Derivation and validation of the NAFLD Cirrhosis Score (NCS) to distinguish bridging fibrosis from cirrhosis

European Journal of Internal Medicine

Derivation and validation of the NAFLD Cirrhosis Score (NCS) to distinguish bridging fibrosis from cirrhosis

--Manuscript Draft--

Manuscript Number:	EJINME-D-21-01702R1
Article Type:	Original Article
Keywords:	liver cirrhosis, advanced fibrosis, liver-related events, prognosis, cardiovascular events
Corresponding Author:	Jörn M. Schattenberg, MD University Medical Center Mainz Mainz, GERMANY
First Author:	Christian Labenz
Order of Authors:	Christian Labenz
	Gerrit Toenges
	Peter R. Galle
	Ming-Hua Zheng
	Dora Ding
	Robert P. Myers
	Angelo Armandi
	Javier Ampuero
	Manuel Romero Gomez
	Elisabetta Bugianesi
	Quentin M. Anstee
	Jörn M. Schattenberg, MD
Abstract:	Separation of bridging fibrosis from cirrhosis in non-alcoholic fatty liver disease (NAFLD) is critical to guide management. The aim of this study was to develop an easy-to-perform score distinguishing F3 and F4 fibrosis in NAFLD. A derivation cohort comprising 251 NAFLD patients with F3 or F4 was used to develop the NAFLD Cirrhosis Score (NCS). The NCS was validated in three independent cohorts with liver histology comprising 1,666 participants from the STELLAR trials, 47 patients from China and 2,058 patients from the European NAFLD Registry. A model including INR, gGT, ALT, platelets and age discriminated best between patients with bridging fibrosis and cirrhosis with an area under the curve (AUC) of 0.733 (95%CI 0.671–0.795). The diagnostic performance of the NCS was similar in the STELLAR studies (AUC 0.700; 95%CI 0.680-0.730). In the European NAFLD Registry, spanning all histological fibrosis stages, the NCS exhibited an AUC of 0.798 (95%CI 0.766-0.830) to detect cirrhosis. We derived two NCS cut-off values (<64.5 and >79.17) to classify patients at low, intermediate, or high risk for the presence of cirrhosis. Using these cut-offs, diagnostic workup could be avoided by ruling in or ruling out cirrhosis in half of the patients. The NCS identified patients at risk for progression to cirrhosis and liver-related outcomes. Conclusion: The NCS is a simple tool to improve the identification of compensated cirrhosis within the group of advanced disease stage and provides prognostic information. The differentiation of F3 from F4 disease using standard laboratory remains difficult.

UNIVERSITÄTSmedizin. MAINZ

I. Department of Internal Medicine
Direktor: Prof. Dr. P. R. Galle

Jörn M. Schattenberg, MD

Director Metabolic Liver Research Program
Gastroenterology and Hepatology

Bldg 601, 1. floor, Room 1.124
Langenbeckstrasse 1
55131 Mainz

Phone: +49 (0) 6131 176074
Fax: +49 (0) 6131 17477282
E-Mail: joern.schattenberg@unimedizin-mainz.de
http://www.unimedizin-mainz.de/1-med

Mainz, 4. Dezember 2021

To the Editors
EJIM

EJINME-D-21-01702 entitled "Derivation and validation of the NAFLD Cirrhosis Score (NCS) to distinguish bridging fibrosis from cirrhosis"

Dear Prof. Agnelli,
dear Prof. van Erpecum

We would like to thank you for returning the reviewers and editors comments and providing us with the opportunity to resubmit after major revisions. We found the suggestions very helpful. All points are highly relevant and overall the study was performed to highlight the challenges of non-invasive test in the distinction of cirrhotic from pre-cirrhotic NASH. We believe that our contribution to the field will underline the currently ongoing developments in the field of NITs in NASH and put the standard lab scores into perspective.

The revised manuscript contains all changes highlighted and a point-to-point response is included below. We believe that the comments and consecutive changes improved the manuscript and would hope that you can further consider the manuscript for publication in the European Journal of Internal Medicine.

Sincerely yours,

Jörn M. Schattenberg, MD

UNIVERSITÄTSmedizin. MAINZ
I. Department of Internal Medicine
Direktor: Prof. Dr. P. R. Galle

Point-by-point reply

The revised manuscript has been modified according to the comments and all changes in the former version of the manuscript are highlighted in red (named R1_Manuscript file). To us, the comments and consecutive changes significantly improved the quality of the manuscript and after careful revision we hope that our manuscript can further be considered for publication in the European Journal of Internal Medicine.

Point-by-point reply to all comments raised by the reviewer

Reviewer 2:

In this work, the authors aim to develop an easy-to-perform score distinguishing bridging fibrosis (F3) from cirrhosis (F4) in non-alcoholic fatty liver disease (NAFLD). A derivation cohort comprising 251 NAFLD patients with F3 or F4 was used to develop the NAFLD Cirrhosis Score (NCS). The NCS was validated in three independent cohorts with liver histology comprising 1,666 participants from the STELLAR trials, 47 patients from China and 2,058 patients from the European NAFLD Registry. A model including INR, gGT, ALT, platelets and age discriminated best between patients with bridging fibrosis and cirrhosis with an area under the curve (AUC) of 0.733. The diagnostic performance of the NCS appeared similar in the STELLAR studies (AUC 0.700 and the small cohort from China (AUC 0.727). In the European NAFLD Registry, spanning all histological fibrosis stages, the NCS exhibited an AUC of 0.798 (95%CI 0.766-0.830) to detect cirrhosis. The authors derived two NCS cut-off values (<64.5 and >79.17) to classify patients at low, intermediate, or high risk for the presence of cirrhosis. The authors claim that using these cut-offs, further diagnostic workup could be avoided by ruling in or ruling out cirrhosis in approximately half of the patients. Furthermore, they claim that NCS identified patients at risk for progression to cirrhosis in the F3 cohort and liver-related outcomes in the F4 cohort.

Detailed comments:

1. The reviewer is not convinced after reading this work, that distinguishing F3 and F4 fibrosis (i.d. severe fibrosis from cirrhosis) would have major impact on the management of NAFLD patients with advanced liver disease. The arguments given by the authors in favour of such relevance for the decision whether or not to screen patients for esophageal varices does not hold true: here the (original or expanded) Baveno criteria (based on results of fibroscan and thrombocyte count have been extensively validated. Please note that patients with early cirrhosis (platelet count >110x109/L and LSM value ≥25 kPa: expanded Baveno criteria) have low risk of large varices and no indication for variceal screening. In contrast, the authors do not provide direct evidence that the NCS score of the current work is reliable to exclude large varices. Similarly, up to 40% of NAFLD patients who develop hepatocellular carcinoma, have no cirrhosis. NAFD is well known as a cause of non-cirrhotic hepatocellular carcinoma. Of course, one can debate if NAFLD patients should be screened for hepatocellular carcinoma anyway, considering the relative low risk of this event in NAFLD (compared to viral hepatitis), even if cirrhosis is present, provided that there are no other cofactors. However, that is another discussion. The NCS proposed here has a quite modest predictive value to distinguish F3 from F4. It would be unwise in the opinion of the reviewer to base screening decisions on the NCS.

Thank you very much for your thoroughness reviewing our manuscript and your thoughts. We agree that the NCS is unlikely to change practice in highly specialized centers where all diagnostic tools are available. However, we believe that the NCS may have merit in settings where especially transient elastography or other more elaborate blood-based tests are not available.

We also agree that the diagnostic accuracy of the NCS is far from perfect. However, this only reflects the dilemma and limitations of all blood-based test in detecting early and compensated cirrhosis in patients with NAFLD. Given that we used some of the best cohorts currently available, we believe that is also a relevant finding that it is not possible to develop a better test than this using readily available parameters. We stressed these facts in our manuscript and reworked our discussion section accordingly.

2. The authors state in the methods for the 251 patients in the derivation cohort that: "All liver biopsies were obtained according to local practice and scored by one experienced liver histopathologist in each center". It would be highly preferable to have judgement by a central pathologist considering the potential for inter-observer variability. This was correctly done in the Stellar study (central reader). However, it is not clear who evaluated the biopsy in the European and Chinese cohorts.
The assessment of liver biopsies was performed by experienced histopathologists that are part of the European NAFLD registry study and are in contact among each other to sync on the histological scoring. In the STELLAR studies all biopsies were centrally read. As stated, all biopsies were scored by (experienced) liver histopathologists at the respective centers. Nevertheless, we agree that this may introduce potential bias and acknowledged this in our discussion section.

3. A major concern is that no details about quality of the liver biopsy in the various cohorts are provided. This is very relevant because the biopsy is used as gold standard and the performance of the biopsy to correctly stage fibrosis is depending on length of the biopsy (see Figure 5 Bedossa et al . Hepatology 2003. The authors should add details of liver biopsy quality (in particular biopsy length) and perform sensitivity analyses by including in an additional exercise with only biopsies > 25 mm length (which is required for accurate staging).
The comment on the quality of liver biopsies is well taken. In the European cohorts (Mainz, Turin, Seville) liver biopsies were obtained as part of the standard of care at expert centers. These three centers are experienced, and all have a standard operating procedure implemented to ensure the highest quality of liver biopsy. Thus, the biopsies obtained as part of the derivation cohort are representative of standard of care. The STELLAR studies have more specifically defined the use of only liver biopsies the were deemed of high enough quality by one central reader. Overall, while not being in the position to revisit the individual quality metric of all 4022 liver biopsies included in in this analysis, represent the standard of care at expert centers or from within phase 2 clinical trials.

4. It is not entirely clear what is the additive value of the Chinese cohort, considering the very small patient numbers (47 pts).

   We agree on the limitation that arise form a small cohort from China. Nonetheless, we wanted to allow the evaluation of the NCS in genetically and culturally diverse backgrounds. Therefore, we included the Chinese cohort to investigate the predictive ability of the NCS. The limitations are acknowledged.

5. It is not clear why the authors decided to use only F3 and F4 patients (according to the liver biopsy) in the derivation cohort of 251 patients. Similarly, it is not clear why the authors did not decide to include all patients who were screened in the stellar 3 and 4 studies. Now only patients who met the F3 or F4 criteria in biopsy were included. Inclusion of all patients (also including in the current work those patients who could not participate in the stellar study because of F1 or F2 in biopsy would have given much more information). In practice, one generally does not know whether there is at least advanced fibrosis (F3 or F4) without biopsy. Of note Fibroscan has very good negative predictive value but only modest positive predictive value for this aim (need for biopsy when fibroscan suggests advanced fibrosis.)

   This comment is at the heart of this analysis. All scores and modalities in clinical practice focus on early vs advanced fibrosis. When looking at current guidelines, there is a clear recommendation to separate patient with cirrhosis from other patients. In addition, certain recommendation related to e.g. screening for HCC are not cost effective before the stage of cirrhosis. We wanted to answer the question, if a blood-base score, costing of standard labs, is capable to separate these cohorts and to determine the test accuracy. The final answer is that the ability of the noninvasive scores is only moderate. To us, this data is very important and underlines the big unmet need, that is currently tried to be overcome by direct fibrosis marker or imaging modalities. The Fibroscan is a tool for experts, that will not be available at large scale outside of expert hands, related to its high cots, the need for a separate room and personal to operate it. Therefore, in clinical routine care, blood-based markers will mainly be utilized. Along this line, the NCS could be used in primary care to further refine the separation of the advanced fibrosis cohort.

6. In follow up of 5. : the authors report in the results section for the derivation cohort very high positive and negative predictive values ("Using the above-mentioned cutoffs the sensitivity, specificity, PPV and NPV of the resulting diagnosis algorithm in the derivation cohort were 90.6%, 90.3%, 85.3% and 93.9%, respectively"). Similar results are given for the Stellar study. However, this is also related to the high a priori risk of these cohorts (only F3/F4 included). This is a major limitation. It would have been preferable to include all available patients (also F0, F1, F2) to obtain results that could be generalized to clinical practice (see also comment 5).
Based on the previous comment, we aimed to develop a score to separate F3 from F4 and therefore chose a cohort that was representative of this population for the development. To determine how it performs in the entire disease spectrum, in particular as a follow-up of the FIB-4, we next employed it to the European NAFLD registry cohort which constitute the entire F0-F4 spectrum. We accept all limitations arising from this model and have highlighted them in the discussion.

7. The results on progression in the STELLAR cohort are interesting. (Prognostic utility of the NAFLD Cirrhosis Score (NCS) to predict liver-related outcomes and MACE). In the stellar 4 study, patients with portal hypertension and varices (i.e. advanced cirrhosis) could be included, provided that there were no previous events of decompensated liver disease (ascites, hepatic encephalopathy, variceal bleeding). The authors should give the results for clinical events also separately for the subgroups with and without a priori evidence for advanced cirrhosis (varices, thrombocytopenia).
Thank you very much for your suggestion. We added these data into the results section accordingly.

8. The data available in the STELLAR cohort merit reporting additional details that could make the current work more valuable (also in the light of the very modest predictive value of the NCS). In many patients included in the STELLAR study, also Fibroscan was performed at baseline (i.e. when the equipment was available in the participating center). The authors should explore whether

combining Fibroscan results with NCS score at baseline could improve the performance.

This is an important comment. Imaging and blood-based biomarkers are increasingly used – either in a hierarchical fashion or even combined. They amplify the results and improve the accuracy. However, the limitations arise from the additional costs and the very limited access to Fibroscan. There will be no Fibroscan machines in primary care due to costs and requirement of personal and room requirements. Therefore, we deliberately developed a score that is based on very basic and standard lab measures. The important message is that the separation of F3 from F4 is difficult to be achieved. The specificity of the test can help to guide clinical decisions. We have included this in the discussion.

Reviewer 3:

Labenz and collegues were interested in finding biomarkers that would allow separation of bridging F3 fibrosis from F4 cirrhosis in non-alcoholic fatty liver disease. They developed a derivation cohort comprising 251 NAFLD patients with F3 or F4 was used to develop the NAFLD Cirrhosis Score (NCS). The NCS was validated in three independent cohorts (3724 patients) with liver histology. The authors build a model including INR, gGT, ALT, platelets and age that discriminated best between patients with bridging fibrosis and cirrhosis with an area under the curve (AUC) of 0.733. The also used the model to identify patients at risk for progression to cirrhosis in the F3 cohort and liver-related outcomes in the F4 cohort.

This is a well written manuscript with a clear message and with a tangible outcome. I have to commend the authors with the use of a rigorous methodology and in particular 3 well defined and characterized cohorts. This adds to the credibility of the results.

1. The TRIPOD Statement aims to improve the transparency of the reporting of a prediction model study I like the use of the TRIPOD guideline. It would be of benefit to the reviewers and readers of this prediction model study that the completed checklist is added to their submission. While this is a technically well performed study I would like to veer back to the question the authors want to address. It was the aim of this study to develop an easy-to-perform score distinguishing F3 and F4 fibrosis in NAFLD. I have a number of questions.

Thank you very much for your kind words. We added the completed checklist at the end of our manuscript.

2. What is the clinical relevance to separate F3 from F4, you could argue that F3 patients do not benefit from hepatocellular carcinoma screening but that only goes that far. The authors offer a number of arguments in the discussion and while I lend credence to their argumentation, the separation in real world is less relevant

   Thank you for this comment. At current, recommendations to primary and secondary care are to screen for HCC and varices in cirrhotic patients. Most settings don't have high-end tools or knowledge to achieve this. Therefore, we developed a tool that combines very basic blood-based markers, to support non-specialist in ruling-out patients that are noncirrhotic. Looking at the significant increase of patient affected, we believe that this tool will is helpful. We agree on the inaccuracy of histology and separation of F3 and F4. Nonetheless, attempting to separate these stages will allow to improve the management of these patients. Prospective evaluation will have to show its benefit and relevance.

3. The gold standard here was the liver biopsy. It is known that the diagnostic performance of the liver biopsy is incomplete, in particular in the distinction of the fibrosis grades. Could the authors highlight the robustness of their gold standard? (ie length of biopsy specimen, how many pathologists scored, independent replication etc)

   This is an important comment, that was also raised by R2. We applied standard of care – including all limitations that liver histology has. In each center an expert pathologist (affiliation with the European NAFLD registry study) scored the biopsies. Based on the standardized assessment within the European NAFLD registry study these pathologists have harmonized previously. The STELLAR studies were read centrally by one pathologist. Importantly, the current analysis included over 4000 liver biopsies – a strength of this analysis in our perception. The limitations of liver biopsies are however inherent to all studies addressing this surrogate instead of outcome. We

have included this in the discussion.

4. The diagnostic performance of the NCS is reasonable (according to the authors "fair accuracy", what should be done to enhance the diagnostic properties of the test
   The NCS reflects the best a simple blood-based score ca achieves. The diagnostic accuracy will unlikely be higher when using standard and readily available parameters. Given the fact that we included some of the most intensively characterized cohorts available, these finding is relevant – and highlights the limitations of NITS to identify cirrhosis within the group of advanced fibrosis.

5. I like the statement that the NCS had an acceptable predictive ability to detect cirrhosis in unselected patients with all grades of fibrosis. This would be beneficial in clinical practice and it would be great to compare this with the FIB-4.
   Not surprisingly, NCS did not exceed the predictive ability of Fib-4 in this unselected cohort (AUC 0.830). We believe that the NCS may especially have merit in settings without specialized tests like transient elastography. Here, stratifying patients with sequential use of Fib-4 and NCS may improve patient management.

Derivation and validation of the NAFLD Cirrhosis Score (NCS) to distinguish bridging fibrosis from cirrhosis

Running head: NAFLD cirrhosis score

Christian Labenz[1, 2, 3], Gerrit Toenges[4], Ming-Hua Zheng[5], Dora Ding[6], Robert P. Myers[6], Peter R. Galle[1,3], Angelo Armandi[7], Javier Ampuero[8], Manuel Romero Gómez[8], Elisabetta Bugianesi[7], Quentin M. Anstee[9,10], Jörn M. Schattenberg[1, 2, 3*]

Affiliations:

[1]Department of Internal Medicine I, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany

[2]Metabolic Liver Research Program, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany

[3]Cirrhosis Center Mainz (CCM), University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

[4]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

[5]NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, China

[6]Gilead Sciences, Inc., Foster City, CA, USA

[7]Department of Medical Sciences, Division of Gastroenterology, AOU Citta della Salute e della Scienza, University of Torino, Italy

[8]Digestive Disease Department, Virgen del Rocio University Hospital, Sevilla, Spain

[9]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom.

[10]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, United Kingdom.

* Corresponding author:

Jörn M. Schattenberg, MD

Metabolic Liver Research Program

I. Department of Internal Medicine

University Medical Centre of the Johannes Gutenberg-University

Langenbeckstrasse 1

55131 Mainz, Germany

Telephone: +49 (0) 6131 17 6074

Telefax: +49 (0) 6131 17 477282

E-Mail: joern.schattenberg@unimedizin-mainz.de

Electronic word count: 3,706; tables 4, figures 4

E-Mails:

Christian Labenz: christian.labenz@unimedizin-mainz.de

Gerrit Toenges: gtoenges@unimedizin-mainz.de

Peter R. Galle: peter.galle@unimedizin-mainz.de

Ming-Hua Zheng: zhengmh@wmu.edu.cn

Dora Ding: dora.ding@gilead.com

Robert P. Myers: rob.myers@gilead.com

Angelo Armandi: angelo.armandi@unito.it

Javier Ampuero: jampuero-ibis@us.es

Manuel Romero Gómez: mromerogomez@us.es

Elisabetta Bugianesi: elisabetta.bugianesi@unito.it

Quentin M. Anstee: quentin.anstee@newcastle.ac.uk

Abbreviations: BMI, body mass index; CP, Child-Pugh; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; MACE, major adverse cardiovascular events

Keywords: liver cirrhosis, advanced fibrosis, liver-related events, prognosis, cardiovascular events

Conflict of interest: JMS has acted as consultant to Boehringer Ingelheim, BMS, Echosens, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Madrigal, Novartis, Novo Nordisk, Nordic Bioscience, Pfizer, Roche, Sanofi, Siemens Healthcare GmbH, Zydus. Research Funding: Gilead Sciences. Speakers Bureau: Falk Foundation MSD Sharp & Dohme GmbH. DD and RPM are employees and shareholders of Gilead Sciences, Inc. EB acted as a consultant to Boehringer Ingelheim, BMS, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Novo Nordisk, Pfizer. Research Funding: Gilead Sciences. QMA reports Research Grant Funding: Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex. Active Research Collaborations: (including research collaborations supported through the EU IMI2 LITMUS Consortium*) Abbvie, Antaros Medical*, Allergan/Tobira*, AstraZeneca*, BMS*, Boehringer Ingelheim International GMBH*, Echosens*, Ellegaard Gottingen Minipigs AS*, Eli Lilly & Company Ltd.*, Exalenz Bioscience Ltd.*, Genfit SA*, Glympse Bio, GlaxoSmithKline, HistoIndex*, Intercept Pharma Europe Ltd.*, iXscient Ltd.*, Nordic Bioscience*, Novartis Pharma AG*, Novo Nordisk A/S*, One Way Liver Genomics SL*, Perspectum Diagnostics*, Pfizer Ltd.*, Resoundant*, Sanofi-Aventis Deutschland GMBH*, SomaLogic Inc.*, Takeda Pharmaceuticals International SA*. Consultancy: 89Bio, Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, BMS, BNN Cardio, Celgene,

Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, PathAI, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Terns, Viking Therapeutics. Speaker: Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, MedScape. Royalties: Elsevier Ltd (Davidson's Principles & Practice of Medicine textbook). The other authors have nothing to disclose.

Funding: MRG, EB, QMA and JMS are supported by the LITMUS (Liver Investigation: Testing Biomarker Utility in Steatohepatitis) consortium funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under Grant Agreement 777377; this Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The STELLAR studies were funded by Gilead Sciences, Inc. MRG, EB, QMA and JMS are European NAFLD Registry investigators.

Abstract

Separation of bridging fibrosis from cirrhosis in non-alcoholic fatty liver disease (NAFLD) is critical to guide management. Therefore, it was the aim of this study to develop an easy-to-perform score distinguishing F3 and F4 fibrosis in NAFLD. A derivation cohort comprising 251 NAFLD patients with F3 or F4 was used to develop the NAFLD Cirrhosis Score (NCS). The NCS was validated in three independent cohorts with liver histology comprising 1,666 participants from the STELLAR trials, 47 patients from China and 2,058 patients from the European NAFLD Registry. A model including INR, gGT, ALT, platelets and age discriminated best between patients with bridging fibrosis and cirrhosis with an area under the curve (AUC) of 0.733 (95%CI 0.671–0.795). The diagnostic performance of the NCS was similar in the STELLAR studies (AUC 0.700; 95%CI 0.680-0.730) and a smaller cohort from China (AUC 0.727; 95%CI 0.533–0.921). In the European NAFLD Registry, spanning all histological fibrosis stages, the NCS exhibited an AUC of 0.798 (95%CI 0.766-0.830) to detect cirrhosis. We derived two NCS cut-off values (<64.5 and >79.17) to classify patients at low, intermediate, or high risk for the presence of cirrhosis. Using these cut-offs, further diagnostic workup could be avoided by ruling in or ruling out cirrhosis in approximately half of the patients. Furthermore, NCS identified patients at risk for progression to cirrhosis in the F3 cohort and liver-related outcomes in the F4 cohort.

Conclusion: The NCS is a simple tool to improve the identification of compensated cirrhosis within the large group of advanced disease stage and provides prognostic information. Overall, the differentiation of F3 from F4 disease using standard laboratory remains difficult and does not exceed moderate accuracy.

Introduction

Non-alcoholic fatty liver disease (NAFLD) has become a major health burden (1). Among the large group of patients that are affected, the subgroup progressing to end-stage disease exhibits the highest mortality and health expenses (2, 3).

A number of blood-based non-invasive surrogate scores have been developed to guide in the management of NAFLD. However, none of these separate bridging fibrosis and cirrhosis – defined as the histological stages F3 and F4 according to the NASH-Clinical Research Network (CRN) staging system (4, 5). The addition of direct fibrosis markers, such as the Enhanced Liver Fibrosis (ELF) (6) or PRO-C3 (7, 8) has improved performance but does not add to the ability to identify cirrhosis. As of today, this challenge is only met by imaging modalities including magnetic resonance elastography and transient elastography (9, 10). In addition, the BARVENO VI criteria have been validated extensively but focus on patients with liver cirrhosis and are used in the context of screening for varices and clinical significant portal hypertension (11).

From a clinician's perspective, the availability of an easy-to-use, low-cost, blood-based test to distinguish patients with compensated cirrhosis from bridging (F3) fibrosis holds value for several reasons. First, most patients are treated in primary care and here costly tests (e.g. transient elastography) are often not readily available. Second, there is strong evidence to support screening and surveillance measures in patients with cirrhosis, including screening for varices and hepatocellular carcinoma (HCC), Also, considering future pharmacotherapy labels, the easy separation of pre-cirrhotic from cirrhotic NAFLD could be of importance to identify the subgroup of patients that benefits most – from an efficacy perspective – or are at greatest risk – from a safety perspective. Therefore, a tool with the ability to distinguish between patients with F3 and F4 fibrosis and additionally provide prognostic information related to the risk of disease progression would be of great value.

Materials and methods

Patient cohorts - derivation cohort

In the derivation cohort, 251 adult patients with biopsy-proven NAFLD and fibrosis stages F3 or F4 according to the NASH CRN were included at the University Medical Centers Mainz (Germany; n=122), Torino (Italy; n=97) and Seville (Spain; n=32), as recently described (12). For the current analysis, only patients with fibrosis stages F3 or F4 according to the NASH CRN classification were selected. Patients with other liver diseases, a Child-Turcotte-Pugh Score >6, or significant alcohol use based on clinical grounds or random urine ethylglucuronid measurements were excluded. All liver biopsies were obtained according to local practice and scored by one experienced liver histopathologist in each center (13). All laboratory tests were obtained within 90 days of liver biopsy.

External validation cohorts

For validation purposes, participants in the STELLAR-3 and -4 clinical trials and a real-world cohort from China were used. The STELLAR trials evaluated selonsertib versus placebo in patients with NASH (defined as a NAS of ≥3 with at least grade 1 for each of steatosis, hepatocellular ballooning and lobular inflammation) and bridging fibrosis (F3 for STELLAR-3; NCT03053050) or compensated cirrhosis (F4 for STELLAR-4; NCT03053063) according to the NASH CRN classification. The primary results and methods of these studies are reported elsewhere (14). In the current analysis, patients recruited at the clinical trial sites in Mainz, Torino and Seville were excluded to avoid duplicate inclusion. In both STELLAR studies, the planned duration of treatment was 240 weeks. However, the studies were halted after pre-planned interim analyses conducted after all patients had completed at least 48 weeks of treatment found that

there were no meaningful differences between the active treatment groups or the placebo group in any efficacy endpoint. Therefore, for the purposes of this analysis, treatment groups were combined.

In both STELLAR trials, liver biopsies performed during screening and at week 48 were evaluated by a single central reader blinded to study treatment. In the STELLAR-3 trial, a key endpoint was progression to cirrhosis at week 48, defined as histologic progression to cirrhosis or the development of hepatic decompensation (defined below) during follow-up. In both studies, time to first liver-related clinical event, defined as hepatic decompensation (clinically apparent ascites requiring treatment, hepatic encephalopathy of Grade 2 or above according to the West Haven criteria requiring treatment, and portal hypertension-related gastrointestinal bleeding), liver transplantation, qualification for transplantation (MELD ≥15), or all-cause mortality, as confirmed by an independent Hepatic Events Adjudication Committee, was evaluated. Finally, a Cardiovascular Events Adjudication Committee reviewed all major adverse cardiovascular events (MACE) including cardiovascular death, myocardial infarction, stroke, hospitalization for unstable angina or cardiac failure, and coronary revascularization.

In addition to patients from the STELLAR studies, a second external real-world validation histologically confirmed NAFLD who were prospectively recruited according to a standardized protocol at the NAFLD Research Center at Wenzhou Medical University. Recruitment (inclusion/exclusion criteria) of patients was conducted in accordance with the derivation cohort as described above. In this cohort blood tests were performed on the same day as liver biopsy and histology was read by one single experienced hepato-pathologist as described elsewhere (15).

For validation purposes patient data from the European NAFLD Registry (European NAFLD cohort) comprising all five fibrosis stages according to the NASH CRN

classification (F0-F4) were analyzed and the protocol of this prospective, controlled registry study has been published (13). Patients were recruited as already described above. Participants in the European NAFLD Registry from Mainz, Torino and Seville were excluded from these analyses to ensure a fully independent cohort was used.

Ethics

The analysis was conducted according to the ethical guidelines of the 1975 Declaration of Helsinki and its later amendments. The study protocols, including those for the STELLAR trials and the European NAFLD registry, were approved by the responsible ethics committees. Written informed consent was obtained from all participants.

Statistical analysis and modelling

Continuous data are expressed as medians with interquartile ranges (IQR) and comparisons between groups were performed using the Wilcoxon rank sum test. Categorical variables are described as frequencies and percentages and for two-between-group comparisons were made using the chi-square or Fisher's exact tests. Missing values in the data of the derivation cohort were replaced with statistical imputation procedures. For development, validation and reporting of the proposed score, we followed the TRIPOD guideline (16). The score was developed in the framework of logistic regression modelling. To avoid overfitting, we applied an automated variable selection procedure that selected the best predictor-subset out of 16 potential explanatory variables. Therein, the Akaike's information criterion (AIC) was chosen as the criterion of effectiveness and liver enzyme values (gGT, ALT, AST, ALP) were log10-transformed due to their skew distributions. The regression coefficients of the selected model were finally converted to an easy-to-use scoring system.

The discrimination of the final model was assessed in both the development and validation cohorts using areas under receiver operating characteristic (ROC) curves (AUC). Moreover, calibration of the model was assessed on the development data. Additionally, as a comparison, the performance of the Fib-4 was also assessed on the development data (ROC-AUC). Fib-4 was calculated as described elsewhere (17).

The score was finally used to build a classifier discriminating between patients at low (score < c_1), intermediate ($c_1 \leq$ score $\leq c_2$) and high risk (score > c_2) for liver cirrhosis. For the choice of the two cut-offs c_1 and c_2 we provide a scenario where only high- and low-risk patients are directly diagnosed using the score while for diagnosis in the intermediate risk group additional investigations rated as gold standard are undertaken (e.g. transient elastography or liver biopsy). By specifying a sensitivity of 90% and a specificity of likewise 90% as a minimal requirement for that diagnostic algorithm on the one hand and by minimizing the size of the intermediate risk group on the other hand, the determination of the two cut-offs becomes unique. The development data were used to choose the two cut-offs and the performance of the resulting diagnostic algorithm was assessed both on the derivation and the validation data in terms of sensitivity, specificity, positive (PPV) and negative predictive values (NPV).

We also evaluated the prognostic significance of the score in the STELLAR studies. Specifically, Cox proportional hazards regression analyses evaluated associations between the score and progression to cirrhosis (in patients with F3 fibrosis at baseline), liver-related clinical events (in patients with F4 fibrosis at baseline), and MACE (in all patients).

All tests performed in our analyses were two-tailed and a p-value <0.05 was considered statistically relevant. Our complete data analysis is exploratory and aims at prediction rather than causal inference. Hence no adjustments for multiple testing were performed. All statistical analyses were performed in R software version 3.6.1 (R Core

Team, 2019, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistic Version 23.0 (Armonk, NY: IBM Corp.).

Results

Cohort description

In total, 251 patients with F3 and F4 fibrosis were included in the derivation cohort; the external validation cohorts included 1666 patients from the STELLAR trials and 47 patients from Wenzhou Medical University in China and 2058 participants from the European NAFLD registry. Demographics and baseline clinical characteristics of these patients are summarized in Table 1.

Development of a risk prediction model based on the derivation cohort data

Supplementary table 1 provides the results of the univariable analyses for all factors that were considered as potential predictors in the NCS. All potential predictors were subsequently included in an automated variable selection process that resulted in a final logistic regression model containing 5 variables: INR, log10(gGT), log10(ALT), platelets and age. Regression coefficients of the selected model are shown in table 2. They were converted to an easy-to-use scoring system, whose result can be calculated as follows: Score = 25.859 x (INR – 0.86) x 1* + 9.555 x (log10(gGT) – 1.0792) x 1* + 22.409 x (2.6848 – log10(ALT)) x 1* + 0.04 x (664 – platelets) x 1* + 0.394 x (Age – 15) x 1*; *exchange 1 with 0, if e.g. INR is < 0.86, log10(gGT) < 1.0792, log10(ALT) > 2.6848, platelets > 664 or age < 15. A plotted nomogram of the scoring system, herein referred to as the NAFLD Cirrhosis Score (NCS), is illustrated in Figure 1.

Diagnostic performance of the NAFLD Cirrhosis Score in the derivation cohort

The diagnostic performance of the NCS in detecting patients with cirrhosis in the derivation cohort are presented in Figure 2 and supplementary Figure 1. The AUC was 0.733 (95%CI 0.671 – 0.795) with a calibration of the selected prediction model as follows: intercept calibration line: 0.008, 95%CI -0.188 – 0.204; Slope calibration line: 0.979, 95%CI 0.520 – 1.438. As a comparison, we tested the discriminative ability of FIB-4 to detect patients with cirrhosis in the derivation cohort and the AUC was 0.682 (95%CI 0.616 - 0.748).

Next, we identified two cutoff values for the score on the derivation cohort data to define three patient groups: a low-risk group to identify patients without cirrhosis with high certainty in whom further diagnostic workup may be omitted (NCS< 64.5 points), an intermediate-risk group in which patients have to go through additional diagnostic workup, and a high-risk group containing patients who are very likely to suffer from cirrhosis and in whom further testing may be omitted (NCS> 79.17 points) (Figure 3). Cutoffs were chosen in order to achieve a minimum sensitivity and specificity of 90% and allow limiting additional diagnostic workup to the intermediate risk group.

Using the above-mentioned cutoffs and considering that patients in the intermediate-risk group subsequently receive additional diagnostic workup including liver biopsy, the sensitivity, specificity, PPV and NPV of the resulting diagnosis algorithm in the derivation cohort were 90.6%, 90.3%, 85.3% and 93.9%, respectively (Table 3). Applying the NCS as a first test in the advanced fibrosis group, further diagnostic testing could be omitted in 50% of the patients.

External validation of the NAFLD Cirrhosis Score (NCS)

In the STELLAR study cohort, the diagnostic performance of the NCS was confirmed with an AUC of 0.700 (95% CI 0.680-0.730) (Supplementary figure 2). Using the above-mentioned cutoffs and considering that patients in the intermediate-risk group (61.6%

of the patients) subsequently receive additional diagnostic workup, the sensitivity, specificity, PPV and NPV of the NCS in this cohort were 88%, 92%, 93% and 88%, respectively (Table 3). In contrast, the discriminative ability of the FIB-4 to detect patients with cirrhosis in the validation cohort showed an AUC of 0.67 (95%CI 0.64 - 0.70).

For additional validation purposes, the discriminative ability of the NCS was assessed in an independent cohort from China. In this cohort, the diagnostic performance of the NCS exhibited an AUC of 0.727 (95%CI 0.533–0.921). Using the established cutoffs and considering that patients in the intermediate-risk group (34.0% of the patients) subsequently receive additional diagnostic workup, the sensitivity, specificity, PPV and NPV of the diagnosis algorithm in this cohort were 80%, 97%, 88% and 92%, respectively (Table 3).

Prognostic utility of the NAFLD Cirrhosis Score (NCS) to predict liver-related outcomes and MACE

To assess progression to cirrhosis we analyzed the patients of the STELLAR-3 study. Here the median follow-up was 16.5 months (16.2; 16.6). During this follow-up, 14.9% (119/798) of patients with F3 fibrosis at baseline progressed to cirrhosis. When separating the cohort according to the NCS cut-offs, patients in the high-risk group had the highest risk of developing cirrhosis (30.6%, 19/62 patients) (Table 4). The incidence of disease progression in the high-risk group was 2.78-fold and 2.30-fold higher compared to the low-risk or intermediate-risk groups, respectively (Table 4). In patients with cirrhosis at baseline (STELLAR-4 study population), 3.1% (27/868 patients) developed a liver-related clinical event during the median follow-up period of 15.8 months (15.5; 16.2). Patients in the high-risk NCS category had the highest risk of events (8.4%, 18/215) and no patient in the low-risk group progressed to cirrhosis. In

Cox regression analysis, the risk of liver-related events was 5.27-fold larger in the high-risk group than in the intermediate-risk group (Table 4). In a separate cox regression analysis, the NCS as a metric variable maintained its association with the incidence of liver-related events (HR 1.07, 95% CI 1.04 – 1.10, p < 0.001), even after adjusting for signs of portal hypertension defined by varices or portal hypertensive gastropathy (HR 3.19, 95% CI 1.46 – 6.98, p = 0.004). This finding was validated by stratifying the cohort into patients with platelets above and below 150 /nl. In patients with platelets ≥ 150 /nl, the frequency of liver-related events was 1.9% and the HR of the NCS was 1.09 (95% CI 1.01 – 1.17, p = 0.021). In patients with platelets < 150 /nl, the frequency of liver-related events was 4.6% and the HR of the NCS was 1.06 (95% CI 1.03 – 1.10, p < 0.001).

With regards to MACE, a total of 16 patients (1%) in the combined STELLAR 3 and 4 study population experienced an event during the median follow-up time of 15.8 months (15.6; 16.0). The incidence of MACE in the low, intermediate, and high-risk groups within the entire STELLAR trial population was 0.6% (2/363 patients), 1.0% (10/1026), and 1.4% (4/277), respectively. The results of the respective Cox regression analyses are displayed in Table 4.

Utility of the NAFLD Cirrhosis Score (NCS) in a real-world cohort of patients with NAFLD

To explore the utility of the NCS in a cohort encompassing all stages of liver fibrosis from NAFLD, we analyzed data from the European NAFLD Registry (13). In total, data of 2,058 patients with biopsy proven NAFLD from across Europe were included (Table 1). In this cohort comprising all fibrosis stages the diagnostic performance of the NCS

to detect cirrhosis exhibited an AUC of 0.798 (95%CI 0.766-0.830). In comparison, Fib-4 exhibited an AUC of 0.830 (95%CI 0.799-0.861).

To refine the detection of cirrhosis patients we investigated a sequential algorithm consisting of FIB-4 followed by NCS (Figure 4). Using a decision tree and considering that patients in the intermediate-risk group of the NCS (23.0% of the patients) subsequently receive additional diagnostic workup including liver biopsy, the test performance for the diagnosis of liver cirrhosis in the European NAFLD registry was as follows: sensitivity 76% (95% CI 69 – 81), specificity 96% (95% CI 94 – 96), PPV 64% (95% CI 57 – 70), and NPV 97% (95% CI 97 – 98). The respective positive and negative likelihood ratios were 17.0 (95% CI 13.6 – 21.3) and 0.26 (95% CI 0.20 – 0.33), respectively. This translates into 18 additional cases of cirrhosis identified by non-invasive testing in the FIB-4 intermediate risk group, while 13 patients without cirrhosis in the FIB-4 high-risk group are identified by this sequential use of FIB-4 and NCS (Figure 4).

Discussion

The current study describes the development of a non-invasive, easy-to-use, blood-based surrogate score - the NCS - with moderate accuracy to distinguish bridging fibrosis from compensated cirrhosis in NAFLD. The accuracy of the NCS was validated in three independent cohorts, including data from two large phase 3 trials and a large prospective registry study. Importantly, the NCS had an acceptable predictive ability to detect cirrhosis in unselected patients with all grades of fibrosis supporting its use in a sequential diagnostic algorithm with FIB-4 followed by NCS. Lastly, the NCS, which incorporates patient age and the four readily available laboratory parameters ALT, GGT, INR, and platelets, was also able to predict liver-related events during a short follow-up period. This makes the NCS a valuable tool not only in the primary care

settings but also as an adjunct to identify progressing patients with baseline liver biopsy.

Despite an increasing knowledge on the pathophysiology of NAFLD, the lack of reliable non-invasive diagnostics has hampered the development of therapeutics and care pathways to prioritize the large group of affected patients. In the context of limited healthcare resources the task is to provide care to the subgroup of patients at greatest need using easy and inexpensive, but reliable tools. The currently available non-invasive surrogate scores have been largely developed to predict advanced fibrosis combining the histological stages F3 and F4 (18). From a health care system perspective, identification of patients with early liver cirrhosis and separation of these from the larger group of advanced fibrosis stages poses an additional benefit for several reasons. The largest increase in health care expenditures arises in cirrhotic patients (3). Moreover, recent studies indicated that patients with cirrhosis (F4) have a remarkably higher risk for decompensation and mortality when compared to F3 fibrosis (2). The available time span to prevent end-stage liver disease and exponential costs from treatment related to liver transplantation is much shorter compared to all other disease stages (19). Importantly, in the cirrhotic population, screening measures for HCC and esophageal varices have been shown to be cost-effective, while this is not the case for pre-cirrhotic disease stages.

The NCS utilizes the addition of INR and gGT to replace AST in the FIB-4 and results in comparable accuracy in the European NAFLD registry study when applying to all fibrosis stages but outperforms it in the identification of cirrhosis in a preselected F3 and F4 population. Using our proposed two cut-offs system allows the separation of a high-risk from an intermediate- and low-risk group and could avoid invasive diagnostic workup to identify cirrhosis in up to 50% of patients. The overall discriminative ability

of the NCS was moderate with an AUC of 0.733 in the development and 0.700 or 0.727 in the two external validation cohorts. While more sophisticated imaging biomarkers or direct fibrosis markers will likely outperform the NCS in cirrhosis detection, the goal to develop an algorithm that is applicable at all levels of health care settings at a low cost was meet.

Currently, one of the biggest challenges in the management of patients with NAFLD is the identification of patients that develop clinical endpoints and liver-related outcomes. A post-hoc analysis of two large phase 2b trials enrolling patients with F3 and compensated F4 stage observed a progression to cirrhosis in 22% of F3 patients, and liver-related clinical events in 19% of patients with cirrhosis at baseline within 96 weeks (19). In this study, an increase in ELF, FIB-4 and APRI predicted liver outcomes. Also, there is early evidence that repeated measurements of the FIB-4 can in a population-based study predicts disease progression over time (20). Vibration-controlled transient elastography (VCTE) may be another valuable tool, however, VCTE is often not available in non-specialized settings. The current demonstrates that the NCS has the ability to identify patients progressing from advanced fibrosis (F3) and early cirrhosis (F4) supporting its use in risk stratification.

NAFLD is an independent predictor for cardiovascular disease (21) and cardiovascular events are the primary cause of death in patients with NAFLD (22, 23). The current analysis observed a numerically higher rate of MACE in patients in the NCS high-risk group – albeit large and overlapping confidence intervals. The lack of a relevant association between NCS and MACE in the current analysis is likely impacted by type-II error with only 16 patients developing a MACE during follow-up. On the other hand it is plausible that the NCS is related to MACE since individual factors – in particular gGT – have been linked to cardiovascular outcomes (24). Prospective studies in

NAFLD patients with an increased risk of CVD events will need to be explored to answer this question.

The biggest strength of the current analysis is the inclusion of multiple large and well characterized cohorts with biopsy proven NAFLD. The derivation cohort compromises NAFLD patients recruited in a real-world multicenter setting in Europe. The first validation cohort reflects study patients recruited for two global, interventional trials and underlines the applicability of the newly derived score. Additionally, we employed a smaller Asian validation cohort demonstrating robustness of the NCS across ethnic backgrounds. Finally, the largest prospective NAFLD registry study was explored to determine the diagnostic accuracy (13). Overall, one of the main findings of the current analysis is the only moderate accuracy to distinguish the histological stages F3 from F4 in NAFLD using blood-based marker. This is in parts related to the variability of liver histology as a reference standard but also the limitations of the available markers.

The following limitations have to be acknowledged. Despite the diverse study cohorts the derivation groups are derived from tertiary care referral centers. Therefore, the applicability of the NCS in primary and secondary care remains to be established. Additionally, the AUC of the NCS to predict the presence of cirrhosis was only moderate. This is related to the use of accessible parameters used in clinical routine to allow for a resource conserving approach. Conversely, this means that about 50% of the patients would need further diagnostic workup to accurately differentiate between F3 and F4. The AUC highlights the difficulties of noninvasive test to separate advanced fibrosis form cirrhosis and it can be expected that no other blood-based algorithm combining indirect markers of hepatic fibrosis will outperform the currently developed surrogate score. With the emergence of novel, direct fibrosis respectively their increasing availability (8, 25, 26), the NCS could be updated in the near future.

Also, the use of artificial intelligence to segment larger datasets will lead to improved patient identification once developed algorithms are implemented in electronic health care records (27). In the interpretation of the data, the pre-selection of advanced disease stages in the cohorts has to be kept in mind. Additionally, liver histology was scored at the enrolling centers, with only the data derived within the STELLAR studies being read centrally. Lastly, the Chinese validation cohort is comparably small and therefore future validation of the NCS in larger Asian cohorts is warranted.

Conclusions

In conclusion, we developed an easy-to-perform, non-invasive score to distinguish compensated cirrhosis from bridging fibrosis in NAFLD. We propose a sequential two-step diagnostic algorithm with potential use in a primary care setting for the detection of compensated cirrhosis in NAFLD using blood-based testing. By applying the NCS, invasive diagnostic workup could be avoided in up to 50% of the patients with advanced fibrosis. Additionally, the NCS predicted outcome and thus allows to identify patients at greatest need for intensified management or pharmacotherapy therapy in the future.

Abbreviations: BMI, body mass index; CP, Child-Pugh; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; MACE, major adverse cardiovascular events

Acknowledgements

Parts of this manuscript were presented at the International Liver Congress 2021.

Author contributions: Performed research: C.L., M.H.Z., R.P.M., M.R.G., A.A., J.A., E.B., Q.M.A., J.M.S.

Designed the experiments and analysed the data: C.L., D.D., Q.M.A, J.M.S.

Contributed reagents/materials/analysis tools: P.R.G., J.M.S.

Wrote the paper: C.L., R.P.M., J.M.S.

Statistical analysis: C.L., G.T., D.D.

All authors approved the final version of the manuscript and the authorship list.

Guarantor of the article: J.M.S.

References

1. Schattenberg JM, Lazarus JV, Newsome PN, Serfaty L, Aghemo A, Augustin S, Tsochatzis E, et al. Disease burden and economic impact of diagnosed non-alcoholic steatohepatitis in five European countries in 2018: A cost-of-illness analysis. Liver Int 2021;41:1227-1242.
2. Taylor RS, Taylor RJ, Bayliss S, Hagstrom H, Nasr P, Schattenberg JM, Ishigami M, et al. Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. Gastroenterology 2020;158:1611-1625 e1612.
3. O'Hara J, Finnegan A, Dhillon H, Ruiz-Casas L, Pedra G, Franks B, Morgan G, et al. Cost of non-alcoholic steatohepatitis in Europe and the USA: The GAIN study. JHEP Rep 2020;2:100142.
4. Schattenberg JM, Loomba R. Refining Noninvasive Diagnostics In Nonalcoholic Fatty Liver Disease: Closing the Gap to Detect Advanced Fibrosis. Hepatology 2019;69:934-936.
5. Schattenberg JM. Reading the stars for hepatic fibrosis or how to predict the severity of liver disease in patients with NASH. Liver Int 2019;39:812-814.
6. Vali Y, Lee J, Boursier J, Spijker R, Loffler J, Verheij J, Brosnan MJ, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis. J Hepatol 2020.
7. Daniels SJ, Leeming DJ, Eslam M, Hashem AM, Nielsen MJ, Krag A, Karsdal MA, et al. ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. Hepatology 2019;69:1075-1086.
8. Boyle M, Tiniakos D, Schattenberg JM, Ratziu V, Bugianessi E, Petta S, Oliveira CP, et al. Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. JHEP Rep 2019;1:188-198.
9. Eddowes PJ, McDonald N, Davies N, Semple SIK, Kendall TJ, Hodson J, Newsome PN, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2018;47:631-644.
10. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, Guha IN, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2019;156:1717-1730.
11. Augustin S, Pons M, Maurice JB, Bureau C, Stefanescu H, Ney M, Blasco H, et al. Expanding the Baveno VI criteria for the screening of varices in patients with compensated advanced chronic liver disease. Hepatology 2017;66:1980-1988.
12. Labenz C, Huber Y, Kalliga E, Nagel M, Ruckes C, Straub BK, Galle PR, et al. Predictors of advanced fibrosis in non-cirrhotic non-alcoholic fatty liver disease in Germany. Aliment Pharmacol Ther 2018;48:1109-1116.
13. Hardy T, Wonders K, Younes R, Aithal GP, Aller R, Allison M, Bedossa P, et al. The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp Clin Trials 2020;98:106175.
14. Harrison SA, Wong VW, Okanoue T, Bzowej N, Vuppalanchi R, Younes Z, Kohli A, et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: Results from randomized phase III STELLAR trials. J Hepatol 2020;73:26-39.

15. Sun DQ, Zheng KI, Xu G, Ma HL, Zhang HY, Pan XY, Zhu PW, et al. PNPLA3 rs738409 is associated with renal glomerular and tubular injury in NAFLD patients with persistently normal ALT levels. Liver Int 2020;40:107-119.
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.
17. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, M SS, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317-1325.
18. Anstee QM, Lawitz EJ, Alkhouri N, Wong VW, Romero-Gomez M, Okanoue T, Trauner M, et al. Noninvasive Tests Accurately Identify Advanced Fibrosis due to NASH: Baseline Data From the STELLAR Trials. Hepatology 2019;70:1521-1530.
19. Sanyal AJ, Harrison SA, Ratziu V, Abdelmalek MF, Diehl AM, Caldwell S, Shiffman ML, et al. The Natural History of Advanced Fibrosis Due to Nonalcoholic Steatohepatitis: Data From the Simtuzumab Trials. Hepatology 2019;70:1913-1927.
20. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. J Hepatol 2020;73:1023-1029.
21. Labenz C, Huber Y, Michel M, Nagel M, Galle PR, Kostev K, Schattenberg JM. Impact of NAFLD on the Incidence of Cardiovascular Diseases in a Primary Care Population in Germany. Dig Dis Sci 2020;65:2112-2119.
22. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, Sebastiani G, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology 2017;65:1557-1565.
23. Hagstrom H, Nasr P, Ekstedt M, Hammar U, Stal P, Hultcrantz R, Kechagias S. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265-1273.
24. Loomba R, Doycheva I, Bettencourt R, Cohen B, Wassel CL, Brenner D, Barrett-Connor E. Serum gamma-glutamyltranspeptidase predicts all-cause, cardiovascular and liver mortality in older adults. J Clin Exp Hepatol 2013;3:4-11.
25. Are VS, Vuppalanchi R, Vilar-Gomez E, Chalasani N. Enhanced Liver Fibrosis Score Can Be Used to Predict Liver-Related Events in Patients With Nonalcoholic Steatohepatitis and Compensated Cirrhosis. Clin Gastroenterol Hepatol 2020.
26. Huber Y, Pfirrmann D, Gebhardt I, Labenz C, Gehrke N, Straub BK, Ruckes C, et al. Improvement of non-invasive markers of NAFLD from an individualised, web-based exercise program. Aliment Pharmacol Ther 2019;50:930-939.
27. Docherty M, Regnier SA, Capkun G, Balp MM, Ye Q, Janssens N, Tietz A, et al. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. J Am Med Inform Assoc 2021;28:1235-1241.

Table 1. Patient demographics

	Derivation cohort (tertiary care)	Validation cohort (STELLAR)	Validation cohort (China)	European NAFLD cohort
Total number of patients	251	1,666	47	2058
Age in years (median and IQR)	58 (50; 64)	59 (53; 65)	51 (32; 62)	51.49 (43; 61)
Male gender	121 (48.2)	676 (40.6)	27 (57.4)	1222 (59.4%)
F3 Fibrosis F4 (Cirrhosis)	154 (61.4) 97 (38.6)	798 (47.9) 868 (52.1)	10 (21.3) 37 (78.7)	F0 547 (26.6) F1 501 (24.4) F2 447 (21.7) F3 371 (18) F4 192 (9.3)
BMI (median and IQR)	32.0 (28.4; 35.5)	32.8 (28.7; 37.2)	26.1 (24.8; 29;3)	31.43 (28.4; 37)
Metabolic comorbidities				
Arterial hypertension	186 (74.1)	1144 (68.7)	16 (34.0)	1027 (49.9)
Diabetes mellitus type 2	155 (61.8)	1227 (73.6)	41 (87.2)	837 (40.7)
Dyslipidemia		1155 (69.3)		
Hypercholesterinemia	112 (44.6)		17 (36.2)	682 (33.1)
Hypertryglyceridemia	99 (39.4)		25 (53.2)	810 (39.4)
Laboratory values				
Alanine aminotransferase, U/l	60 (40; 94)	48 (33; 70)	40 (25; 78)	57 (38; 86)
Aspartate aminotransferase, U/l	47 (36; 68)	46 (34; 63)	43 (27; 58)	39 (29; 56)
Alkaline phosphatase, U/l	91 (70; 113)	86 (70; 108)	83 (63; 105)	78 (60; 103)
Gamma glutamyl transferase, U/l	90 (45; 201)	68 (42; 120)	52 (32; 75)	66 (38; 125)
Bilirubin, mg/dl	0.70 (0.50; 1.0)	0.57 (0.45; 0.8)	1.4 (0.9; 1.8)	0.58 (0.41; 0.77)
INR	1.0 (1.0; 1.1)	1.0 (1.0; 1.1)	1.0 (1.0; 1.0)	1.0 (0.91; 1.07)
Albumin, g/dl	4.1 (3.9; 4.4)	4.5 (4.3; 4.7)	4.5 (4.1; 4.8)	4.4 (4.2; 4.7)
Platelets, /nl	198 (153; 244)	181 (137; 230)	212 (172; 248)	233 (190; 283)

Clinical characteristics of the derivation and two validation cohorts. Continuous data are expressed as medians and interquartile ranges or frequencies and percentages. Comparisons between groups were performed using the Wilcoxon rank sum test. For two-between-group comparisons chi-square or Fisher's exact tests were used. BMI, body mass index. Missing values in the data of the derivation cohort were replaced with statistical imputation procedures. Platelets x1, bilirubin x5, AST x1, ALT x1, ALP x8, gGT x2, albumin x9, INR x9. Missing values in the European NAFLD cohort: BMI x57, AST x59, ALP x97, bilirubin x21, albumin x110.

Table 2. Regression coefficients and odds ratios

Variable	Odds ratios (95% CI)	Regression coefficient	p-value
INR	8.763 (0.470 – 185.268)	2.171	0.153
log(gGT)	2.230 (1.125 – 4.478)	0.802	0.022
log(ALT)	0.152 (0.048 – 0.459)	-1.881	0.001
Platelets	0.997 (0.993 – 1.000)	-0.003	0.096
Age	1.034 (1.007 – 1.062)	0.033	0.015

Regression coefficients and odds ratios of the final logistic regression model for prediction of liver cirrhosis in the derivation cohort. Unit of platelets: /nl; age is expressed in years. 95% CI, 95% confidence interval.

Table 3. Diagnostic performance in the F3 and F4 population

	Derivation cohort	Validation cohort (STELLAR)	Validation cohort (China)
Sensitivity	91% (83 – 95)	88% (86 – 90)	80% (44 – 96)
Specificity	90% (84 – 94)	92% (90 – 94)	97% (84 – 100)
PPV	85% (76 – 92)	93% (90 – 94)	88% (47 – 99)
NPV	94% (89 – 97)	88% (85 – 90)	92% (78 – 98)
LR+	5.80 (3.61 – 9.32)	12.34 (9.70 – 15.69)	7.00 (1.10 – 44.61)
LR-	0.06 (0.03 – 0.12)	0.14 (0.12 – 0.17)	0.08 (0.03 – 0.25)
Intermediate-risk group	50.4%	61.6%	34.0%
Cutoff 64.5 points:			
Sensitivity	91% (83 – 95)	88% (86 – 90)	80% (44 – 96)
Specificity	48% (40 – 56)	33% (29 – 36)	60% (42 – 75)
Cutoff 79.17 points:			
Sensitivity	27% (19 – 37)	25% (22 – 28)	50% (20 – 80)
Specificity	90% (84 – 94)	92% (90 – 94)	97% (84 – 100)

Diagnostic performance of the NAFLD cirrhosis score for the detection of liver cirrhosis in the derivation and validation cohorts considering cutoff values of <64.5 and >79.17 points with subsequent specialized testing (liver biopsy) in the intermediate group. Data are given as percentages and 95% confidence intervals. NPV, negative predictive value; PPV, positive predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio. Likelihood ratios are given weighted by prevalence.

Table 4. Liver-related and cardiovascular events in the validation cohort.

	NAFLD Cirrhosis Score category		
	Low-risk	**Intermediate-risk**	**High-risk**
Progression to cirrhosis *F3 population*	31/260 (11.9%)	69/476 (14.5%)	19/62 (30.6%)
Liver related events *F4 population*	0/103 (0%)	9/550 (1.6%)	18/215 (8.4%)
F3 (HR vs. low) (outcome: progression to cirrhosis)	N/A	1.21 (0.79; 1.85) p = 0.374	2.78 (1.57; 4.93) p < 0.001
F4 (HR vs. low) (outcome: liver related events)	N/A	Not possible*	Not possible*
F3 (HR vs. intermediate) (outcome: progression to cirrhosis)	0.83 (0.54; 1.26) p = 0.374	N/A	2.30 (1.38; 3.82) p = 0.001
F4 (HR vs. intermediate) (outcome: liver related events)	Not possible*	N/A	5.27 (2.37; 11.73) p < 0.001
MACE	2/363 (0.6%)	10/1026 (1.0%)	4/227 (1.4%)
Total cohort (HR vs. Low)	N/A	1.77 (0.39; 8.08) p= 0.461	2.68 (0.49; 14.62) p= 0.256

Ability of the NAFLD Cirrhosis score (NCS) to predict liver related events and major adverse cardiovascular events (MACE) *regression analyses could not be conducted (0 events in the low-risk group). Data are given as frequencies and percentages. We used a 0.05 level to define statistically relevant deviations from the respective null hypothesis. 95% CI are given in brackets.

NCS, NAFLD Cirrhosis Score; 95% CI, 95% confidence interval; HR, hazard ratio;

Figure 1. Nomogram of the NAFLD Cirrhosis Score

The nomogram depicts the combination of the individual factors of the NAFLD cirrhosis score (NCS). Alanine aminotransferase (ALT), International normalized ratio (INR) gamma glutamyl transferase (gGT).

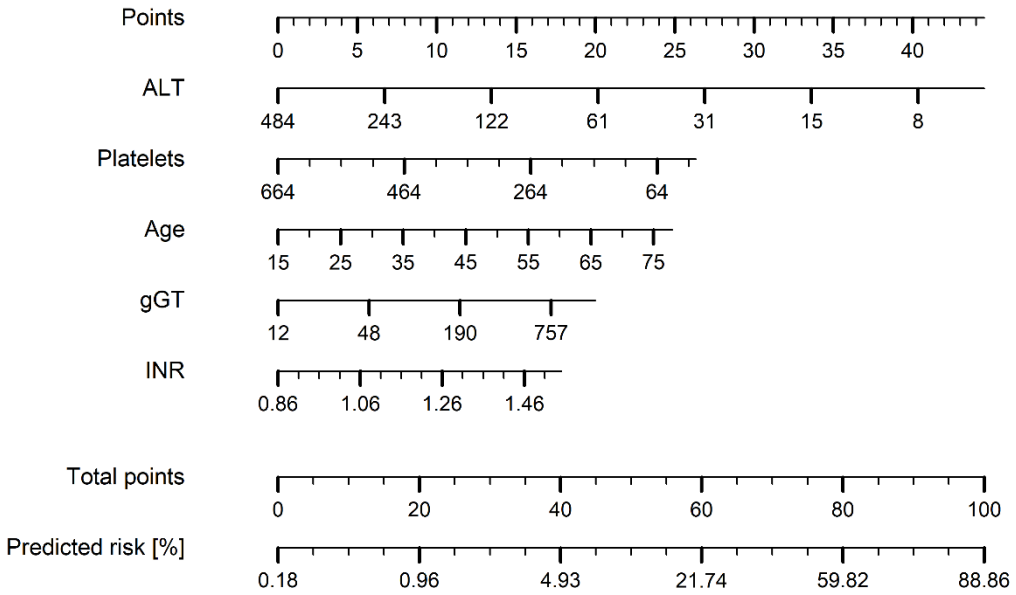


Figure 2. Receiver operating characteristic curve for the NAFLD Cirrhosis Score

Receiver operating characteristic (ROC) curve for the NAFLD Cirrhosis Score in patients with cirrhosis (F4) or advanced fibrosis (F3) in the derivation cohort (n=251).

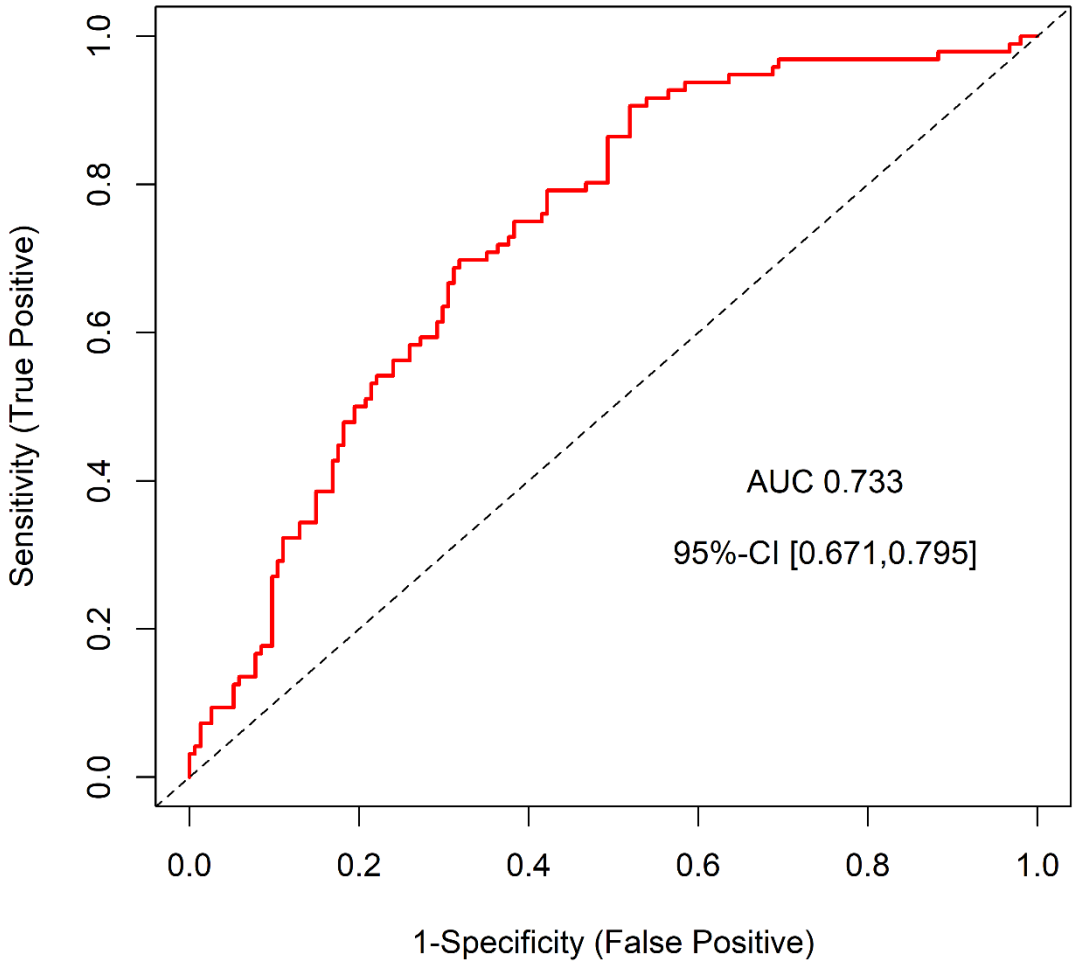


Figure 3. Distribution of the NCS

Distribution of individual patients scores in the F3 (red dot) and F4 (blue dot) population across the NCS bandwidth. Cut-off separate the low (c1) and high (c2) NCS subgroups.

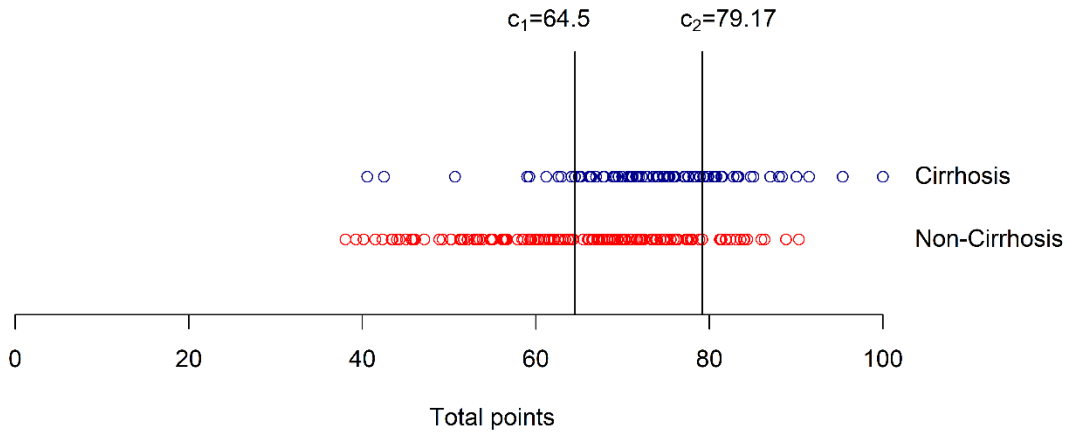
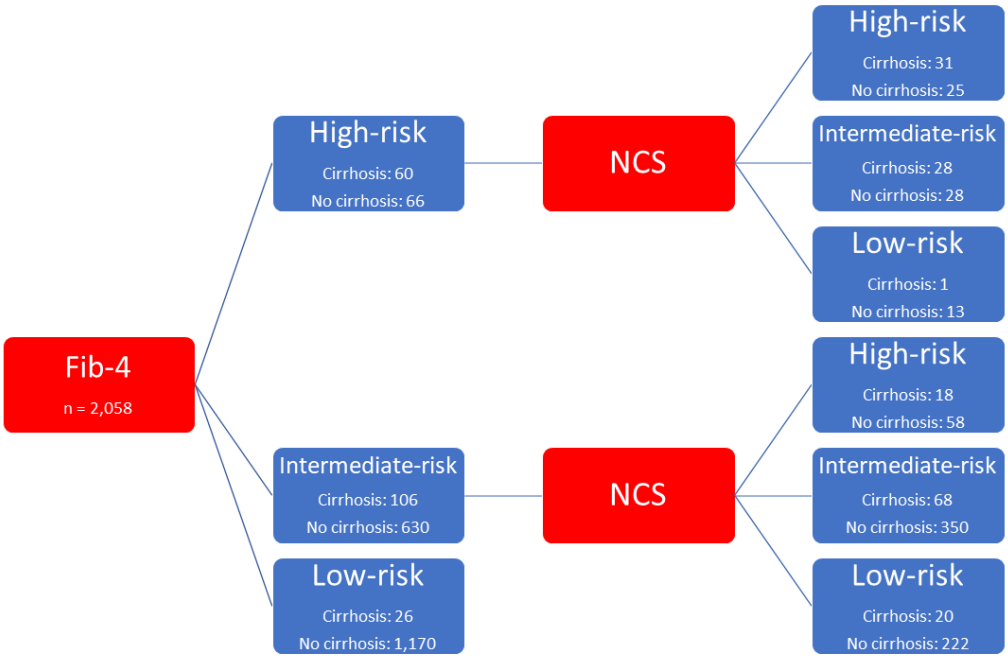


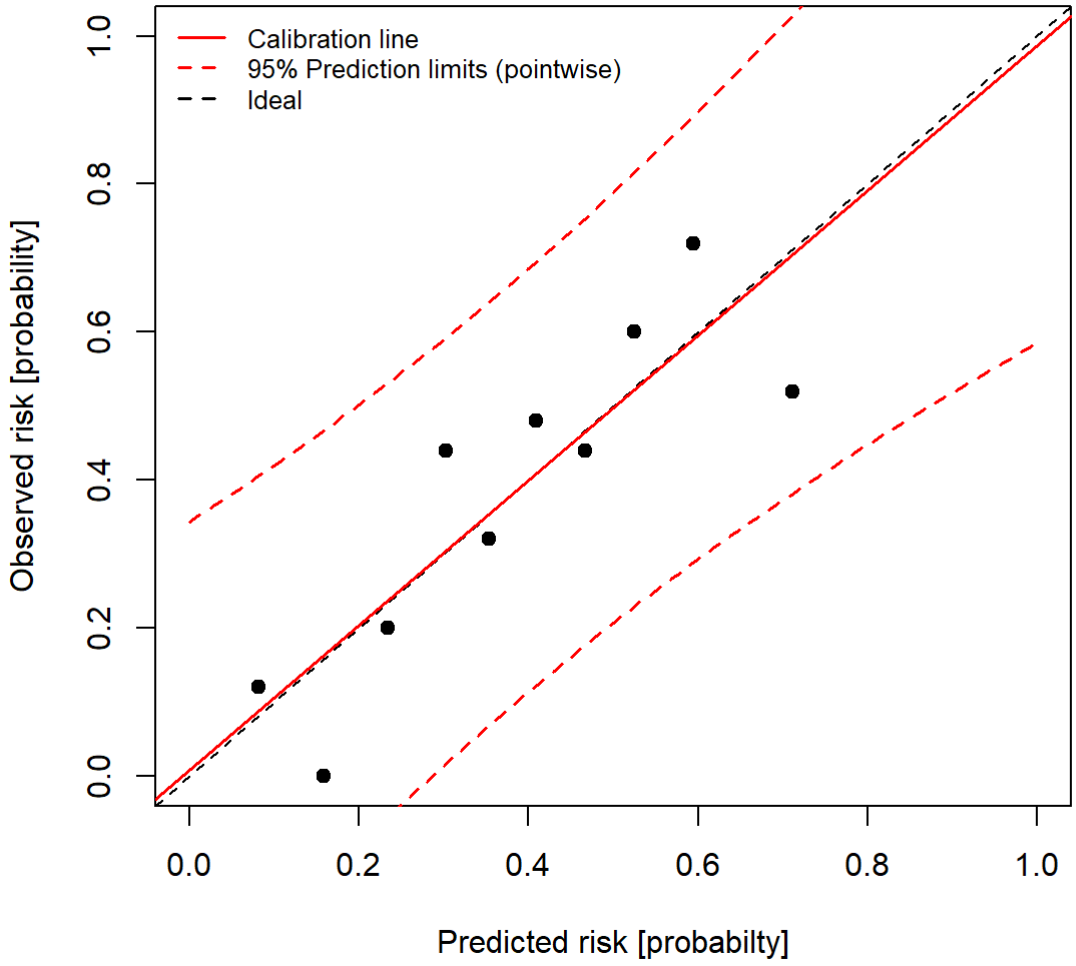
Figure 4. Decision tree including Fib-4 and NCS in the European NAFLD cohort

Step-wise algorithm including Fib-4 and NCS to identify patients with liver cirrhosis in a real-world cohort (European NAFLD cohort) with all grades of liver fibrosis.

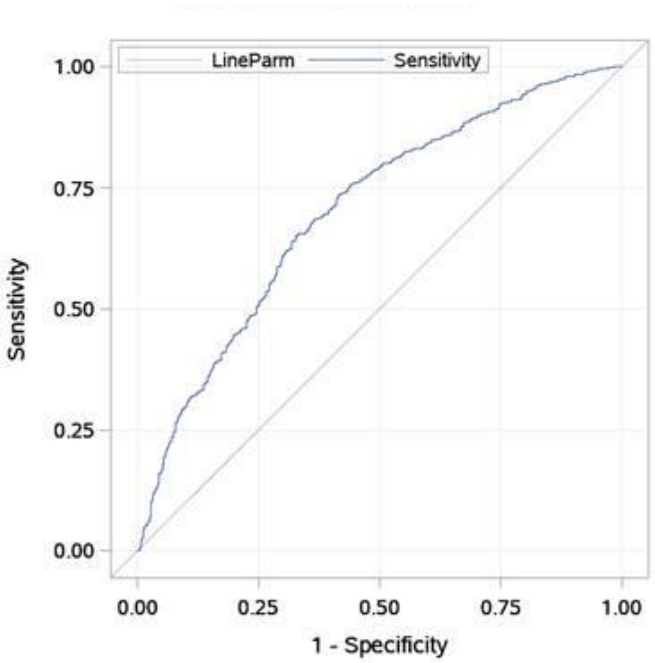


Section/Topic		Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	5
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7, 8
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7, 8
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7, 8
	5b	Describe eligibility criteria for participants.	7, 8
	5c	Give details of treatments received, if relevant.	7, 8
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7, 8
	6b	Report any actions to blind assessment of the outcome to be predicted.	7, 8
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7, 8
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	7, 8
Sample size	8	Explain how the study size was arrived at.	7, 8
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9, 10
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	9, 10
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	9, 10
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9, 10
Risk groups	11	Provide details on how risk groups were created, if done.	n/a
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	11, 12
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	11, 12
Model development	14a	Specify the number of participants and outcome events in each analysis.	11, 12
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	11-14
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	11-14
	15b	Explain how to the use the prediction model.	11-14
Model performance	16	Report performance measures (with CIs) for the prediction model.	11-14
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	18
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	15-17
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15-17
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Supp. doc
Funding	22	Give the source of funding and the role of the funders for the present study.	4

Supplementary figure 1. Calibration of the NAFLD cirrhosis score.



Supplementary figure 2. Receiver operating characteristic curve for the NAFLD cirrhosis score in patients with cirrhosis (F4) or advanced fibrosis (F3) in the STEALLAR validation cohort (AUC 0.700).



Supplementary table 1. Two-group comparisons (advanced fibrosis F3 vs. liver cirrhosis F4) for potential predictors in the derivation cohort.

Variable	p-value
Gender	0.118
Age	<0.001
BMI	0.771
Hypertension	0.010
Diabetes	0.044
Hypercholesterinemia	0.896
Hypertriglyceridemia	0.692
AST	0.143
ALT	<0.001
ALP	0.116
gGT	0.045
Bilirubin	0.767
INR	0.005
Albumin	0.471
Platelets	<0.001

BMI, body mass index.

Derivation and validation of the NAFLD Cirrhosis Score (NCS) to distinguish bridging fibrosis from cirrhosis

Running head: NAFLD cirrhosis score

Christian Labenz[1, 2, 3], Gerrit Toenges[4], Ming-Hua Zheng[5], Dora Ding[6], Robert P. Myers[6], Peter R. Galle[1,3], Angelo Armandi[7], Javier Ampuero[8], Manuel Romero Gómez[8], Elisabetta Bugianesi[7], Quentin M. Anstee[9,10], Jörn M. Schattenberg[1, 2, 3*]

Affiliations:

[1]Department of Internal Medicine I, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany

[2]Metabolic Liver Research Program, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany

[3]Cirrhosis Center Mainz (CCM), University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

[4]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

[5]NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, China

[6]Gilead Sciences, Inc., Foster City, CA, USA

[7]Department of Medical Sciences, Division of Gastroenterology, AOU Citta della Salute e della Scienza, University of Torino, Italy

[8]Digestive Disease Department, Virgen del Rocio University Hospital, Sevilla, Spain

[9]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom.

[10]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, United Kingdom.

* Corresponding author:

Jörn M. Schattenberg, MD

Metabolic Liver Research Program

I. Department of Internal Medicine

University Medical Centre of the Johannes Gutenberg-University

Langenbeckstrasse 1

55131 Mainz, Germany

Telephone: +49 (0) 6131 17 6074

Telefax: +49 (0) 6131 17 477282

E-Mail: joern.schattenberg@unimedizin-mainz.de

Electronic word count: 3,706; tables 4, figures 4

E-Mails:

Christian Labenz: christian.labenz@unimedizin-mainz.de

Gerrit Toenges: gtoenges@unimedizin-mainz.de

Peter R. Galle: peter.galle@unimedizin-mainz.de

Ming-Hua Zheng: zhengmh@wmu.edu.cn

Dora Ding: dora.ding@gilead.com

Robert P. Myers: rob.myers@gilead.com

Angelo Armandi: angelo.armandi@unito.it

Javier Ampuero: jampuero-ibis@us.es

Manuel Romero Gómez: mromerogomez@us.es

Elisabetta Bugianesi: elisabetta.bugianesi@unito.it

Quentin M. Anstee: quentin.anstee@newcastle.ac.uk

Abbreviations: BMI, body mass index; CP, Child-Pugh; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; MACE, major adverse cardiovascular events

Keywords: liver cirrhosis, advanced fibrosis, liver-related events, prognosis, cardiovascular events

Conflict of interest: JMS has acted as consultant to Boehringer Ingelheim, BMS, Echosens, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Madrigal, Novartis, Novo Nordisk, Nordic Bioscience, Pfizer, Roche, Sanofi, Siemens Healthcare GmbH, Zydus. Research Funding: Gilead Sciences. Speakers Bureau: Falk Foundation MSD Sharp & Dohme GmbH. DD and RPM are employees and shareholders of Gilead Sciences, Inc. EB acted as a consultant to Boehringer Ingelheim, BMS, Genfit, Gilead Sciences, Intercept Pharmaceuticals, Novo Nordisk, Pfizer. Research Funding: Gilead Sciences. QMA reports Research Grant Funding: Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex. Active Research Collaborations: (including research collaborations supported through the EU IMI2 LITMUS Consortium*) Abbvie, Antaros Medical*, Allergan/Tobira*, AstraZeneca*, BMS*, Boehringer Ingelheim International GMBH*, Echosens*, Ellegaard Gottingen Minipigs AS*, Eli Lilly & Company Ltd.*, Exalenz Bioscience Ltd.*, Genfit SA*, Glympse Bio, GlaxoSmithKline, HistoIndex*, Intercept Pharma Europe Ltd.*, iXscient Ltd.*, Nordic Bioscience*, Novartis Pharma AG*, Novo Nordisk A/S*, One Way Liver Genomics SL*, Perspectum Diagnostics*, Pfizer Ltd.*, Resoundant*, Sanofi-Aventis Deutschland GMBH*, SomaLogic Inc.*, Takeda Pharmaceuticals International SA*. Consultancy: 89Bio, Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, BMS, BNN Cardio, Celgene,

Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, PathAI, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Terns, Viking Therapeutics. Speaker: Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, MedScape. Royalties: Elsevier Ltd (Davidson's Principles & Practice of Medicine textbook). The other authors have nothing to disclose.

Funding: MRG, EB, QMA and JMS are supported by the LITMUS (Liver Investigation: Testing Biomarker Utility in Steatohepatitis) consortium funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under Grant Agreement 777377; this Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The STELLAR studies were funded by Gilead Sciences, Inc. MRG, EB, QMA and JMS are European NAFLD Registry investigators.

Abstract

Separation of bridging fibrosis from cirrhosis in non-alcoholic fatty liver disease (NAFLD) is critical to guide management. Therefore, it was the aim of this study to develop an easy-to-perform score distinguishing F3 and F4 fibrosis in NAFLD. A derivation cohort comprising 251 NAFLD patients with F3 or F4 was used to develop the NAFLD Cirrhosis Score (NCS). The NCS was validated in three independent cohorts with liver histology comprising 1,666 participants from the STELLAR trials, 47 patients from China and 2,058 patients from the European NAFLD Registry. A model including INR, gGT, ALT, platelets and age discriminated best between patients with bridging fibrosis and cirrhosis with an area under the curve (AUC) of 0.733 (95%CI 0.671–0.795). The diagnostic performance of the NCS was similar in the STELLAR studies (AUC 0.700; 95%CI 0.680-0.730) and a smaller cohort from China (AUC 0.727; 95%CI 0.533–0.921). In the European NAFLD Registry, spanning all histological fibrosis stages, the NCS exhibited an AUC of 0.798 (95%CI 0.766-0.830) to detect cirrhosis. We derived two NCS cut-off values (<64.5 and >79.17) to classify patients at low, intermediate, or high risk for the presence of cirrhosis. Using these cut-offs, further diagnostic workup could be avoided by ruling in or ruling out cirrhosis in approximately half of the patients. Furthermore, NCS identified patients at risk for progression to cirrhosis in the F3 cohort and liver-related outcomes in the F4 cohort.

Conclusion: The NCS is a simple tool to improve the identification of compensated cirrhosis within the large group of advanced disease stage and provides prognostic information. Overall, the differentiation of F3 from F4 disease using standard laboratory remains difficult and does not exceed moderate accuracy.

Introduction

Non-alcoholic fatty liver disease (NAFLD) has become a major health burden (1). Among the large group of patients that are affected, the subgroup progressing to end-stage disease exhibits the highest mortality and health expenses (2, 3).

A number of blood-based non-invasive surrogate scores have been developed to guide in the management of NAFLD. However, none of these separate bridging fibrosis and cirrhosis – defined as the histological stages F3 and F4 according to the NASH-Clinical Research Network (CRN) staging system (4, 5). The addition of direct fibrosis markers, such as the Enhanced Liver Fibrosis (ELF) (6) or PRO-C3 (7, 8) has improved performance but does not add to the ability to identify cirrhosis. As of today, this challenge is only met by imaging modalities including magnetic resonance elastography and transient elastography (9, 10). In addition, the BARVENO VI criteria have been validated extensively but focus on patients with liver cirrhosis and are used in the context of screening for varices and clinical significant portal hypertension (11).

From a clinician's perspective, the availability of an easy-to-use, low-cost, blood-based test to distinguish patients with compensated cirrhosis from bridging (F3) fibrosis holds value for several reasons. First, most patients are treated in primary care and here costly tests (e.g. transient elastography) are often not readily available. Second, there is strong evidence to support screening and surveillance measures in patients with cirrhosis, including screening for varices and hepatocellular carcinoma (HCC), Also, considering future pharmacotherapy labels, the easy separation of pre-cirrhotic from cirrhotic NAFLD could be of importance to identify the subgroup of patients that benefits most – from an efficacy perspective – or are at greatest risk – from a safety perspective. Therefore, a tool with the ability to distinguish between patients with F3 and F4 fibrosis and additionally provide prognostic information related to the risk of disease progression would be of great value.

Materials and methods

Patient cohorts - derivation cohort

In the derivation cohort, 251 adult patients with biopsy-proven NAFLD and fibrosis stages F3 or F4 according to the NASH CRN were included at the University Medical Centers Mainz (Germany; n=122), Torino (Italy; n=97) and Seville (Spain; n=32), as recently described (12). For the current analysis, only patients with fibrosis stages F3 or F4 according to the NASH CRN classification were selected. Patients with other liver diseases, a Child-Turcotte-Pugh Score >6, or significant alcohol use based on clinical grounds or random urine ethylglucuronid measurements were excluded. All liver biopsies were obtained according to local practice and scored by one experienced liver histopathologist in each center (13). All laboratory tests were obtained within 90 days of liver biopsy.

External validation cohorts

For validation purposes, participants in the STELLAR-3 and -4 clinical trials and a real-world cohort from China were used. The STELLAR trials evaluated selonsertib versus placebo in patients with NASH (defined as a NAS of ≥3 with at least grade 1 for each of steatosis, hepatocellular ballooning and lobular inflammation) and bridging fibrosis (F3 for STELLAR-3; NCT03053050) or compensated cirrhosis (F4 for STELLAR-4; NCT03053063) according to the NASH CRN classification. The primary results and methods of these studies are reported elsewhere (14). In the current analysis, patients recruited at the clinical trial sites in Mainz, Torino and Seville were excluded to avoid duplicate inclusion. In both STELLAR studies, the planned duration of treatment was 240 weeks. However, the studies were halted after pre-planned interim analyses conducted after all patients had completed at least 48 weeks of treatment found that

there were no meaningful differences between the active treatment groups or the placebo group in any efficacy endpoint. Therefore, for the purposes of this analysis, treatment groups were combined.

In both STELLAR trials, liver biopsies performed during screening and at week 48 were evaluated by a single central reader blinded to study treatment. In the STELLAR-3 trial, a key endpoint was progression to cirrhosis at week 48, defined as histologic progression to cirrhosis or the development of hepatic decompensation (defined below) during follow-up. In both studies, time to first liver-related clinical event, defined as hepatic decompensation (clinically apparent ascites requiring treatment, hepatic encephalopathy of Grade 2 or above according to the West Haven criteria requiring treatment, and portal hypertension-related gastrointestinal bleeding), liver transplantation, qualification for transplantation (MELD ≥15), or all-cause mortality, as confirmed by an independent Hepatic Events Adjudication Committee, was evaluated. Finally, a Cardiovascular Events Adjudication Committee reviewed all major adverse cardiovascular events (MACE) including cardiovascular death, myocardial infarction, stroke, hospitalization for unstable angina or cardiac failure, and coronary revascularization.

In addition to patients from the STELLAR studies, a second external real-world validation histologically confirmed NAFLD who were prospectively recruited according to a standardized protocol at the NAFLD Research Center at Wenzhou Medical University. Recruitment (inclusion/exclusion criteria) of patients was conducted in accordance with the derivation cohort as described above. In this cohort blood tests were performed on the same day as liver biopsy and histology was read by one single experienced hepato-pathologist as described elsewhere (15).

For validation purposes patient data from the European NAFLD Registry (European NAFLD cohort) comprising all five fibrosis stages according to the NASH CRN

classification (F0-F4) were analyzed and the protocol of this prospective, controlled registry study has been published (13). Patients were recruited as already described above. Participants in the European NAFLD Registry from Mainz, Torino and Seville were excluded from these analyses to ensure a fully independent cohort was used.

Ethics

The analysis was conducted according to the ethical guidelines of the 1975 Declaration of Helsinki and its later amendments. The study protocols, including those for the STELLAR trials and the European NAFLD registry, were approved by the responsible ethics committees. Written informed consent was obtained from all participants.

Statistical analysis and modelling

Continuous data are expressed as medians with interquartile ranges (IQR) and comparisons between groups were performed using the Wilcoxon rank sum test. Categorical variables are described as frequencies and percentages and for two-between-group comparisons were made using the chi-square or Fisher's exact tests. Missing values in the data of the derivation cohort were replaced with statistical imputation procedures. For development, validation and reporting of the proposed score, we followed the TRIPOD guideline (16). The score was developed in the framework of logistic regression modelling. To avoid overfitting, we applied an automated variable selection procedure that selected the best predictor-subset out of 16 potential explanatory variables. Therein, the Akaike's information criterion (AIC) was chosen as the criterion of effectiveness and liver enzyme values (gGT, ALT, AST, ALP) were log10-transformed due to their skew distributions. The regression coefficients of the selected model were finally converted to an easy-to-use scoring system.

The discrimination of the final model was assessed in both the development and validation cohorts using areas under receiver operating characteristic (ROC) curves (AUC). Moreover, calibration of the model was assessed on the development data. Additionally, as a comparison, the performance of the Fib-4 was also assessed on the development data (ROC-AUC). Fib-4 was calculated as described elsewhere (17).

The score was finally used to build a classifier discriminating between patients at low (score < c_1), intermediate ($c_1 \leq$ score $\leq c_2$) and high risk (score > c_2) for liver cirrhosis. For the choice of the two cut-offs c_1 and c_2 we provide a scenario where only high- and low-risk patients are directly diagnosed using the score while for diagnosis in the intermediate risk group additional investigations rated as gold standard are undertaken (e.g. transient elastography or liver biopsy). By specifying a sensitivity of 90% and a specificity of likewise 90% as a minimal requirement for that diagnostic algorithm on the one hand and by minimizing the size of the intermediate risk group on the other hand, the determination of the two cut-offs becomes unique. The development data were used to choose the two cut-offs and the performance of the resulting diagnostic algorithm was assessed both on the derivation and the validation data in terms of sensitivity, specificity, positive (PPV) and negative predictive values (NPV).

We also evaluated the prognostic significance of the score in the STELLAR studies. Specifically, Cox proportional hazards regression analyses evaluated associations between the score and progression to cirrhosis (in patients with F3 fibrosis at baseline), liver-related clinical events (in patients with F4 fibrosis at baseline), and MACE (in all patients).

All tests performed in our analyses were two-tailed and a p-value <0.05 was considered statistically relevant. Our complete data analysis is exploratory and aims at prediction rather than causal inference. Hence no adjustments for multiple testing were performed. All statistical analyses were performed in R software version 3.6.1 (R Core

Team, 2019, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistic Version 23.0 (Armonk, NY: IBM Corp.).

Results

Cohort description

In total, 251 patients with F3 and F4 fibrosis were included in the derivation cohort; the external validation cohorts included 1666 patients from the STELLAR trials and 47 patients from Wenzhou Medical University in China and 2058 participants from the European NAFLD registry. Demographics and baseline clinical characteristics of these patients are summarized in Table 1.

Development of a risk prediction model based on the derivation cohort data

Supplementary table 1 provides the results of the univariable analyses for all factors that were considered as potential predictors in the NCS. All potential predictors were subsequently included in an automated variable selection process that resulted in a final logistic regression model containing 5 variables: INR, log10(gGT), log10(ALT), platelets and age. Regression coefficients of the selected model are shown in table 2. They were converted to an easy-to-use scoring system, whose result can be calculated as follows: Score = 25.859 x (INR – 0.86) x 1* + 9.555 x (log10(gGT) – 1.0792) x 1* + 22.409 x (2.6848 – log10(ALT)) x 1* + 0.04 x (664 – platelets) x 1* + 0.394 x (Age – 15) x 1*; *exchange 1 with 0, if e.g. INR is < 0.86, log10(gGT) < 1.0792, log10(ALT) > 2.6848, platelets > 664 or age < 15. A plotted nomogram of the scoring system, herein referred to as the NAFLD Cirrhosis Score (NCS), is illustrated in Figure 1.

Diagnostic performance of the NAFLD Cirrhosis Score in the derivation cohort

The diagnostic performance of the NCS in detecting patients with cirrhosis in the derivation cohort are presented in Figure 2 and supplementary Figure 1. The AUC was 0.733 (95%CI 0.671 – 0.795) with a calibration of the selected prediction model as follows: intercept calibration line: 0.008, 95%CI -0.188 – 0.204; Slope calibration line: 0.979, 95%CI 0.520 – 1.438. As a comparison, we tested the discriminative ability of FIB-4 to detect patients with cirrhosis in the derivation cohort and the AUC was 0.682 (95%CI 0.616 - 0.748).

Next, we identified two cutoff values for the score on the derivation cohort data to define three patient groups: a low-risk group to identify patients without cirrhosis with high certainty in whom further diagnostic workup may be omitted (NCS< 64.5 points), an intermediate-risk group in which patients have to go through additional diagnostic workup, and a high-risk group containing patients who are very likely to suffer from cirrhosis and in whom further testing may be omitted (NCS> 79.17 points) (Figure 3). Cutoffs were chosen in order to achieve a minimum sensitivity and specificity of 90% and allow limiting additional diagnostic workup to the intermediate risk group.

Using the above-mentioned cutoffs and considering that patients in the intermediate-risk group subsequently receive additional diagnostic workup including liver biopsy, the sensitivity, specificity, PPV and NPV of the resulting diagnosis algorithm in the derivation cohort were 90.6%, 90.3%, 85.3% and 93.9%, respectively (Table 3). Applying the NCS as a first test in the advanced fibrosis group, further diagnostic testing could be omitted in 50% of the patients.

External validation of the NAFLD Cirrhosis Score (NCS)

In the STELLAR study cohort, the diagnostic performance of the NCS was confirmed with an AUC of 0.700 (95% CI 0.680-0.730) (Supplementary figure 2). Using the above-mentioned cutoffs and considering that patients in the intermediate-risk group (61.6%

of the patients) subsequently receive additional diagnostic workup, the sensitivity, specificity, PPV and NPV of the NCS in this cohort were 88%, 92%, 93% and 88%, respectively (Table 3). In contrast, the discriminative ability of the FIB-4 to detect patients with cirrhosis in the validation cohort showed an AUC of 0.67 (95%CI 0.64 - 0.70).

For additional validation purposes, the discriminative ability of the NCS was assessed in an independent cohort from China. In this cohort, the diagnostic performance of the NCS exhibited an AUC of 0.727 (95%CI 0.533–0.921). Using the established cutoffs and considering that patients in the intermediate-risk group (34.0% of the patients) subsequently receive additional diagnostic workup, the sensitivity, specificity, PPV and NPV of the diagnosis algorithm in this cohort were 80%, 97%, 88% and 92%, respectively (Table 3).

Prognostic utility of the NAFLD Cirrhosis Score (NCS) to predict liver-related outcomes and MACE

To assess progression to cirrhosis we analyzed the patients of the STELLAR-3 study. Here the median follow-up was 16.5 months (16.2; 16.6). During this follow-up, 14.9% (119/798) of patients with F3 fibrosis at baseline progressed to cirrhosis. When separating the cohort according to the NCS cut-offs, patients in the high-risk group had the highest risk of developing cirrhosis (30.6%, 19/62 patients) (Table 4). The incidence of disease progression in the high-risk group was 2.78-fold and 2.30-fold higher compared to the low-risk or intermediate-risk groups, respectively (Table 4). In patients with cirrhosis at baseline (STELLAR-4 study population), 3.1% (27/868 patients) developed a liver-related clinical event during the median follow-up period of 15.8 months (15.5; 16.2). Patients in the high-risk NCS category had the highest risk of events (8.4%, 18/215) and no patient in the low-risk group progressed to cirrhosis. In

Cox regression analysis, the risk of liver-related events was 5.27-fold larger in the high-risk group than in the intermediate-risk group (Table 4). In a separate cox regression analysis, the NCS as a metric variable maintained its association with the incidence of liver-related events (HR 1.07, 95% CI 1.04 – 1.10, p < 0.001), even after adjusting for signs of portal hypertension defined by varices or portal hypertensive gastropathy (HR 3.19, 95% CI 1.46 – 6.98, p = 0.004). This finding was validated by stratifying the cohort into patients with platelets above and below 150 /nl. In patients with platelets ≥ 150 /nl, the frequency of liver-related events was 1.9% and the HR of the NCS was 1.09 (95% CI 1.01 – 1.17, p = 0.021). In patients with platelets < 150 /nl, the frequency of liver-related events was 4.6% and the HR of the NCS was 1.06 (95% CI 1.03 – 1.10, p < 0.001).

With regards to MACE, a total of 16 patients (1%) in the combined STELLAR 3 and 4 study population experienced an event during the median follow-up time of 15.8 months (15.6; 16.0). The incidence of MACE in the low, intermediate, and high-risk groups within the entire STELLAR trial population was 0.6% (2/363 patients), 1.0% (10/1026), and 1.4% (4/277), respectively. The results of the respective Cox regression analyses are displayed in Table 4.

Utility of the NAFLD Cirrhosis Score (NCS) in a real-world cohort of patients with NAFLD

To explore the utility of the NCS in a cohort encompassing all stages of liver fibrosis from NAFLD, we analyzed data from the European NAFLD Registry (13). In total, data of 2,058 patients with biopsy proven NAFLD from across Europe were included (Table 1). In this cohort comprising all fibrosis stages the diagnostic performance of the NCS

to detect cirrhosis exhibited an AUC of 0.798 (95%CI 0.766-0.830). In comparison, Fib-4 exhibited an AUC of 0.830 (95%CI 0.799-0.861).

To refine the detection of cirrhosis patients we investigated a sequential algorithm consisting of FIB-4 followed by NCS (Figure 4). Using a decision tree and considering that patients in the intermediate-risk group of the NCS (23.0% of the patients) subsequently receive additional diagnostic workup including liver biopsy, the test performance for the diagnosis of liver cirrhosis in the European NAFLD registry was as follows: sensitivity 76% (95% CI 69 – 81), specificity 96% (95% CI 94 – 96), PPV 64% (95% CI 57 – 70), and NPV 97% (95% CI 97 – 98). The respective positive and negative likelihood ratios were 17.0 (95% CI 13.6 – 21.3) and 0.26 (95% CI 0.20 – 0.33), respectively. This translates into 18 additional cases of cirrhosis identified by non-invasive testing in the FIB-4 intermediate risk group, while 13 patients without cirrhosis in the FIB-4 high-risk group are identified by this sequential use of FIB-4 and NCS (Figure 4).

Discussion

The current study describes the development of a non-invasive, easy-to-use, blood-based surrogate score - the NCS - with moderate accuracy to distinguish bridging fibrosis from compensated cirrhosis in NAFLD. The accuracy of the NCS was validated in three independent cohorts, including data from two large phase 3 trials and a large prospective registry study. Importantly, the NCS had an acceptable predictive ability to detect cirrhosis in unselected patients with all grades of fibrosis supporting its use in a sequential diagnostic algorithm with FIB-4 followed by NCS. Lastly, the NCS, which incorporates patient age and the four readily available laboratory parameters ALT, GGT, INR, and platelets, was also able to predict liver-related events during a short follow-up period. This makes the NCS a valuable tool not only in the primary care

settings but also as an adjunct to identify progressing patients with baseline liver biopsy.

Despite an increasing knowledge on the pathophysiology of NAFLD, the lack of reliable non-invasive diagnostics has hampered the development of therapeutics and care pathways to prioritize the large group of affected patients. In the context of limited healthcare resources the task is to provide care to the subgroup of patients at greatest need using easy and inexpensive, but reliable tools. The currently available non-invasive surrogate scores have been largely developed to predict advanced fibrosis combining the histological stages F3 and F4 (18). From a health care system perspective, identification of patients with early liver cirrhosis and separation of these from the larger group of advanced fibrosis stages poses an additional benefit for several reasons. The largest increase in health care expenditures arises in cirrhotic patients (3). Moreover, recent studies indicated that patients with cirrhosis (F4) have a remarkably higher risk for decompensation and mortality when compared to F3 fibrosis (2). The available time span to prevent end-stage liver disease and exponential costs from treatment related to liver transplantation is much shorter compared to all other disease stages (19). Importantly, in the cirrhotic population, screening measures for HCC and esophageal varices have been shown to be cost-effective, while this is not the case for pre-cirrhotic disease stages.

The NCS utilizes the addition of INR and gGT to replace AST in the FIB-4 and results in comparable accuracy in the European NAFLD registry study when applying to all fibrosis stages but outperforms it in the identification of cirrhosis in a preselected F3 and F4 population. Using our proposed two cut-offs system allows the separation of a high-risk from an intermediate- and low-risk group and could avoid invasive diagnostic workup to identify cirrhosis in up to 50% of patients. The overall discriminative ability

of the NCS was moderate with an AUC of 0.733 in the development and 0.700 or 0.727 in the two external validation cohorts. While more sophisticated imaging biomarkers or direct fibrosis markers will likely outperform the NCS in cirrhosis detection, the goal to develop an algorithm that is applicable at all levels of health care settings at a low cost was meet.

Currently, one of the biggest challenges in the management of patients with NAFLD is the identification of patients that develop clinical endpoints and liver-related outcomes. A post-hoc analysis of two large phase 2b trials enrolling patients with F3 and compensated F4 stage observed a progression to cirrhosis in 22% of F3 patients, and liver-related clinical events in 19% of patients with cirrhosis at baseline within 96 weeks (19). In this study, an increase in ELF, FIB-4 and APRI predicted liver outcomes. Also, there is early evidence that repeated measurements of the FIB-4 can in a population-based study predicts disease progression over time (20). Vibration-controlled transient elastography (VCTE) may be another valuable tool, however, VCTE is often not available in non-specialized settings. The current demonstrates that the NCS has the ability to identify patients progressing from advanced fibrosis (F3) and early cirrhosis (F4) supporting its use in risk stratification.

NAFLD is an independent predictor for cardiovascular disease (21) and cardiovascular events are the primary cause of death in patients with NAFLD (22, 23). The current analysis observed a numerically higher rate of MACE in patients in the NCS high-risk group – albeit large and overlapping confidence intervals. The lack of a relevant association between NCS and MACE in the current analysis is likely impacted by type-II error with only 16 patients developing a MACE during follow-up. On the other hand it is plausible that the NCS is related to MACE since individual factors – in particular gGT – have been linked to cardiovascular outcomes (24). Prospective studies in

NAFLD patients with an increased risk of CVD events will need to be explored to answer this question.

The biggest strength of the current analysis is the inclusion of multiple large and well characterized cohorts with biopsy proven NAFLD. The derivation cohort compromises NAFLD patients recruited in a real-world multicenter setting in Europe. The first validation cohort reflects study patients recruited for two global, interventional trials and underlines the applicability of the newly derived score. Additionally, we employed a smaller Asian validation cohort demonstrating robustness of the NCS across ethnic backgrounds. Finally, the largest prospective NAFLD registry study was explored to determine the diagnostic accuracy (13). Overall, one of the main findings of the current analysis is the only moderate accuracy to distinguish the histological stages F3 from F4 in NAFLD using blood-based marker. This is in parts related to the variability of liver histology as a reference standard but also the limitations of the available markers.

The following limitations have to be acknowledged. Despite the diverse study cohorts the derivation groups are derived from tertiary care referral centers. Therefore, the applicability of the NCS in primary and secondary care remains to be established. Additionally, the AUC of the NCS to predict the presence of cirrhosis was only moderate. This is related to the use of accessible parameters used in clinical routine to allow for a resource conserving approach. Conversely, this means that about 50% of the patients would need further diagnostic workup to accurately differentiate between F3 and F4. The AUC highlights the difficulties of noninvasive test to separate advanced fibrosis form cirrhosis and it can be expected that no other blood-based algorithm combining indirect markers of hepatic fibrosis will outperform the currently developed surrogate score. With the emergence of novel, direct fibrosis respectively their increasing availability (8, 25, 26), the NCS could be updated in the near future.

Also, the use of artificial intelligence to segment larger datasets will lead to improved patient identification once developed algorithms are implemented in electronic health care records (27). In the interpretation of the data, the pre-selection of advanced disease stages in the cohorts has to be kept in mind. Additionally, liver histology was scored at the enrolling centers, with only the data derived within the STELLAR studies being read centrally. Lastly, the Chinese validation cohort is comparably small and therefore future validation of the NCS in larger Asian cohorts is warranted.

Conclusions

In conclusion, we developed an easy-to-perform, non-invasive score to distinguish compensated cirrhosis from bridging fibrosis in NAFLD. We propose a sequential two-step diagnostic algorithm with potential use in a primary care setting for the detection of compensated cirrhosis in NAFLD using blood-based testing. By applying the NCS, invasive diagnostic workup could be avoided in up to 50% of the patients with advanced fibrosis. Additionally, the NCS predicted outcome and thus allows to identify patients at greatest need for intensified management or pharmacotherapy therapy in the future.

Abbreviations: BMI, body mass index; CP, Child-Pugh; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; MACE, major adverse cardiovascular events

Acknowledgements

Parts of this manuscript were presented at the International Liver Congress 2021.

Author contributions: Performed research: C.L., M.H.Z., R.P.M., M.R.G., A.A., J.A., E.B., Q.M.A., J.M.S.

Designed the experiments and analysed the data: C.L., D.D., Q.M.A, J.M.S.

Contributed reagents/materials/analysis tools: P.R.G., J.M.S.

Wrote the paper: C.L., R.P.M., J.M.S.

Statistical analysis: C.L., G.T., D.D.

All authors approved the final version of the manuscript and the authorship list.

Guarantor of the article: J.M.S.

References

1. Schattenberg JM, Lazarus JV, Newsome PN, Serfaty L, Aghemo A, Augustin S, Tsochatzis E, et al. Disease burden and economic impact of diagnosed non-alcoholic steatohepatitis in five European countries in 2018: A cost-of-illness analysis. Liver Int 2021;41:1227-1242.
2. Taylor RS, Taylor RJ, Bayliss S, Hagstrom H, Nasr P, Schattenberg JM, Ishigami M, et al. Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. Gastroenterology 2020;158:1611-1625 e1612.
3. O'Hara J, Finnegan A, Dhillon H, Ruiz-Casas L, Pedra G, Franks B, Morgan G, et al. Cost of non-alcoholic steatohepatitis in Europe and the USA: The GAIN study. JHEP Rep 2020;2:100142.
4. Schattenberg JM, Loomba R. Refining Noninvasive Diagnostics In Nonalcoholic Fatty Liver Disease: Closing the Gap to Detect Advanced Fibrosis. Hepatology 2019;69:934-936.
5. Schattenberg JM. Reading the stars for hepatic fibrosis or how to predict the severity of liver disease in patients with NASH. Liver Int 2019;39:812-814.
6. Vali Y, Lee J, Boursier J, Spijker R, Loffler J, Verheij J, Brosnan MJ, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis. J Hepatol 2020.
7. Daniels SJ, Leeming DJ, Eslam M, Hashem AM, Nielsen MJ, Krag A, Karsdal MA, et al. ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. Hepatology 2019;69:1075-1086.
8. Boyle M, Tiniakos D, Schattenberg JM, Ratziu V, Bugianessi E, Petta S, Oliveira CP, et al. Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. JHEP Rep 2019;1:188-198.
9. Eddowes PJ, McDonald N, Davies N, Semple SIK, Kendall TJ, Hodson J, Newsome PN, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2018;47:631-644.
10. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, Guha IN, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2019;156:1717-1730.
11. Augustin S, Pons M, Maurice JB, Bureau C, Stefanescu H, Ney M, Blasco H, et al. Expanding the Baveno VI criteria for the screening of varices in patients with compensated advanced chronic liver disease. Hepatology 2017;66:1980-1988.
12. Labenz C, Huber Y, Kalliga E, Nagel M, Ruckes C, Straub BK, Galle PR, et al. Predictors of advanced fibrosis in non-cirrhotic non-alcoholic fatty liver disease in Germany. Aliment Pharmacol Ther 2018;48:1109-1116.
13. Hardy T, Wonders K, Younes R, Aithal GP, Aller R, Allison M, Bedossa P, et al. The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp Clin Trials 2020;98:106175.
14. Harrison SA, Wong VW, Okanoue T, Bzowej N, Vuppalanchi R, Younes Z, Kohli A, et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: Results from randomized phase III STELLAR trials. J Hepatol 2020;73:26-39.

15. Sun DQ, Zheng KI, Xu G, Ma HL, Zhang HY, Pan XY, Zhu PW, et al. PNPLA3 rs738409 is associated with renal glomerular and tubular injury in NAFLD patients with persistently normal ALT levels. Liver Int 2020;40:107-119.
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.
17. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, M SS, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 2006;43:1317-1325.
18. Anstee QM, Lawitz EJ, Alkhouri N, Wong VW, Romero-Gomez M, Okanoue T, Trauner M, et al. Noninvasive Tests Accurately Identify Advanced Fibrosis due to NASH: Baseline Data From the STELLAR Trials. Hepatology 2019;70:1521-1530.
19. Sanyal AJ, Harrison SA, Ratziu V, Abdelmalek MF, Diehl AM, Caldwell S, Shiffman ML, et al. The Natural History of Advanced Fibrosis Due to Nonalcoholic Steatohepatitis: Data From the Simtuzumab Trials. Hepatology 2019;70:1913-1927.
20. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. J Hepatol 2020;73:1023-1029.
21. Labenz C, Huber Y, Michel M, Nagel M, Galle PR, Kostev K, Schattenberg JM. Impact of NAFLD on the Incidence of Cardiovascular Diseases in a Primary Care Population in Germany. Dig Dis Sci 2020;65:2112-2119.
22. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, Sebastiani G, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology 2017;65:1557-1565.
23. Hagstrom H, Nasr P, Ekstedt M, Hammar U, Stal P, Hultcrantz R, Kechagias S. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265-1273.
24. Loomba R, Doycheva I, Bettencourt R, Cohen B, Wassel CL, Brenner D, Barrett-Connor E. Serum gamma-glutamyltranspeptidase predicts all-cause, cardiovascular and liver mortality in older adults. J Clin Exp Hepatol 2013;3:4-11.
25. Are VS, Vuppalanchi R, Vilar-Gomez E, Chalasani N. Enhanced Liver Fibrosis Score Can Be Used to Predict Liver-Related Events in Patients With Nonalcoholic Steatohepatitis and Compensated Cirrhosis. Clin Gastroenterol Hepatol 2020.
26. Huber Y, Pfirrmann D, Gebhardt I, Labenz C, Gehrke N, Straub BK, Ruckes C, et al. Improvement of non-invasive markers of NAFLD from an individualised, web-based exercise program. Aliment Pharmacol Ther 2019;50:930-939.
27. Docherty M, Regnier SA, Capkun G, Balp MM, Ye Q, Janssens N, Tietz A, et al. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. J Am Med Inform Assoc 2021;28:1235-1241.

Table 1. Patient demographics

	Derivation cohort (tertiary care)	Validation cohort (STELLAR)	Validation cohort (China)	European NAFLD cohort
Total number of patients	251	1,666	47	2058
Age in years (median and IQR)	58 (50; 64)	59 (53; 65)	51 (32; 62)	51.49 (43; 61)
Male gender	121 (48.2)	676 (40.6)	27 (57.4)	1222 (59.4%)
F3 Fibrosis F4 (Cirrhosis)	154 (61.4) 97 (38.6)	798 (47.9) 868 (52.1)	10 (21.3) 37 (78.7)	F0 547 (26.6) F1 501 (24.4) F2 447 (21.7) F3 371 (18) F4 192 (9.3)
BMI (median and IQR)	32.0 (28.4; 35.5)	32.8 (28.7; 37.2)	26.1 (24.8; 29;3)	31.43 (28.4; 37)
Metabolic comorbidities				
Arterial hypertension	186 (74.1)	1144 (68.7)	16 (34.0)	1027 (49.9)
Diabetes mellitus type 2	155 (61.8)	1227 (73.6)	41 (87.2)	837 (40.7)
Dyslipidemia		1155 (69.3)		
Hypercholesterinemia	112 (44.6)		17 (36.2)	682 (33.1)
Hypertryglyceridemia	99 (39.4)		25 (53.2)	810 (39.4)
Laboratory values				
Alanine aminotransferase, U/l	60 (40; 94)	48 (33; 70)	40 (25; 78)	57 (38; 86)
Aspartate aminotransferase, U/l	47 (36; 68)	46 (34; 63)	43 (27; 58)	39 (29; 56)
Alkaline phosphatase, U/l	91 (70; 113)	86 (70; 108)	83 (63; 105)	78 (60; 103)
Gamma glutamyl transferase, U/l	90 (45; 201)	68 (42; 120)	52 (32; 75)	66 (38; 125)
Bilirubin, mg/dl	0.70 (0.50; 1.0)	0.57 (0.45; 0.8)	1.4 (0.9; 1.8)	0.58 (0.41; 0.77)
INR	1.0 (1.0; 1.1)	1.0 (1.0; 1.1)	1.0 (1.0; 1.0)	1.0 (0.91; 1.07)
Albumin, g/dl	4.1 (3.9; 4.4)	4.5 (4.3; 4.7)	4.5 (4.1; 4.8)	4.4 (4.2; 4.7)
Platelets, /nl	198 (153; 244)	181 (137; 230)	212 (172; 248)	233 (190; 283)

Clinical characteristics of the derivation and two validation cohorts. Continuous data are expressed as medians and interquartile ranges or frequencies and percentages. Comparisons between groups were performed using the Wilcoxon rank sum test. For two-between-group comparisons chi-square or Fisher's exact tests were used. BMI, body mass index. Missing values in the data of the derivation cohort were replaced with statistical imputation procedures. Platelets x1, bilirubin x5, AST x1, ALT x1, ALP x8, gGT x2, albumin x9, INR x9. Missing values in the European NAFLD cohort: BMI x57, AST x59, ALP x97, bilirubin x21, albumin x110.

Table 2. Regression coefficients and odds ratios

Variable	Odds ratios (95% CI)	Regression coefficient	p-value
INR	8.763 (0.470 – 185.268)	2.171	0.153
log(gGT)	2.230 (1.125 – 4.478)	0.802	0.022
log(ALT)	0.152 (0.048 – 0.459)	-1.881	0.001
Platelets	0.997 (0.993 – 1.000)	-0.003	0.096
Age	1.034 (1.007 – 1.062)	0.033	0.015

Regression coefficients and odds ratios of the final logistic regression model for prediction of liver cirrhosis in the derivation cohort. Unit of platelets: /nl; age is expressed in years. 95% CI, 95% confidence interval.

Table 3. Diagnostic performance in the F3 and F4 population

	Derivation cohort	Validation cohort (STELLAR)	Validation cohort (China)
Sensitivity	91% (83 – 95)	88% (86 – 90)	80% (44 – 96)
Specificity	90% (84 – 94)	92% (90 – 94)	97% (84 – 100)
PPV	85% (76 – 92)	93% (90 – 94)	88% (47 – 99)
NPV	94% (89 – 97)	88% (85 – 90)	92% (78 – 98)
LR+	5.80 (3.61 – 9.32)	12.34 (9.70 – 15.69)	7.00 (1.10 – 44.61)
LR-	0.06 (0.03 – 0.12)	0.14 (0.12 – 0.17)	0.08 (0.03 – 0.25)
Intermediate-risk group	50.4%	61.6%	34.0%
Cutoff 64.5 points:			
Sensitivity	91% (83 – 95)	88% (86 – 90)	80% (44 – 96)
Specificity	48% (40 – 56)	33% (29 – 36)	60% (42 – 75)
Cutoff 79.17 points:			
Sensitivity	27% (19 – 37)	25% (22 – 28)	50% (20 – 80)
Specificity	90% (84 – 94)	92% (90 – 94)	97% (84 – 100)

Diagnostic performance of the NAFLD cirrhosis score for the detection of liver cirrhosis in the derivation and validation cohorts considering cutoff values of <64.5 and >79.17 points with subsequent specialized testing (liver biopsy) in the intermediate group. Data are given as percentages and 95% confidence intervals. NPV, negative predictive value; PPV, positive predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio. Likelihood ratios are given weighted by prevalence.

Table 4. Liver-related and cardiovascular events in the validation cohort.

	NAFLD Cirrhosis Score category		
	Low-risk	**Intermediate-risk**	**High-risk**
Progression to cirrhosis *F3 population*	31/260 (11.9%)	69/476 (14.5%)	19/62 (30.6%)
Liver related events *F4 population*	0/103 (0%)	9/550 (1.6%)	18/215 (8.4%)
F3 (HR vs. low) (outcome: progression to cirrhosis)	N/A	1.21 (0.79; 1.85) p = 0.374	2.78 (1.57; 4.93) p < 0.001
F4 (HR vs. low) (outcome: liver related events)	N/A	Not possible*	Not possible*
F3 (HR vs. intermediate) (outcome: progression to cirrhosis)	0.83 (0.54; 1.26) p = 0.374	N/A	2.30 (1.38; 3.82) p = 0.001
F4 (HR vs. intermediate) (outcome: liver related events)	Not possible*	N/A	5.27 (2.37; 11.73) p < 0.001
MACE	2/363 (0.6%)	10/1026 (1.0%)	4/227 (1.4%)
Total cohort (HR vs. Low)	N/A	1.77 (0.39; 8.08) p= 0.461	2.68 (0.49; 14.62) p= 0.256

Ability of the NAFLD Cirrhosis score (NCS) to predict liver related events and major adverse cardiovascular events (MACE) *regression analyses could not be conducted (0 events in the low-risk group). Data are given as frequencies and percentages. We used a 0.05 level to define statistically relevant deviations from the respective null hypothesis. 95% CI are given in brackets.

NCS, NAFLD Cirrhosis Score; 95% CI, 95% confidence interval; HR, hazard ratio;

Figure 1. Nomogram of the NAFLD Cirrhosis Score

The nomogram depicts the combination of the individual factors of the NAFLD cirrhosis score (NCS). Alanine aminotransferase (ALT), International normalized ratio (INR) gamma glutamyl transferase (gGT).

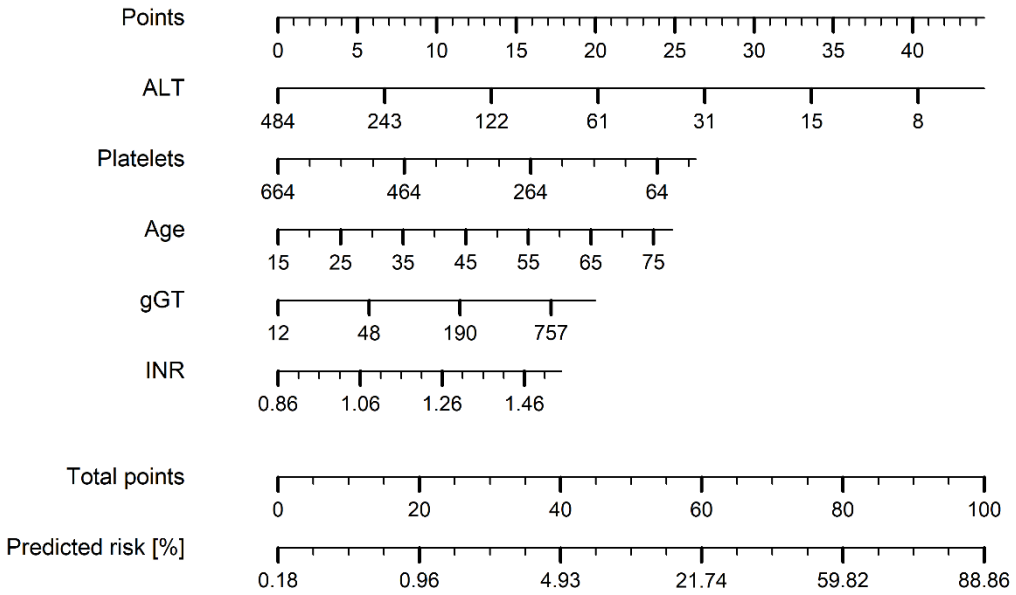


Figure 2. Receiver operating characteristic curve for the NAFLD Cirrhosis Score

Receiver operating characteristic (ROC) curve for the NAFLD Cirrhosis Score in patients with cirrhosis (F4) or advanced fibrosis (F3) in the derivation cohort (n=251).

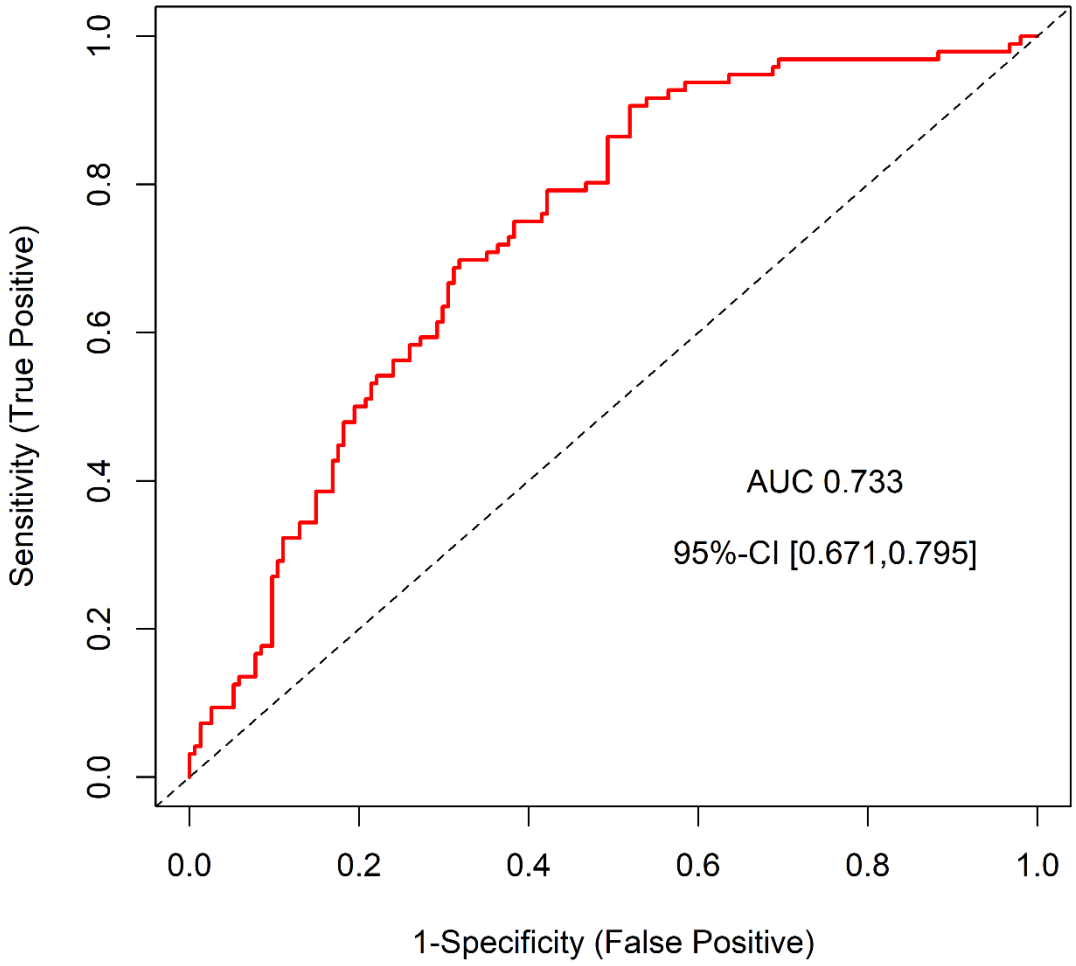


Figure 3. Distribution of the NCS

Distribution of individual patients scores in the F3 (red dot) and F4 (blue dot) population across the NCS bandwidth. Cut-off separate the low (c1) and high (c2) NCS subgroups.

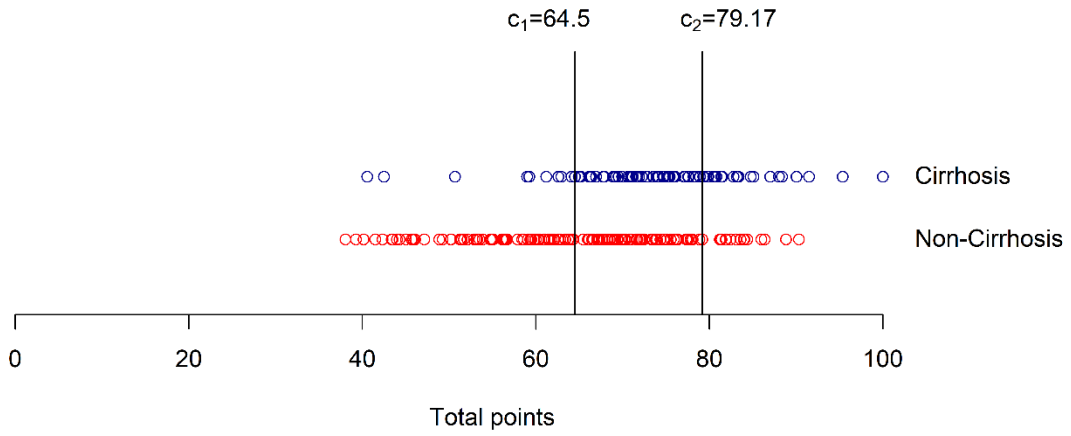
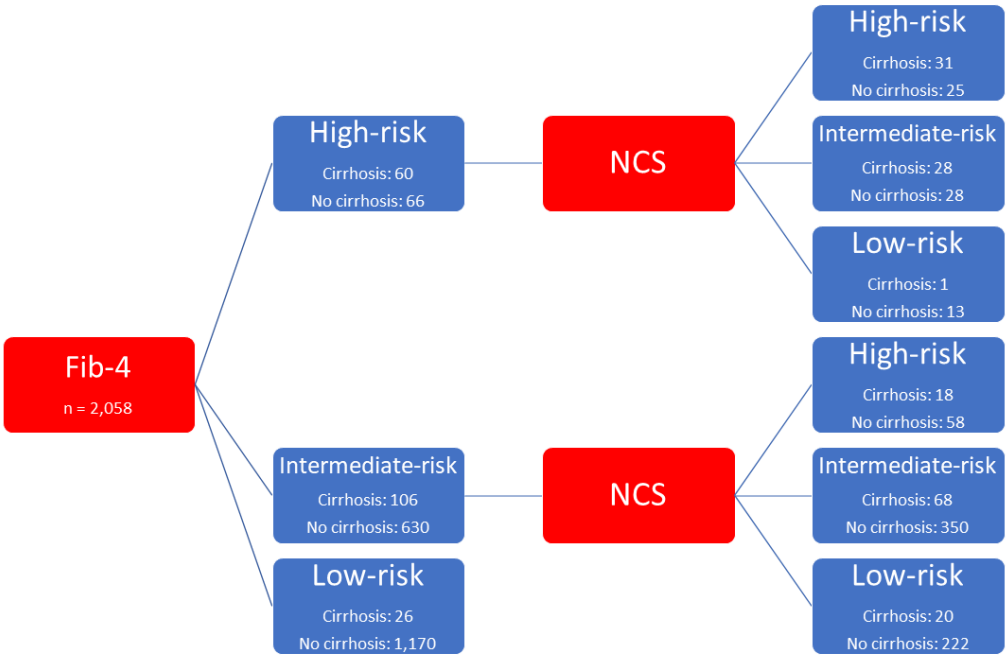


Figure 4. Decision tree including Fib-4 and NCS in the European NAFLD cohort

Step-wise algorithm including Fib-4 and NCS to identify patients with liver cirrhosis in a real-world cohort (European NAFLD cohort) with all grades of liver fibrosis.



Section/Topic		Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	5
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7, 8
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7, 8
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7, 8
	5b	Describe eligibility criteria for participants.	7, 8
	5c	Give details of treatments received, if relevant.	7, 8
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7, 8
	6b	Report any actions to blind assessment of the outcome to be predicted.	7, 8
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7, 8
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	7, 8
Sample size	8	Explain how the study size was arrived at.	7, 8
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9, 10
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	9, 10
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	9, 10
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9, 10
Risk groups	11	Provide details on how risk groups were created, if done.	n/a
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	11, 12
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	11, 12
Model development	14a	Specify the number of participants and outcome events in each analysis.	11, 12
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	11-14
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	11-14
	15b	Explain how to the use the prediction model.	11-14
Model performance	16	Report performance measures (with CIs) for the prediction model.	11-14
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	18
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	15-17
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15-17
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Supp. doc
Funding	22	Give the source of funding and the role of the funders for the present study.	4

Study Highlights:

- Patients with cirrhotic NAFLD exhibit the highest diseases burden among the large population affected.
- The NCS discriminates between F3 and F4 with an AUC of 0.733.
- The NCS identifies patients at risk of disease progression within 48 weeks.
- The NCS exhibits an AUC of 0.798 to detect cirrhosis in referred patients with NAFLD and all histological fibrosis stages.