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

622

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

626

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**

Tests	3-year liver-related events			5-year liver-related events		
	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

635 **Figure legends**

636 ***Figure 1: Study participant flow***

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

639

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

641 A, AUROC and AUPRC for the prediction of liver-related events at 3 and 5 years. B,

642 Cumulative incidence of liver-related events stratified by Agile 3+ score in the baseline
643 model.

644 In panel B, the cut points for Agile 3+ score were based on the original publication. The low
645 cut point (0.451) achieved sensitivity of $\geq 85\%$ to rule-out patients of fibrosis stage ≥ 3 , the
646 high cut point (0.679) achieved specificity of $\geq 90\%$ to rule-in patients of fibrosis stage ≥ 3 .

647 The median follow-up duration of each group was listed in the legend.

648 Abbreviations: AST, aspartate aminotransferase. ALT, alanine aminotransferase ratio.

649 AUROC, area under the receiver-operating characteristic curve. AUPRC, area under the
650 precision-recall curve. CI, confidence interval. LRE, liver-related event. NAFLD, non-
651 alcoholic fatty liver disease

652

653

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

655 A, Change in the Agile 3+ between two vibration-controlled transient elastography

656 examinations. B, Relative change in the Agile 3+ score and incident liver-related events after
657 the last test.

658 In panel A, the numbers in the middle represent the percentages of patients in each group.
659 Patients who developed liver-related events before the last examination are shown in the top
660 of the Sankey diagram.
661 Abbreviations: CI, confidence interval. LREs, liver-related events. PY, person-year.
662

**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)

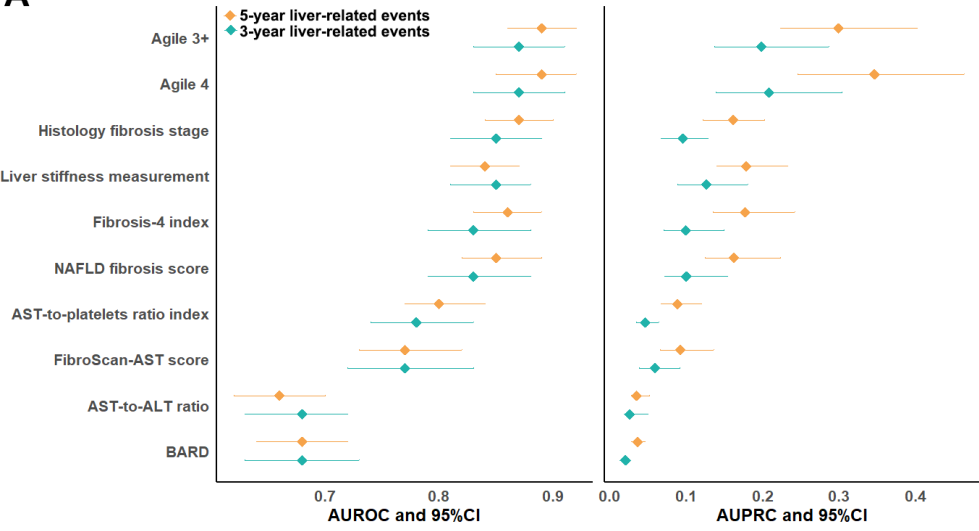
Baseline model
(N = 16603)

Excluded:

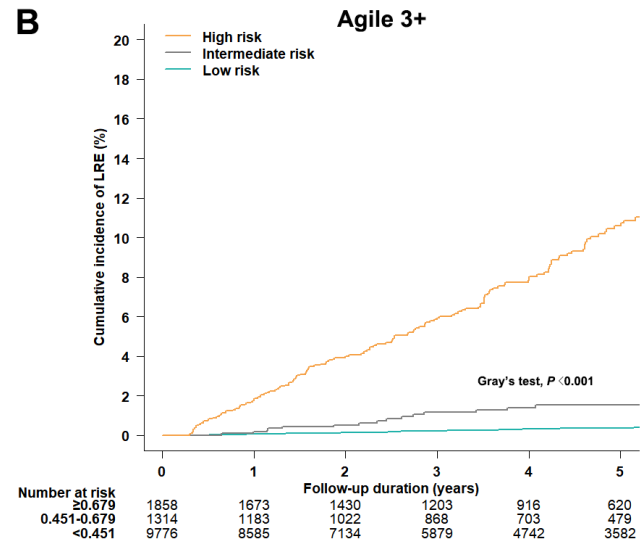
Without repeat test (N = 4157);
Time interval between two tests <6 months or >5
years (N = 1409);
HCC or decompensation occurred between two
tests (N = 117)

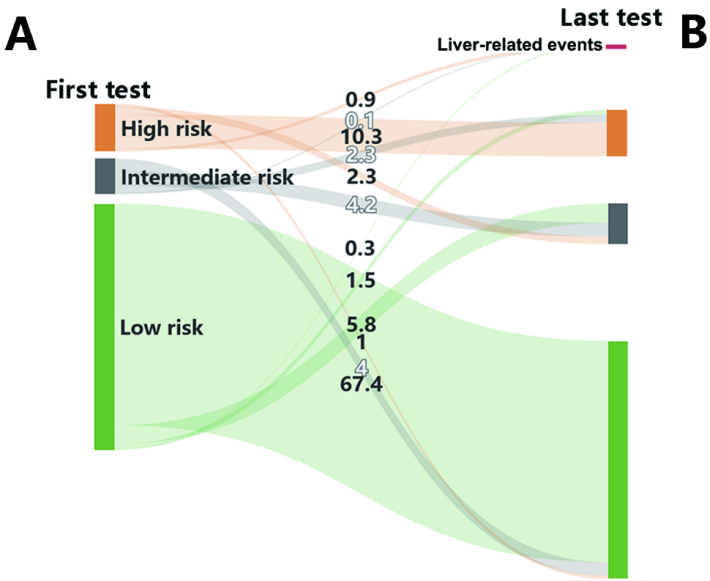
Serial model
(N = 10920)

A



B





B

First test	Relative change	N (%)	5-year LRE (%)	Relative change	N (%)	5-year LRE (%)	Relative change	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)	(-)
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:

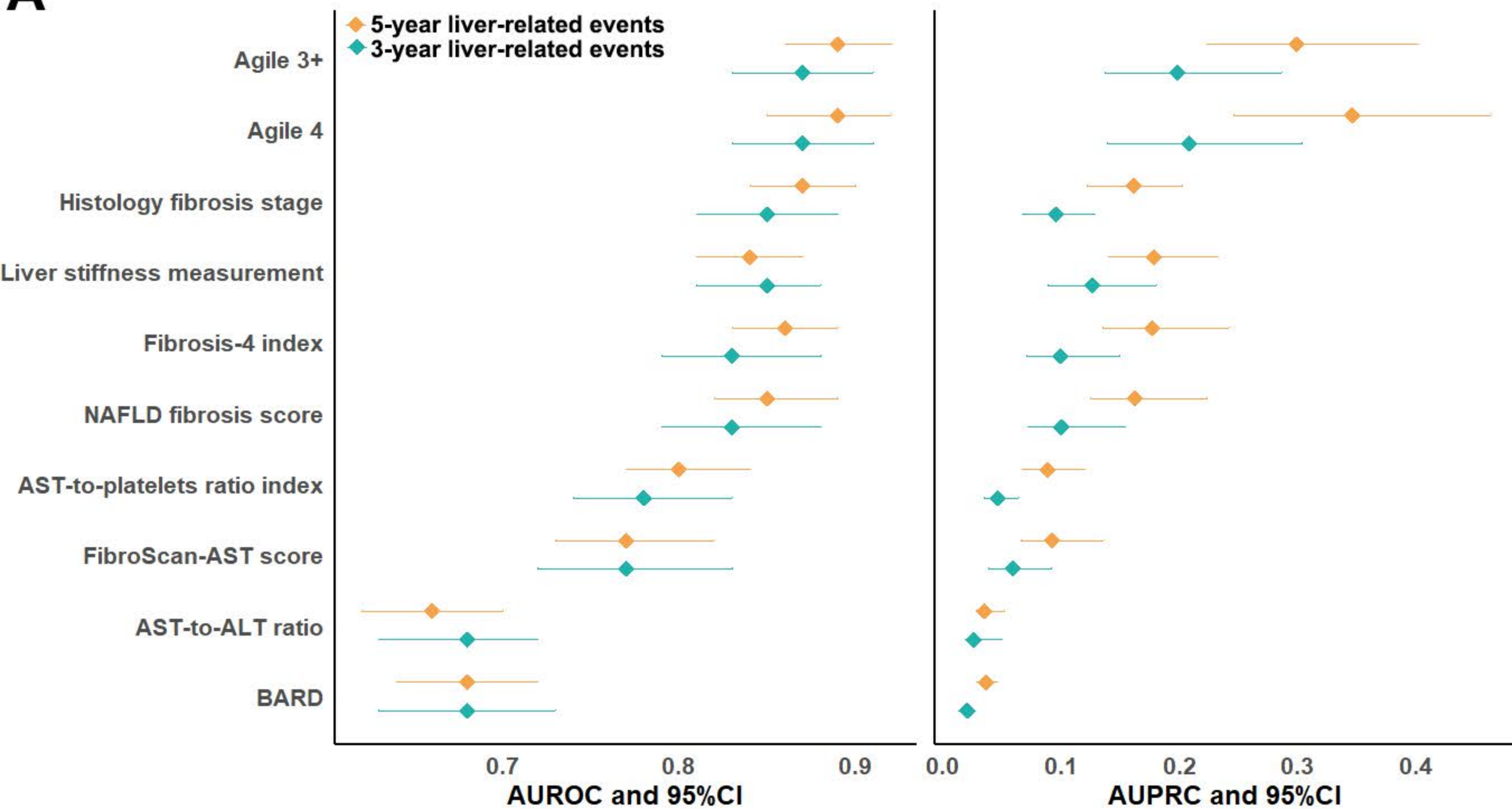
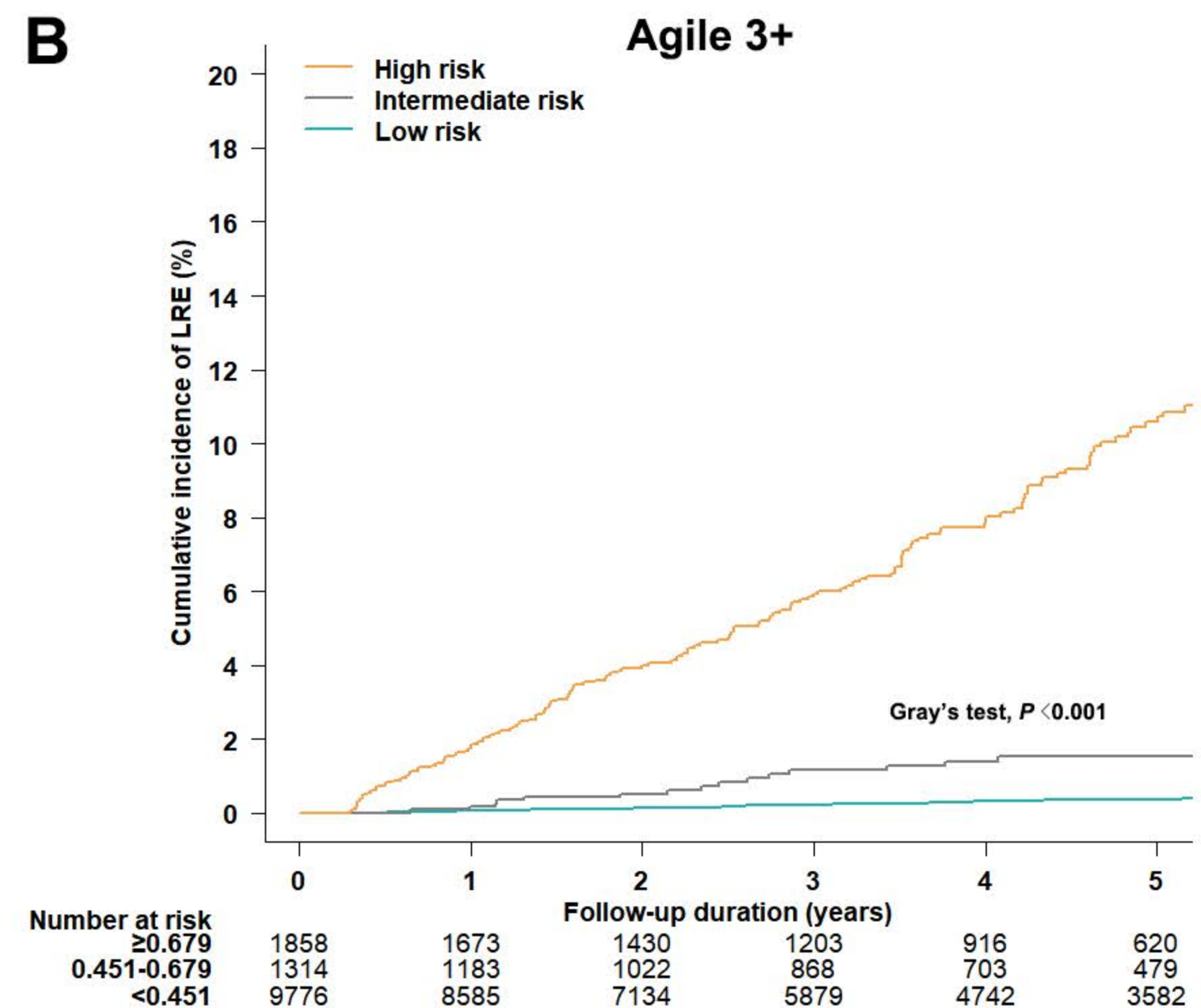
Age <18 years or age unknown (N = 679);
HCC or decompensation before VCTE or No follow-
up data (N = 598);
HCC or decompensation within 3 months after
VCTE (N = 69)

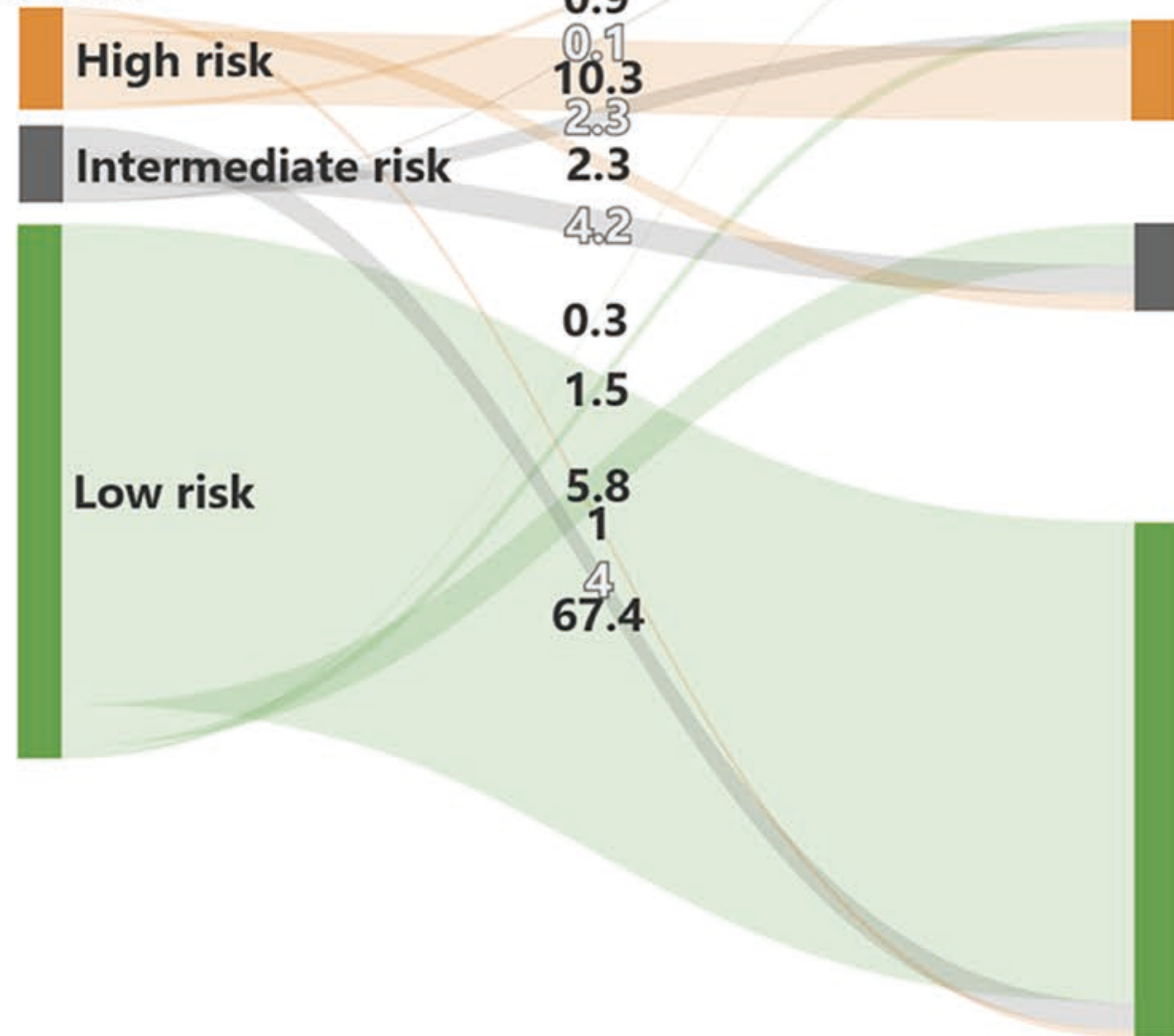
Baseline model
(N = 16603)

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**B**

A**First test****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)