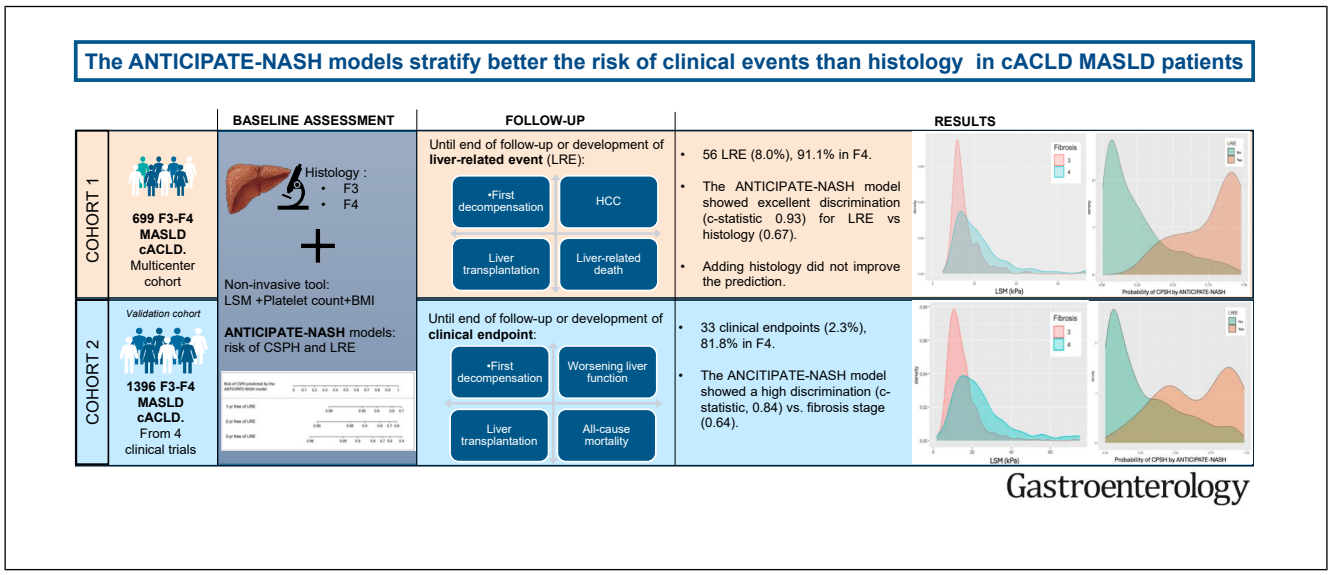
Gastroenterology 2025;■:1–10

The ANTICIPATE-NASH Models Stratify Better the Risk of Clinical Events Than Histology in Metabolic Dysfunction-Associated Steatotic Liver Disease Patients With Advanced Chronic Liver Disease

Laia Aceituno,[1,*] **Juan Bañares**,[1,2,*] Mònica Pons,[1,2] Jesús Rivera-Esteban,[3] Clara Sabiote,[1] Calogero Cammà,[4] Giacinta Ciancimino,[4] Grazia Pennisi,[4] Adele Tulone,[4] Mang M. Ma,[5] Xiangyu Liu,[6] Timothy R. Watkins,[6] Andrew N. Billin,[6] Salvatore Petta,[4] Juan M. Pericàs,[1,2] Juan G. Abraldes,[5] and Joan Genescà[1,2]

[1]Liver Unit, Digestive Diseases Division, Hospital Universitari Vall d'Hebron, Vall d'Hebron Institut de Recerca, Vall d'Hebron Barcelona Hospital Campus, Universitat Autònoma de Barcelona, Barcelona, Spain; [2]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III, Madrid, Spain; [3]Department of Gastroenterology and Hepatology, Hospital Universitario Puerta de Hierro Majadahonda, Instituto de Investigación Sanitaria Puerta Hierro-Segovia Arana (IDIPHISA), Universidad Autónoma de Madrid, Madrid, Spain; [4]Section of Gastroenterology, Dipartimento di Promozione della Salute, Materno-Infantile, di Medicina Interna e Specialistica di Eccellenza "G. D'Alessandro" (PROMISE), University of Palermo, Unità Operativa Complessa (UOC) of Gastroenterology, Azienda Ospedaliera Universitaria Policlinico (AOUP) Paolo Giaccone of Palermo, Palermo, Italy; [5]Liver Unit, Division of Gastroenterology, University of Alberta, Edmonton, Alberta, Canada; and [6]Gilead Sciences, Inc, Foster City, California



BACKGROUND & AIMS: The reference for risk stratification and clinical trial selection of metabolic dysfunction-associated steatotic liver disease (MASLD) patients is fibrosis degree by histology. The noninvasive ANTICIPATE-NASH models have been validated for risk prediction of clinically significant portal hypertension (CSPH) and liver-related events (LRE). We assessed whether these models provide better risk stratification of events than histology. **METHODS:** A multicenter cohort 1, including 699 biopsy specimen-proven F3-F4 patients with MASLD was evaluated. The end point was LRE (hepatic decompensation, hepatocellular carcinoma, transplantation, or liver-related death). We assessed (Cox regression) whether histology provided added value to ANTICIPATE-NASH and whether model predictions differed in F3/F4 patients. Results were validated in cohort 2 (1396 F3-F4 patients) from 4 clinical trials using the clinical regulatory end point. **RESULTS:** In cohort 1, F3 and F4 were equally distributed. There were 56 LREs (8.0%) during follow-up, concentrated in F4 (51 LREs). The ANTICIPATE-NASH model showed excellent discrimination (C statistic, 0.93) for LRE, higher than histology (C statistic, 0.67). Model calibration was excellent. Adding histology did not improve model prediction. Thresholds of ANTICIPATE-NASH above which F3 patients developed LREs and below which F4 patients did not were identified. Results were reproduced in cohort 2 with the regulatory end point, with higher model discrimination (C statistic, 0.84) compared with histology (C statistic, 0.64). **CONCLUSIONS:** In MASLD patients with F3/F4, the noninvasive ANTICIPATE-NASH models

provide better risk stratification of clinical events than histologic classification. These models could be very useful for clinical trials by selecting patients at risk of clinical events and patients with higher chances of observed cirrhosis regression.

Keywords: Predictive Models; Clinical Events; Steatotic Liver Disease.

Predicting the risk of having a clinical condition or clinical events related to that condition is essential for clinical practice. Very often, patients are classified in subgroups with different risks of presenting events (risk stratification) based on characteristics or parameters related to the disease.[1] Then these subgroups, frequently based on numerical thresholds, provide an average estimate of the risk of events. The lower the number of subgroups, the less accurate is the prediction for an individual patient.[2] By contrast, predictive models can provide individual prediction for the whole spectrum of the disease condition. These models to be successful should be validated, easy to use, clinically sound, and if possible, applied to a clinical situation in which interventional or therapeutic decisions must be taken.

Metabolic dysfunction–associated steatotic liver disease (MASLD)[3] is one of the most common causes of chronic liver diseases, currently affecting ∼30% of the worldwide adult population.[4] It has emerged as one of the leading causes of cirrhosis and hepatocellular carcinoma in middle- and high-income countries.[5] MASLD includes a continuum of hepatic lesions from steatosis, with or without inflammation, to fibrosis and ultimately, cirrhosis. A liver biopsy specimen has served as the primary tool for risk stratification and selection of patients for clinical trials in MASLD patients. It has been established that only fibrosis stages F3 and F4 are associated with death due to liver disease.[6] However, histologic classification is invasive, with a high degree of misclassification,[7] and importantly, very inaccurate in predicting an individual risk, because the variability of risks inside each subgroup (F3 or F4) is considerable.

The development of clinically significant portal hypertension (CSPH) is a key driver and a surrogate marker of decompensation and liver-related events (LREs) in patients with compensated advanced chronic liver disease (cACLD),[8,9] including MASLD patients.[10] The development of CSPH and LREs in cACLD MASLD patients seems to follow a more subtle and slow progression than with other etiologies,[6,11] complicating predictions regarding disease outcomes and clinical trial design.[12]

Among the noninvasive tests to predict the presence of CSPH, liver stiffness measurement (LSM), alone or combined with other markers, has been proposed as a preferred tool to identify patients at risk of LRE. The ANTICIPATE model, based on LSM by transient elastography and platelet count, is a validated tool for predicting CSPH in cACLD patients of different etiologies.[13–15] By adding body mass index (BMI) to the model, the ANTICIPATE-NASH model was constructed and validated for predicting CSPH in MASLD patients.[14] Both models are currently recommended in daily clinical practice according to Baveno VII guidelines.[11] Finally, the ANTICIPATE-NASH-LRE model was a subsequently validated adaptation of the original model to predict 3-year risk of LREs in MASLD patients.[16]

We hypothesized that by using the ANTICIPATE-NASH models, a more accurate prediction of LREs compared with liver histology (F3 vs F4) would be achieved in MASLD patients, providing better risk stratification and an

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Stratification of the risk of clinical events and selection of patients for clinical trials with clinical end points in metabolic dysfunction-associated steatotic liver disease patients is usually done by assessing fibrosis degree by histology. Validated predictive models for assessing the risk of clinically significant portal hypertension and clinical events could potentially provide better and more granular risk stratification.

NEW FINDINGS

In metabolic dysfunction-associated steatotic liver disease patients with advanced fibrosis (F3 and F4 fibrosis degree), the ANTICIPATE-NASH models that include liver stiffness, platelet count, and body mass index, stratify the risk of clinical events much better than liver biopsy specimen. Once prediction of events by the model was known, histology was unable to improve it.

LIMITATIONS

The cohorts used were retrospective and partially used in prior studies. Central liver biopsy specimen reading was not available in all patients. The number of clinical events was low.

CLINICAL RESEARCH RELEVANCE

The use of the ANTICIPATE-NASH models for risk stratification and clinical trial selection of metabolic dysfunction-associated steatotic liver disease patients could facilitate risk assessment at point-of-care and reduce sample size and duration of clinical trials with clinical endpoints.

BASIC RESEARCH RELEVANCE

Liver fibrosis and portal hypertension can be reliably assessed by noninvasive methods sparing direct liver tissue analysis.

*** Authors share co-first authorship.**

Abbreviations used in this paper: **BMI, body mass index; cACLD, compensated advanced chronic liver disease; CSPH, clinically significant portal hypertension; LRE, liver-related event; LSM, liver stiffness measurement; MASLD, metabolic dysfunction-associated steatotic liver disease; NASH, nonalcoholic steatohepatitis.**

0016-5085
https://doi.org/10.1053/j.gastro.2025.08.020

improved patient selection for clinical trials. Further, we hypothesized that once ANTICIPATE-NASH is known, histology does not provide additional refinement in risk prediction. Finally, we aimed at testing whether ANTICIPATE-NASH predictions have a different interpretation when coming from patients with F3 and F4. To this end, 2 large cohorts of MASLD patients with F3/F4 fibrosis were consecutively evaluated.

Materials and Methods

Study Cohorts

The first cohort (cohort 1) was a multicentered retrospective cohort composed of MASLD patients from Vall d'Hebron University Hospital in Barcelona and patients from a multicenter international cohort from different European centers, plus 1 each from China and Canada, that had been published before.[17] The patients from Barcelona were part of MASLD patients enrolled from January 2016 to December 2021 and monitored to July 2023 as outpatients, as previously published.[16] During this period, 358 MASLD patients underwent a liver biopsy, and 149 showed F3/F4 fibrosis (Supplementary Figure 1). All patients had MASLD based on current criteria, no decompensation, LSM, and blood tests obtained within 3 months of the liver biopsy.

The multicenter international patients of cohort 1 were MASLD patients with F3/F4 fibrosis on histology or LSM $\geq$10 kPa.[17] In this study, LSM was obtained within 6 months of the histology assessment. The initial cohort included 1039 patients recruited in the different centers from 2004 to 2019, of which 550 with a liver biopsy specimen showing F3/F4 fibrosis were selected. All patients in these 2 studies were monitored at 6-month intervals with laboratory tests and abdominal ultrasound imaging; LREs were recorded during follow-up visits. These 2 sets of patients were merged due to similar origin, composition, and clinical characteristics.

The second large cohort (cohort 2) was used to validate the results obtained by using a slightly different definition of LREs. This cohort was built with F3/F4 biopsy specimen–proven MASLD patients included in 4 different trials already reported in several publications (GS-US-321-0105; GS-US-321-0106; GS-US-384-1943/Safety and Efficacy of Selonsertib in Adults With Nonalcoholic Steatohepatitis [NASH] and Bridging [F3] Fibrosis [STELLAR-3]; GS-US-384-1944/Safety and Efficacy of Selonsertib in Subjects with Compensated Cirrhosis due to Nonalcoholic Steatohepatitis [NASH] STELLAR-4).[18,19] A summary of the main characteristics of these 4 trials is described in Supplementary Table 1.

The 4 trials accounted for 2153 patients, from which 1396 could be included in the present analysis because they had LSM available to calculate the ANTICIPATE-NASH models. Length of follow-up and clinical end point rates were similar between patients with or without LSM (Supplementary Table 2). Additional information regarding cohort composition and overlapping is in the Supplementary Material.

Study Design and Definitions

We aimed to demonstrate the superior capability of the ANTICIPATE-NASH models to predict LREs and the higher accuracy to stratify the risk of LREs in MASLD patients compared with liver histology. Specifically, we wanted to show that once the risk of LREs by the ANTICIPATE models is known, the information of whether the patient presents F3 or F4 in the liver biopsy specimen is irrelevant.

The first cohort with 699 F3/F4 patients from 2 different studies was used to again validate the ANTICIPATE-NASH-LRE model in this histology-driven population and compare the performance of both the model and fibrosis degree for predicting LREs and classifying subgroups of patients at different risk of LRE. For this, we used the LREs definition that was used in the original publication of the model,[16] a composite of liver decompensation (ascites, portal hypertensive bleeding, and hepatic encephalopathy) and hepatocellular carcinoma. Liver transplantation and liver-related death are also included, but they rarely present as primary events. In every case, only the first LRE was considered.

The second cohort with 1396 F3/F4 patients from the clinical trials was used to prove that the same results could be achieved in a very selected population of MASLD patients with a different follow-up time and by using a different clinical event definition. The United States Food and Drug Administration regulatory composite end point was used, defined as hepatic decompensation (as previously defined), liver transplantation, worsening liver function with qualification for transplantation (Model for End-Stage Liver Disease $\geq$15), or all-cause mortality. The use of a similar but different end point or definition of clinical events was intentional to prove the adaptability and high predictability of the ANTICIPATE-NASH models.

Also, to demonstrate the ability of the models to reclassify F3 and F4 patients based on the predicted (and observed) risk of clinical events, we used the median values of ANTICIPATE-NASH of F3 and F4 patients to analyze F4 and F3 patients, respectively.

Finally, we explored the ANTICIPATE-NASH values of the following subgroups of the STELLAR 3 and 4 trials: (1) 84 F3 patients who improved $\geq$1 fibrosis stage at the 48-week follow-up liver biopsy specimen; (2) 88 F3 patients who progressed to F4 at the 48-week follow-up liver biopsy specimen; and (3) 122 F4 patients who improved $\geq$1 fibrosis stage at the 48-week follow-up liver biopsy specimen.[18] Comparisons were made with F3 and F4 patients who changed fibrosis stage.

Assessments

Transient elastography by FibroScan (Echosens, Paris, France) was performed by using the standards of quality and the devices and procedures described in each of the different publications.[16–19] LSM values from all studies were optimized by assuring a <30% (interquartile range/median ratio) variability among the different measures. Test-retest variability was not assessed in the present study.

Histology assessments were performed according to the usual standards, being also extensively described in the different publications.[17–19] For the Barcelona patients, histology was reviewed by our expert liver pathologist, and patients were classified as F3 or F4 according to the NASH Clinical Research Network. Liver biopsy specimens from patients from the multicenter cohort[17] were not centrally assessed, and the histologic information was based on the report from each

participating center. Specimens from patients from the 4 clinical trials of cohort 2 were centrally read as is usually performed in these studies.

The ANTICIPATE-NASH and ANTICIPATE-NASH-LRE values were calculated by using the original formulas as described in the publications,[14,16] also available as online calculators (https://www.bcn-liverhuvh.com/resources).

Statistical Analysis

Analysis in cohort 1 was conducted in R (R Foundation for Statistical Computing) using the packages rms (Harrell Jr FE [2022] rms: Regression Modeling Strategies. R package 6.3-0) and Hmisc (Harrell Jr F [2022]. Hmisc: Harrell Miscellaneous. R package 4.7-1). We used Cox regression to develop risk-prediction models for LREs. We used a time horizon limit of 3 years as in our previous study.[16] The added value of fibrosis to ANTICIPATE-NASH was tested by quantifying the contribution of each variable to the model with the χ^2 test and by comparing models with and without the fibrosis stage with the likelihood ratio, as previously described (https://www.fharrell.com/post/addvalue/).

To assess whether fibrosis stage modifies ANTICIPATE-NASH predictions, we tested the interaction of both variables. Discrimination of the models, which reflects how predictions separate high-risk from low-risk patients (patients with an earlier LRE time should exhibit a higher risk and those with later LRE time or no event a lower risk) was assessed with the C statistic, that was derived from the Somers' Dxy rank correlation (for a censored response variable) computed with the formula C statistic = Dxy/2 + 0.5. Calibration of the model was tested graphically by plotting a smooth calibration curve of the observed event rates against the predicted risks at 3 years. Predicted risk was calculated based on the formulas provided in Pons et al.[16] To quantify calibration we provide (1) the integrated calibration index (mean absolute difference between smoothed observed proportions and predicted probabilities) and (2) the E50 and E90 (median and 90th percentile absolute difference between observed and predicted probabilities of the outcome), using methods and code described in McLernon et al.[20]

The analysis of cohort 2 was conducted in-house by the Gilead statistical team, following a similar strategy. Calibration could not be assessed in this cohort, because follow-up was much shorter than 3 years, and definition of the end point was more inclusive in cohort 2 than the one used to construct ANTICIPATE-NASH-LRE.

Results

Cohort 1 Evaluation

The main characteristics and LRE prevalence during follow-up for the 699 MASLD patients in cohort 1 are shown in Supplementary Table 3. The proportion of F3 and F4 patients was 46.5% vs 53.5%, respectively. As expected, F3 patients were different from F4 patients in many epidemiologic and clinical characteristics. LREs developed during follow-up in 56 patients. Most LREs occurred in F4 patients (51 of 56). Figure 1 shows the distribution of LSM values in F3 and F4 patients. F4 patients presented higher mean LSM values, but with considerable overlap with F3 patients. Accordingly, the predicted risk of CSPH by the ANTICIPATE-NASH model was higher in F4 patients (Supplementary Table 3), but also with high overlap with F3 patients (Figure 1).

Finally, the 56 patients who developed an LRE during follow-up compared with patients who did not, presented higher LSM (34.4 kPa vs 18.1 kPa, $P < .001$), and higher predicted risks of CSPH (62.6% vs 23.32%, $P < .001$) and 3-year LREs (17.7% vs 3.43%, $P < .001$). Figure 2 shows the distribution of ANTICIPATE-NASH in patients with and without LRE within 3 years of follow-up, showing a markedly different distribution of ANTICIPATE-NASH values in patients with and without LREs.

Cohort 1 ANTICIPATE-NASH Models Performance and Added Value of Fibrosis Stage

The ANTICIPATE-NASH-LRE model showed excellent discrimination (C statistic, 0.93) (Table 1) and calibration

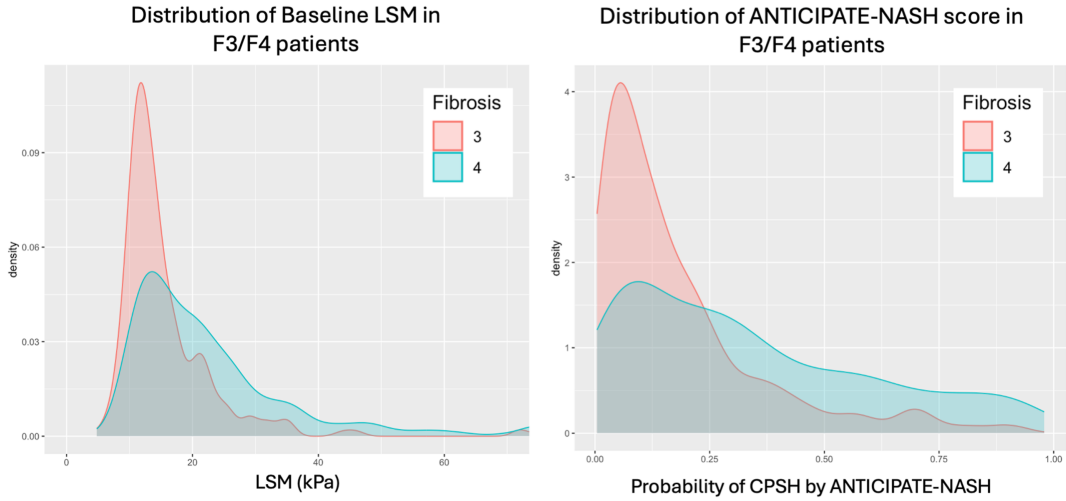


Figure 1. Distribution of (*left*) LSM and (*right*) ANTICIPATE-NASH model values (probability of CSPH) in F3 and F4 patients from cohort 1.

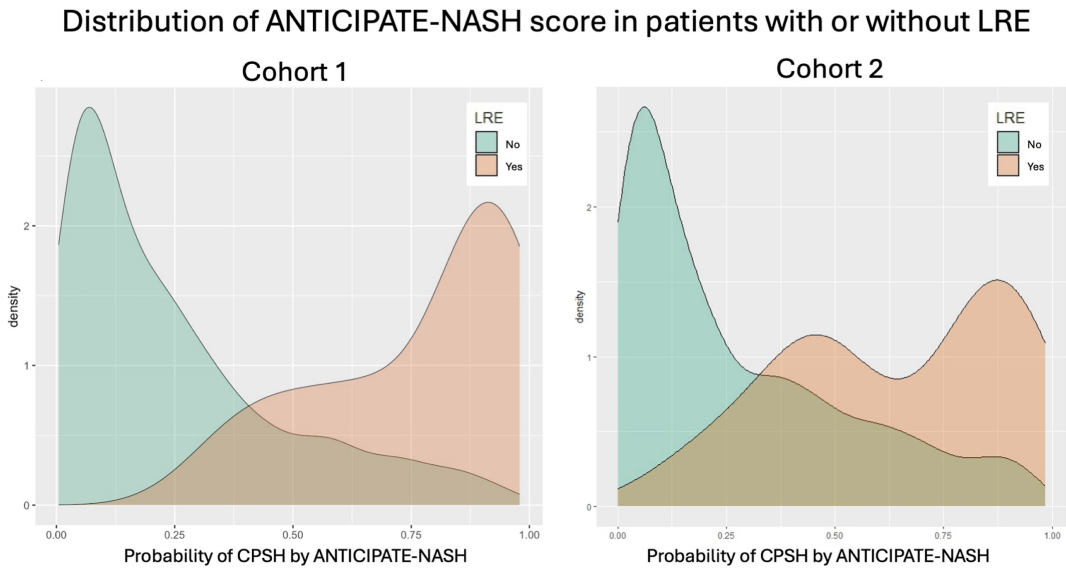


Figure 2. Distribution of the ANTICIPATE-NASH values (probability of CSPH) in patients with and without clinical events (LRE) during follow-up from (*left*) cohort 1 and (*right*) cohort 2.

for LRE at 3 years (Supplementary Figure 2). In contrast, the C statistic for fibrosis stage (F4 vs F3) was 0.67. Supplementary Figures 3, 4, and 5, respectively, show that the model performs well in the subgroup of patients with BMI $\geq$35 kg/m^2, equally in men and women, and better than other noninvasive tests, although the sample size was limited.

We then tested the added value of fibrosis stage to the ANTICIPATE-NASH model by including both in a Cox regression model and comparing them with a model including only ANTICIPATE-NASH. Fibrosis stage did not add relevant information to the ANTICIPATE-NASH model (likelihood ratio tests for the comparison, $P = .55$). Table 1 summarizes the relative contribution of fibrosis stage and ANTICIPATE-NASH to the model.

Next, we tested whether predictions of ANTICIPATE-NASH were affected by fibrosis stage (ie, whether the association between ANTICIPATE-NASH and LREs would change by knowing if it comes from a patient with F3 or F4 stage), by testing the interaction of both in a Cox regression model. Again, neither fibrosis stage ($P = .77$) nor its interaction with ANTICIPATE-NASH ($P = .63$) added value to ANTICIPATE-NASH predictions (Table 1).

Finally, F3 patients with an ANTICIPATE-NASH value of >25% (68 of 324 [20%]), that was the median value of F4 patients, accounted for the 5 LREs seen in F3 patients (Figure 3). By contrast, F4 patients with an ANTICIPATE-NASH value of <12% (104 of 375 [28.5%]), that was the median value of F3 patients, presented no LREs during follow-up. In summary, the ANTICIPATE-NASH model was able to reclassify, based on outcomes, an important number of F3 patients who behaved as F4 patients and reclassify F4 patients who behaved as F3. Supplementary Table 4 compares patients reclassified by the model with patients not reclassified based on ANTICIPATE-NASH values and LRE.

Table 1. Cox Regression Models for the Prediction of Liver-Related Events Up to 3 Years in Cohort 1

Variable	C statistic	Variables	Information contributed by each variable (χ^2)[a]	*P* value
ANTICIPATE-NASH	0.93	ANTICIPATE-NASH	39.8	<.0001
ANTICIPATE-NASH and fibrosis stage	0.93	ANTICIPATE-NASH	35.0	<.0001
		Fibrosis stage	0.4	.55
ANTICIPATE-NASH and fibrosis stage (model with interaction)	0.93	ANTICIPATE-NASH	35.2	<.0001
		Fibrosis stage	0.5	.77
		Interaction	0.2	.63
Fibrosis stage	0.67	Fibrosis stage	7.52	.0061

NOTE. The added value of fibrosis stage (F3/F4) to the noninvasive ANTICIPATE-NASH score was tested. An additive model and a model with interaction were both tested.
[a]The χ^2 statistic quantifies the contribution of information to each parameter in the model.

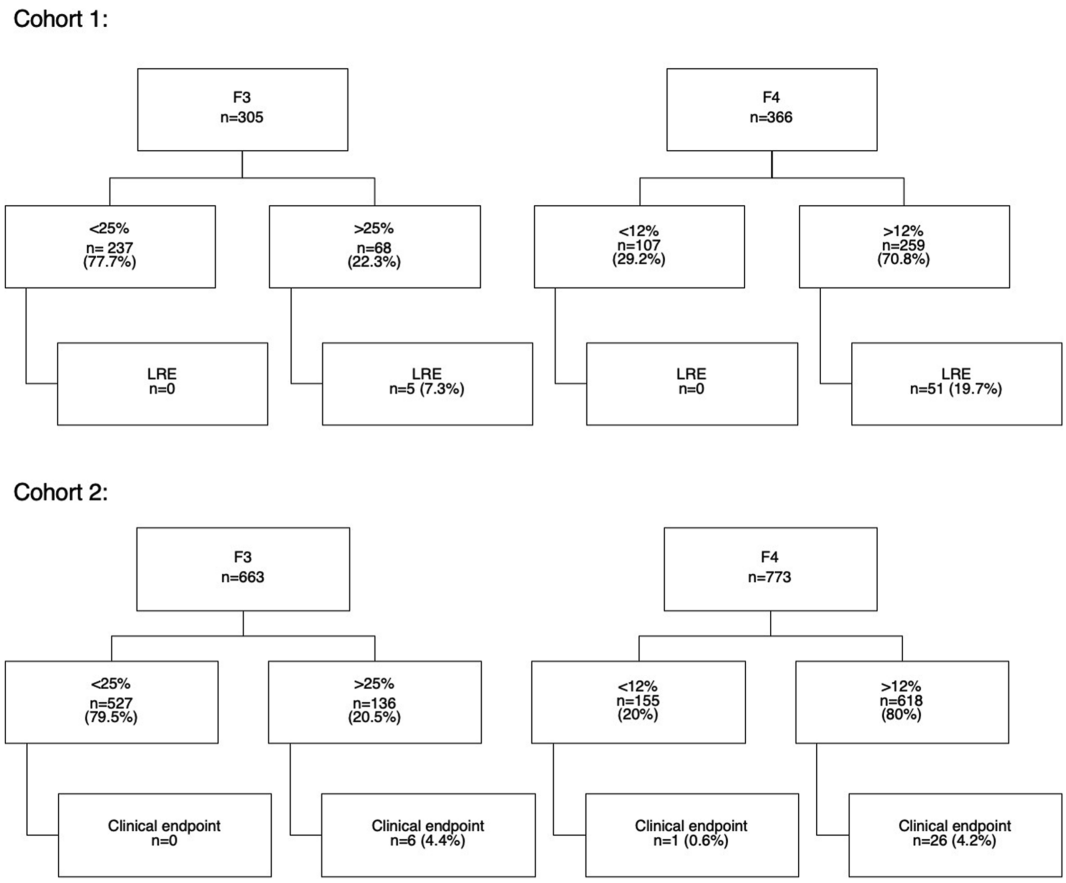


Figure 3. Risk stratification for clinical events of both cohorts, divided by the fibrosis degree (F3 and F4 fibrosis), based on ANTICIPATE-NASH thresholds expressed as the risk (percentage) of having CSPH. Clinical end point: the United States Food and Drug Administration regulatory end point for clinical events.

Cohort 2 Validation

Cohort 2 included 1396 MASLD patients enrolled in the 4 clinical trials. As shown in Supplementary Figure 6, the composite end point developed in 33 patients (2.3%). The median follow-up of this cohort was 16.1 months (interquartile range, 13.8–18.7 months). The distribution of F3 and F4 patients was similar to cohort 1 (47.5% F3 vs 52.5% F4).

Similar to cohort 1, most end points concentrated in F4 patients (27 of 33). The distribution of LSM and ANTICIPATE-NASH values between F3 and F4 patients (Figure 4) was similar to that of cohort 1 (Figure 1). The ANTICIPATE-NASH values were very different between patients with and without the composite end point (Figure 2).

The ANTICIPATE-NASH model also showed a high discrimination for the composite end point in this cohort (C statistic, 0.84), although the C statistic for fibrosis stage (F4 vs F3) was lower (0.64). Similar to what was observed in cohort 1, adding fibrosis stage to the ANTICIPATE-NASH model did not improve performance (C statistic, 0.84).

In addition, as presented in Figure 3, by using the same ANTICIPATE-NASH thresholds for cohort 1, F3 patients with a value >25%, concentrated the 6 clinical events observed in F3 patients, whereas F4 patients with a value <12% developed only 1 of the 27 events in this group.

Finally, Table 2 summarizes the median ANTICIPATE-NASH values of patients from the STELLAR trials according to the occurrence of fibrosis stage changes at the 48-week follow-up liver biopsy specimen. As seen, patients with F3 at baseline who progressed had higher baseline ANTICIPATE-NASH values than patients who remained F3 or who improved to F0-2 in the follow-up biopsy specimen. On the other hand, in patients with F4 at baseline, those regressing to F3 or lower, had a much lower ANTICIPATE-NASH score than those that showed again an F4 biopsy specimen in the follow-up.

Discussion

The results of the present study confirmed our original hypothesis showing that by using the ANTICIPATE-NASH models, a more accurate prediction of LRE compared with liver histology can be achieved in MASLD patients. These models, therefore, could be considered a preferred tool for risk assessment and candidate selection for clinical trials in the setting of cACLD with clinical outcomes as the main end point. Also, our results have shown robustness, considering that they were reproduced in 2 different large cohorts, with different lengths of follow-up, and by using 2 comparable but slightly different end point definitions. Hepatocellular carcinoma was part of our LREs definition

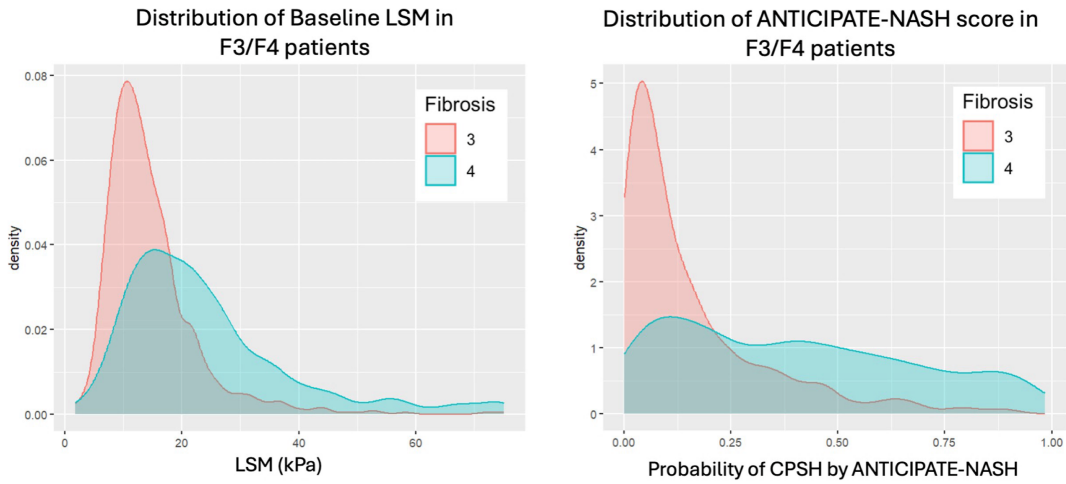


Figure 4. Distribution of (*left*) LSM and (*right*) ANTICIPATE-NASH model values (probability of CSPH) in F3 and F4 patients from cohort 2.

used in cohort 1,[16] whereas it was absent in the regulatory outcome used in cohort 2. Similarly, the regulatory end point included Model for End-Stage Liver Disease ≥15 and all-cause mortality, excluded in our LREs definition. This, in our opinion, reflects the adaptability and versatility of the ANTICIPATE-NASH models in predicting clinical events in cACLD MASLD patients.

The increasing use of noninvasive tools in the assessment and monitoring of chronic liver disease is causing a profound change in its management. Many different noninvasive tools have been developed and applied to assess the stage and prognosis of MASLD patients.[21]

Table 2. ANTICIPATE-NASH Median Values of the Risk of Clinically Significant Portal Hypertension in F3 and F4 Patients

		ANTICIPATE-NASH values
Patient group	No.	(%)
STELLAR-3	619	9.2 (3.9-22)
(a) F3 to F0-2	84	5.5 (2.6-10.9)
(b) F3 unchanged	412	9.3 (4-19.4)
(c) F3 to F4	88	20.9 (8.1-43.4)
STELLAR 4	693	36.8 (15.3-60.2)
(a) F4 to F0-3	122	16.4 (5.6-34.3)
(b) F4 unchanged	507	39.6 (18.8-63.9)

NOTE. Data are presented as the median (25th–75th percentiles). Patients are from the Safety and Efficacy of Selonsertib in Adults With Nonalcoholic Steatohepatitis (NASH) and Bridging (F3) Fibrosis (STELLAR 3) and Safety and Efficacy of Selonsertib in Subjects with Compensated Cirrhosis due to Nonalcoholic Steatohepatitis (NASH) (STELLAR 4) studies. Subgroups indicate improved, unchanged, or progressive fibrosis stage. STELLAR-3: a vs b, $P = .0001$; b vs c, $P < .0001$; STELLAR 4: a vs b, $P < .0001$. All P values were obtained using Wilcoxon's rank sum test.

Although most of them were originally built to stage the liver disease, trying to identify patients with steatohepatitis and fibrosis of different degrees, many have also been adapted to evaluate the risk of clinical events. The Fibrosis 4 (FIB-4) score, nonalcoholic fatty liver disease fibrosis score, and Aspartate Aminotransferase to Platelet Ratio Index score showed good accuracy for predicting LREs in a systematic review including 13 studies,[22] and these results have been repeated in other series. Agile 3+ and Agile 4 are a combination of noninvasive parameters[23,24] developed to detect advanced fibrosis or cirrhosis in MASLD patients, having demonstrated similar accuracy compared with LSM alone to predict LREs. The enhanced liver fibrosis panel has demonstrated a good accuracy for predicting LREs in different etiologies[25]; the <9.8 cutoff identifies a subgroup of patients with a very low risk of LREs.

The ability of LSM for predicting LREs is known since 2011, when Robic et al[26] showed a similar performance of LSM compared with the hepatic venous pressure gradient. Since then, LSM by transient elastography has been validated as a very useful tool to assess prognosis in patients with chronic liver disease.[27–29] The risk of LREs increases in parallel to LSM; however, as shown in 2 meta-analyses[30,31] and by Pons et al,[16] the relationship between LSM and the risk of LREs is not linear, and the slope of the risk of LREs flattens above LSM values of 20 kPa. This indicates that the discriminative capacity for LREs decreases but does not completely disappear, as shown by recent studies.[32]

The ruling-in and ruling-out thresholds of the Baveno VII recommendations for detecting CSPH have also been explored for predicting the risk of LREs with good performance.[33–35] Spleen stiffness measurement alone can also be used for predicting liver decompensation, and by combining Baveno VII criteria with spleen stiffness measurement, it was possible to better classify all the decompensation events in the rule-in group.[33,36] Another approach using the VITRO score (von Willebrand factor/

platelet count), in combination with the Baveno VII criteria, achieved an adequate classification of patients at risk of decompensation.[37]

As explained, risk prediction models might explain better what could be expected for an individual patient based on a parameter or a combination of them rather than subgroups of risk strata. Examples of prognostic models for MASLD patients, include the ABIDE (albumin, bilirubin, international normalized ratio, diabetes, and esophageal varices), that uses varices by endoscopy as a variable and the NAFLD outcomes score (NOS).[38,39] The NOS is a validated model in MASLD patients that showed a good accuracy (area under the curve, 0.90) for predicting LREs at 5 years.

In this regard, our ANTICIPATE-NASH model is a validated tool with good accuracy that can predict both CSPH and LREs at 3 years in MASLD patients. One of the properties of the model is its simplicity, with only 3 easy-to-obtain variables that can be calculated by a nomogram or an online tool. Indeed, one of the important findings of the present work was the ability of the ANTICIPATE-NASH model to reclassify patients. By simply using a binary stratification with certain ANTICIPATE-NASH value thresholds in both F3 and F4 patients and based on clinical events, ∼30% of F4 patients behaved as F3, and 20% of F3 might be classified as F4. The low accuracy and reproducibility of liver biopsy specimen is also evidenced by showing that patients who changed fibrosis stage during follow-up possessed a very different baseline risk of CSPH.

Those who improved fibrosis stage (either F3 or F4) presented lower risk than those who did not change or progressed, whereas patients progressing from F3 to F4 already showed higher risk of CSPH. This probable misclassification of liver biopsy specimen both at baseline and during follow-up might explain why patients from these trials with baseline F4[40] who had cirrhosis regression at follow-up presented many fewer clinical events than patients who maintained F4 stage. Our results suggest that those patients were already different at baseline, with a much lower risk of CSPH and therefore much earlier in the natural history of the disease.

Our study has several strengths. It includes 2 large cohorts of biopsy specimen–proven F3/F4 MASLD patients and results from cohort 1 were reproducible in cohort 2, despite differences in follow-up and clinical end points. We provide enough evidence to demonstrate the superior capacity of the ANTICIPATE-NASH models over liver biopsy for risk stratification, considering, in addition, the easiness, availability, and repeatability of the model compared with the invasiveness of liver biopsy. Although in unselected cACLD MASLD patients, the risk of LREs at 3 years of follow-up is low, ANTICIPATE models may identify a subgroup of patients with higher-risk LREs at 3 years. Indeed, this tool could be used instead of histology to better select patients for the evaluation of clinical events in phase 3 MASLD clinical trials, sensibly decreasing sample size and length of follow-up by entering patients at higher risk of events. An example of the effect on sample size of using the model to enrich patients at higher risk of events is shown on Supplementary Table 5. On the other hand, it could serve to select the F4 patients in whom observing regression to F3 is more likely.

Limitations of our study include that our cohorts have been retrospectively analyzed and recruited from other studies or from ongoing databases not specifically designed for the present study. In most of the cohort 1 patients, central biopsy specimen reading was lacking, and information regarding portal hypertension data (esophageal varices and portal pressure) was not available. The number of events is relatively low. The use of LSM is always an inconvenience for settings or countries lacking this tool, requiring adequate training and dedicated personnel.

Patients from cohort 2 were enrolled in clinical trials for MASLD treatments that were considered ineffective, although a possible effect on the natural history of the disease cannot be excluded. Future studies are needed to prospectively evaluate the use of the ANTICIPATE-NASH models for risk prediction in relation to changes in BMI or the other parameters included in the model. Adaptability in repeated assessments to changes after interventions is an important part of the process of clinical validation of a model.

Conclusion

The ANTICIPATE-NASH model is a noninvasive, point-of-care, accessible method for individual LRE prediction in cACLD-MASLD patients. Knowledge of fibrosis stage (F3 vs F4) does not add relevant prognostic information to ANTICIPATE-NASH. This model could be considered a valuable tool for risk assessment in daily clinical practice and for patient selection or risk stratification in clinical trials.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://dx.doi.org/10.1053/j.gastro.2025.08.020.

References

1. Cowley LE, Farewell DM, Maguire S, et al. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res 2019;3:1–23.
2. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. BJOG 2017;124:423–432.
3. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology 2023;78:1966–1986.
4. Wong VWS, Ekstedt M, Wong GLH, et al. Changing epidemiology, global trends and implications for outcomes of NAFLD. J Hepatol 2023;79:842–852.

5. Powell EE, Wong VWS, Rinella M. Non-alcoholic fatty liver disease. Lancet 2021;397:2212–2224.
6. Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med 2021;385:1559–1569.
7. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol 2020;73:1322–1332.
8. Garcia-Tsao G, Abraldes JG, Berzigotti A, et al. Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the study of liver diseases. Hepatology 2017;65:310–335.
9. Ripoll C, Groszmann R, Garcia-Tsao G, et al. Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. Gastroenterology 2007;133:481–488.
10. Paternostro R, Kwanten WJ, Hofer BS, et al. Hepatic venous pressure gradient predicts risk of hepatic decompensation and liver-related mortality in patients with MASLD. J Hepatol 2024;81:827–836.
11. de Franchis R, Bosch J, Garcia-Tsao G, et al. Baveno VII—renewing consensus in portal hypertension. J Hepatol 2022;76:959–974.
12. Allen AM, Therneau TM, Ahmed OT, et al. Clinical course of non-alcoholic fatty liver disease and the implications for clinical trial design. J Hepatol 2022;77:1237–1245.
13. Abraldes JG, Bureau C, Stefanescu H, et al. Noninvasive tools and risk of clinically significant portal hypertension and varices in compensated cirrhosis: the “Anticipate” study. Hepatology 2016;64:2173–2184.
14. Pons M, Augustin S, Scheiner B, et al. Noninvasive diagnosis of portal hypertension in patients with compensated advanced chronic liver disease. Am J Gastroenterol 2021;116:723–732.
15. Rabiee A, Deng Y, Ciarleglio M, et al. Noninvasive predictors of clinically significant portal hypertension in NASH cirrhosis: validation of ANTICIPATE models and development of a lab-based model. Hepatol Commun 2022;6:3324–3334.
16. Pons M, Esteban JR, Ma MM, et al. Point-of-care noninvasive prediction of liver-related events in patients with NAFLD. Clin Gastroenterol Hepatol 2024; 22:1637–1645.e9.
17. Petta S, Sebastiani G, Viganò M, et al. Monitoring occurrence of liver-related events and survival by transient elastography in patients with nonalcoholic fatty liver disease and compensated advanced chronic liver disease. Clin Gastroenterol Hepatol 2021;19:806–815. e5.
18. Harrison SA, Wong VWS, Okanoue T, 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.
19. Harrison SA, Abdelmalek MF, Caldwell S, et al. Simtuzumab Is Ineffective for patients with bridging fibrosis or compensated cirrhosis caused by nonalcoholic steatohepatitis. Gastroenterology 2018; 155:1140–1153.
20. McLernon DJ, Giardiello D, Van Calster B, et al. Assessing performance and clinical usefulness in prediction models with survival outcomes: practical guidance for Cox proportional hazards models. Ann Intern Med 2023;176:105–114.
21. Dawod S, Brown K. Non-invasive testing in metabolic dysfunction-associated steatotic liver disease. Front Med (Lausanne) 2024;11:1–7.
22. Lee J, Vali Y, Boursier J, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: a systematic review. Liver Int 2021; 41:261–270.
23. Pennisi G, Enea M, Pandolfo A, et al. AGILE 3+ score for the diagnosis of advanced fibrosis and for predicting liver-related events in NAFLD. Clin Gastroenterol Hepatol 2023;21:1293–1302.e5.
24. Sanyal AJ, Foucquier J, Younossi ZM, et al. Enhanced diagnosis of advanced fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile scores. J Hepatol 2023;78:247–259.
25. Are VS, Vuppalanchi R, Vilar-Gomez E, et al. Enhanced liver fibrosis score can be used to predict liver-related events in patients with nonalcoholic steatohepatitis and compensated cirrhosis. Clin Gastroenterol Hepatol 2021;19:1292–1293.e3.
26. Robic MA, Procopet B, Métivier S, et al. Liver stiffness accurately predicts portal hypertension related complications in patients with chronic liver disease: a prospective study. J Hepato 2011;55:1017–1024.
27. Mózes FE, Lee JA, Vali Y, et al. Performance of non-invasive tests and histology for the prediction of clinical outcomes in patients with non-alcoholic fatty liver disease: an individual participant data meta-analysis. Lancet Gastroenterol Hepatol 2023;8:704–713.
28. Lin H, Lee HW, Yip TCF, et al. Vibration-controlled transient elastography scores to predict liver-related events in steatotic liver disease. JAMA 2024; 331:1287–1297.
29. Gawrieh S, Vilar-Gomez E, Wilson LA, et al. Increases and decreases in liver stiffness measurement are independently associated with the risk of liver-related events in NAFLD. J Hepatol 2024;81:600–608.
30. Wang J, Li J, Zhou Q, et al. Liver stiffness measurement predicted liver-related events and all-cause mortality: a systematic review and nonlinear dose-response meta-analysis. Hepatol Commun 2018;2:467–476.
31. Shen Y, Wu SDi, Wu L, et al. The prognostic role of liver stiffness in patients with chronic liver disease: a systematic review and dose-response meta-analysis. Hepatol Int 2019;13:560–572.
32. Vutien P, Barnard Giustini A, et al. Validation and expansion of Baveno VII criteria for cACLD and CSPH based on liver stiffness and platelet count: correlation with risk of hepatic decompensation and death. Hepatology 2025;82:422–437.
33. Lin H, Lai JCT, Wong GLH, et al. Risk and predictors of hepatic decompensation in grey zone patients by the Baveno VII criteria: a competing risk analysis. Aliment Pharmacol Ther 2023;58:920–928.

34. Song BG, Goh MJ, Kang W, et al. Validation of non-invasive diagnosis of CSPH in patients with compensated advanced chronic liver disease according to Baveno VII. Liver Int 2023;43:1966–1974.
35. Wong YJ, Zhaojin C, Tosetti G, et al. Baveno-VII criteria to predict decompensation and initiate non-selective beta-blocker in compensated advanced chronic liver disease patients. Clin Mol Hepatol 2023;29:135–145.
36. Dajti E, Ravaioli F, Marasco G, et al. A combined Baveno VII and spleen stiffness algorithm to improve the noninvasive diagnosis of clinically significant portal hypertension in patients with compensated advanced chronic liver disease. Am J Gastroenterol 2022;117:1825–1833.
37. Jachs M, Hartl L, Simbrunner B, et al. The sequential application of Baveno VII criteria and VITRO score improves diagnosis of clinically significant portal hypertension. Clin Gastroenterol Hepatol 2023;21:1854–1863.e10.
38. Calzadilla-Bertot L, Vilar-Gomez E, Wong VWS, et al. ABIDE: an accurate predictive model of liver decompensation in patients with nonalcoholic fatty liver-related cirrhosis. Hepatology 2021;73:2238–2250.
39. Calzadilla-Bertot L, Jeffrey GP, Wang Z, et al. Predicting liver-related events in NAFLD: a predictive model. Hepatology 2023;78:1240–1251.
40. Sanyal AJ, Anstee QM, Trauner M, et al. Cirrhosis regression is associated with improved clinical outcomes in patients with nonalcoholic steatohepatitis. Hepatology 2022;75:1235–1246.

Received April 24, 2025. Accepted August 12, 2025.

Correspondence
Address correspondence to: Juan G. Abraldes, MD, PhD, Liver Unit, Division of Gastroenterology, University of Alberta, 1-38 Zeidler Ledcor Centre, 8540 112 Street NW, Edmonton, Alberta, T6G 2X8, Canada. e-mail: juan.g.abraldes@ualberta.ca.

CRediT Authorship Contributions
Laia Aceituno Sierra, MD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Project administration: Lead; Resources: Lead; Software: Lead; Supervision: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Juan Bañares, MD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Resources: Equal; Software: Equal; Validation: Equal; Visualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Mònica Pons, MD (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Software: Supporting; Supervision: Supporting; Validation: Supporting; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Jesús Rivera-Esteban, MD (Conceptualization: Supporting; Data curation: Supporting; Project administration: Supporting; Supervision: Supporting; Validation: Supporting; Visualization: Supporting; Writing – review & editing: Supporting)

Clara Sabiote, BSc (Data curation: Supporting; Project administration: Supporting)

Calogero Cammà, MD (Data curation: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Giancita Ciancimino, MD (Data curation: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Grazia Pennisi, MD (Data curation: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Adele Tulone, MD (Data curation: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Mang Ma, MD (Data curation: Supporting; Investigation: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Xiangyu Liu, MD (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Project administration: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Timothy Watkins, MD (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Software: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Andrew Billin, MD (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Project administration: Supporting; Software: Supporting; Supervision: Supporting; Validation: Supporting; Writing – review & editing: Supporting)

Salvatore Petta, MD (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Software: Supporting; Supervision: Supporting; Validation: Supporting; Visualization: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Juan Manuel Pericàs, MD (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Software: Supporting; Supervision: Supporting; Validation: Supporting; Visualization: Supporting; Writing – review & editing: Equal)

Juan G. Abraldes, MD, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Supporting; Investigation: Lead; Methodology: Lead; Project administration: Lead; Resources: Lead; Software: Lead; Supervision: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Joan Genescà, MD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Project administration: Lead; Resources: Lead; Software: Lead; Supervision: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Conflicts of interest
These authors disclose the following: Joan Genescà has received consulting fees from Boehringer Ingelheim and speaking fees from Echosens. Mònica Pons reports speaking fees from CSL Behring and consulting fees from Takeda. Juan G. Abraldes consults for Boehringer Ingelheim, Agomab, 89bio, Boston Pharmaceutics, AstraZeneca, and Novo Nordisk, and received personal speaker fees from Terumo and grants from Gilead and Cook, paid to the University of Alberta. Juan M. Pericàs has received consulting fees from Madrigal, Novo Nordisk, Boehringer-Ingelheim, and MSD, has received speaking fees from Madrigal, Novo Nordisk, Boehringer-Ingelheim, Novartis, and Gilead, travel support from Madrigal, Novo Nordisk and Boehringer-Ingelheim, and educational and research support from Siemens, Gilead, Pfizer, Astellas, Accelerate, Novartis, AbbVie, ViiV, and MSD. Jesús Rivera-Esteban has received consulting and speaking fees from Novo Nordisk. Xiangyu Liu, Timothy R. Watkins, and Andrew N. Billin are employees and shareholders of Gilead Sciences, Inc. The remaining authors disclose no conflicts.

Funding
This work was partially funded by grant PI21/00691 to Joan Genescà from Instituto de Salud Carlos III and cofunded by the European Regional Development Fund/European Social Fund (ERDF/ESF, "A way to make Europe/Investing in your future"). Mònica Pons was funded by a grant contract Joan Rodés from the Instituto de Salud Carlos III. Juan Manuel Pericàs has received a contract grant for Research Intensification from Hospital Universitari Vall d'Hebron and from project PI22/01770 from Instituto de Salud Carlos III. Juan Bañares is funded by a grant contract Río Hortega from the Instituto de Salud Carlos III. Salvatore Petta has received funding from Ministero dell'Istruzione, Università e Ricerca (MIUR) under Planul Național de Redresare și Reziliență (PNRR) M4C2I1.3 Heal Italia project PE00000019 CUP B73C22001250006, PNRR-MAD-2022-12375656 project, Progetti di Rilevante Interesse Nazionale (PRIN) 2022 2022L273C9, and RF-2021-12372399.

Data Availability
The data underlying this study are protected by research agreements between the involved institutions and can only be made available through new agreements with the study investigators. Data from the Gilead trials are the property of Gilead, and sharing these data would require a separate agreement with Gilead. The analytic methods are described in detail within the paper to facilitate reproducibility. The authors are willing to share the analysis code upon request, subject to the necessary terms and conditions.

Gastroenterology, Volume ■

Supplemental information

The ANTICIPATE-NASH Models Stratify Better the Risk of Clinical Events Than Histology in Metabolic Dysfunction-Associated Steatotic Liver Disease Patients With Advanced Chronic Liver Disease

Laia Aceituno, Juan Bañares, Mònica Pons, Jesús Rivera-Esteban, Clara Sabiote, Calogero Cammà, Giacinta Ciancimino, Grazia Pennisi, Adele Tulone, Mang M. Ma, Xiangyu Liu, Timothy R. Watkins, Andrew N. Billin, Salvatore Petta, Juan M. Pericàs, Juan G. Abraldes, and Joan Genescà

Supplementary data

Study cohorts overlapping

The ANTICIPATE-NASH model for CSPH was developed in a multicenter cohort (Pons, et al. Am J Gastro 2021-Ref 17) without any of the patients included in the present study. This model was subsequently validated in many studies (including Rabiee, et al. Hepatol Commun 2022-Ref 16). Then the ANTICIPATE-NASH model was shown to predict 3-year LRE (Pons, et al. Clin Gastro and Hepatol 2024-Ref 18). Of note, the ANTICIPATE-NASH model was used as it was previously developed, without any model re-fitting. What we called ANTICIPATE-NASH-LRE was the formulation to predict LREs. That study included in a large sample of MASLD patients (2638 patients) from Spain-Vall d'Hebron University Hospital (348 patients), Canada-University of Alberta (72 patients) and from a Chinese/French cohort previously published (Shili-Masmoudi, et al. Liver Int 2020-2218 patients).
In the present study, 106 patients of 149 from Vall d'Hebron University Hospital overlap with the ones used in the Clin Gastro Hepatol paper.
For the validation cohort of the study, 679 cACLD MASLD patients of the total of 1039 patients from the multicenter cohort of Petta, et al. (Ref 19) were used. In the present study, 332 patients of the 550 from Pettta, et al. study overlap with the ones used in the Clin Gastro Hepatol paper.
In total, in the present study 261 are new patients and 438 were used before in the Pons 2023 study. Again, we emphasize that none of the patients in the present study were used in the development of ANTICIPATE-NASH.

Tables and figures

Table 1 – Main characteristics of the 4 clinical trials included in the validation cohort

Study	Endpoint	Inclusion criteria		N	Patients included in the analysis*
GS-US-321-0105 (Simtuzumab)	Change in hepatic collagen content (study 105) and change in HVPG (106) + Event free survival: Time to progression to cirrhosis in study 105 and the first liver-related clinical event in study 106	18–65 years old. Body mass index of at least 18 kg/m^2 Chronic liver disease due to NASH, defined as macrovesicular steatosis involving > 5% of hepatocytes on a liver biopsy with associated lobular inflammation. Aspartate and alanine aminotransferase ≤ 10 x Central Laboratory Upper Limit of Normal. Serum creatinine < 2.0 mg/dL.	Bridging fibrosis (stage 3) or marked bridging fibrosis (stage 4) based on a modified Ishak classification.	218	45
GS-US-321-0106 (Simtuzumab)			Compensated cirrhosis, defined histologically as Ishak fibrosis stage 5 (early or incomplete cirrhosis) or 6 (established or advanced cirrhosis). Or cirrhosis without these histologic features but with at least 1 clinical feature suggestive of NASH (eg, diabetes mellitus, overweight or obesity, dyslipidemia).	256	39
GS-US-384-1943/STELLAR-3 (Selonsertib)	≥1 stage improvement in fibrosis on liver biopsy without worsening of NASH + Time to first clinical event + Progression to cirrhosis in STELLAR 3	18 - 70 years old. Histologic diagnosis of NASH (NAFLD activity score [NAS] of ≥ with at least each of grade 1 steatosis, hepatocellular ballooning, and lobular inflammation). Alanine aminotransferase levels no more than 8 times the upper limit of normal, creatinine clearance as estimated by the Cockcroft-Gault equation of at least 30 ml/min, hemoglobin A1c (HbA1c) of no more than 9.5%, a platelet count of at least 100,000 per µl, and an international normalized ratio (INR) of no more than 1.4. A Model for end-stage liver disease score ≤12	Bridging fibrosis (F3 fibrosis according to the NASH Clinical Research Network [CRN] classification) Child-Pugh score ≤6	802	619
GS-US-384-1944/STELLAR-4 (Selonsertib)			Compensated cirrhosis (F4 fibrosis). Child-Pugh score ≤7	877	693

**Patients excluded from the analysis had some data missing, and the ANTICIPATE-NASH was not able to be calculated.*
Patients from the simtuzumab trials were recruited between 2013 and 2014 and STELLAR patients between 2017 and 2018.

Table 2- Comparison of length of follow-up and rate of LRE in patients from the simtuzumab and STELLAR trials who could be included in the present study or not.

	Simtuzumab trials N=474		**Stellar trials N=1679**	
	Included N=84	Excluded N=390	Included N=1312	Excluded N=367
Follow-up (months)	27.5	29	15.8	15.7
Clinical endpoint	6 (7.1%)	42 (10.7%)	27 (2%)	5 (1.3%)

Differences between included and excluded patients are not significant

Table 3- Cohort 1 baseline characteristics.

	F3 fibrosis (N=324)	**F4 fibrosis (N=375)**	**p-value***
Male, n (%)	175 (54)	190 (51)	0.377
Age, years	57.6 (9.9)	60.17 (9.8)	0.001
BMI, kg/m^2	31.5 (5.1)	31.8 (5.4)	0.427
Arterial hypertension, n (%)	209 (64.5)	276 (73.6)	0.009
Diabetes Mellitus, n (%)	187 (57.7)	247 (65.9)	0.027
Dyslipidemia, n (%)	210 (64.8%)	220 (58.5%)	0.095
Cardiovascular events during the follow-up, n (%)	13 (4.0)	19 (5.1)	0.50
Neoplasia (excluding hepatocelular carcinoma), n (%)	14 (4.3)	27 (7.2)	0.106
Follow-up, months	46.55 (35.1)	54.08 (35.3)	0.005
Liver stiffness, kPa	15.8 (8.4)	22.7 (13.3)	<0.001
Platelet count, $x10^9/L$	211 (69)	192 (75)	0.001
INR	1.04 (0.3)	1.07 (0.2)	0.195
Bilirubin, mg/dL	0.74 (0.4)	0.75 (0.4)	0.829
Albumin, g/dL	4.3 (0.4)	4.2 (0.4)	<0.001
Liver-related event, n (%)	5 (1.5)	51 (13.5)	<0.001
Hepatocellular carcinoma, n (%)	2 (0.6)	14 (3.7)	0.006
Death, n (%)	9 (2.8)	36 (9.6)	<0.001
Liver-related death, n (%)	3 (0.9)	19 (5.1)	0.002
Risk of CPSH, %	12 (5.2-22.7)	26.2 (9.6-52.3)	<0.001
Risk of LRE, %	0.8 (0.5-1.4)	1.7 (0.7-6.4)	<0.001

BMI: Body mass index; INR: International normalized ratio. Liver-related event: hepatic decompensation, hepatocellular carcinoma, transplantation and liver-related death.
Variables are expressed as mean (standard deviation) or median (P_{25}-P_{75}).
*Comparisons between groups for continuous variables were performed using Student's T test and for categorical variables using chi-square test or Fisher's test, when appropriate.

Table 4- Comparison of different parameters between F3 patients from cohort 1 who behave as F4 patients with F4 patients (upper table) and F4 patients who behave as F3 patients with F3 patients (lower table). With the data available, we show that F3 patients with ANTICIPATE-NASH >25% are more similar to F4 patients than F3 patients with ANTICIPATE-NASH <25% and that F4 patients with ANTICIPATE-NASH <12% resemble more F3 patients than F4 patients with ANTICIPATE-NASH >12%.

	F3 CSPH≥25% (n=68)	**F4 CSPH≥12% (n=259)**	**P value**
Age, years	59.8 (10.1)	60.2 (10.4)	0.748
Male, n (%)	35 (51.5)	134 (51.7)	0.969
Body mass index, kg/m^2	31.7 (5.5)	31.7 (5.5)	0.948
Liver stiffness, kPa	22.6 (10.3)	26.2 (13.3)	0.038
Platelet count, x10^9/L	142.4 (44.2)	163.8 (53.2)	0.001
Albumin, g/dL	4.2 (0.4)	4.2 (0.4)	0.600
Bilirubin, mg/dL	0.9 (0.5)	0.8 (0.5)	0.257
INR	1.1 (0.3)	1.1 (0.2)	0.757

Variables expressed as mean (standard deviation)

	F3 CSPH<25% (n=237)	**F4 CSPH<12% (n=107)**	**P value**
Age, years	57.1 (9.7)	59.6 (57.8)	0.024
Male, n (%)	129 (54.4)	50 (46.7)	0.187
Body mass index, kg/m^2	31.5 (5.0)	32.5 (5.3)	0.093
Liver stiffness, kPa	13.3 (3.9)	13.1 (3.0)	0.711
Platelet count, x10^9/L	231.9 (62.3)	260.9 (76.7)	<0.001
Albumin, g/dL	4.4 (0.4)	4.3 (0.4)	0.029
Bilirubin, mg/dL	0.71 (0.3)	0.6 (0.4)	0.081
INR	1.0 (0.3)	1.0 (0.1)	0.971

Variables expressed as mean (standard deviation)

Table 5- Theoretical projection of how many patients would be needed for a clinical trial with 2 arms using a therapy with a hypothetical effect of 30% decrease in LRE depending on the selection criteria used (alpha 0.05, power 0.8)

Selection criteria	Observed LRE	Expected LRE	Sample size per arm
Liver biopsy F4	51/375 (13.5%)	11.5%	4300
ANTICIPATE>25%	52/263 (20%)	14%	615
ANTICIPATE>50%	36/120 (30%)	21%	365

Observed LRE in cohort 1

Figures

Figure 1- Flowchart of patients from the cohort 1. Patients from Barcelona all had a liver biopsy.

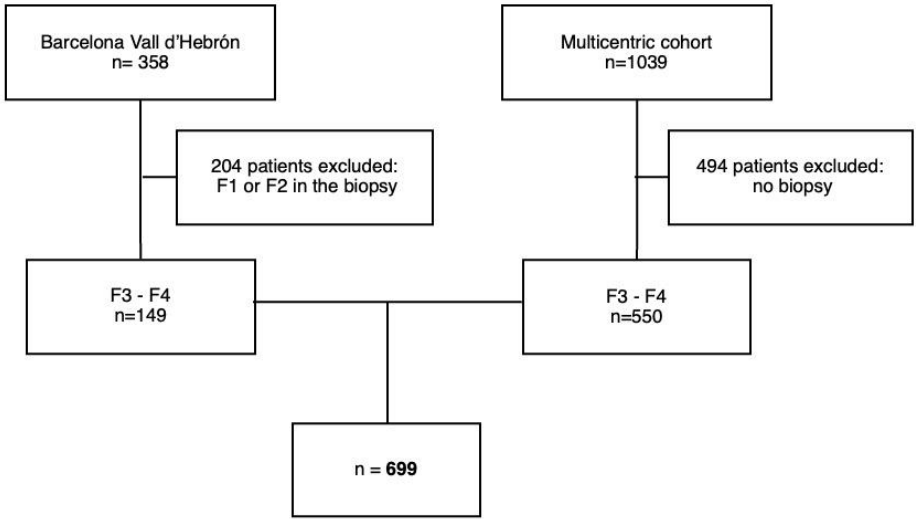


Figure 2- Calibration of the ANTICIPATE-NASH-LRE model in cohort 1. The red dashed line represents the perfect calibration (identity between predicted and observed probabilities). The black solid line represents the smooth calibration curve of the observed liver-related event rates at 3 years of follow-up against the predicted risks. Calibration was assessed by the integrated calibration index (ICI) (mean absolute difference between smoothed observed proportions and predicted probabilities) and b) the E50 and E90 (median and 90th percentile absolute difference between observed and predicted probabilities of the outcome). Results were excellent, showing ICI 0.010, E50 0.009 and E90 0.016.

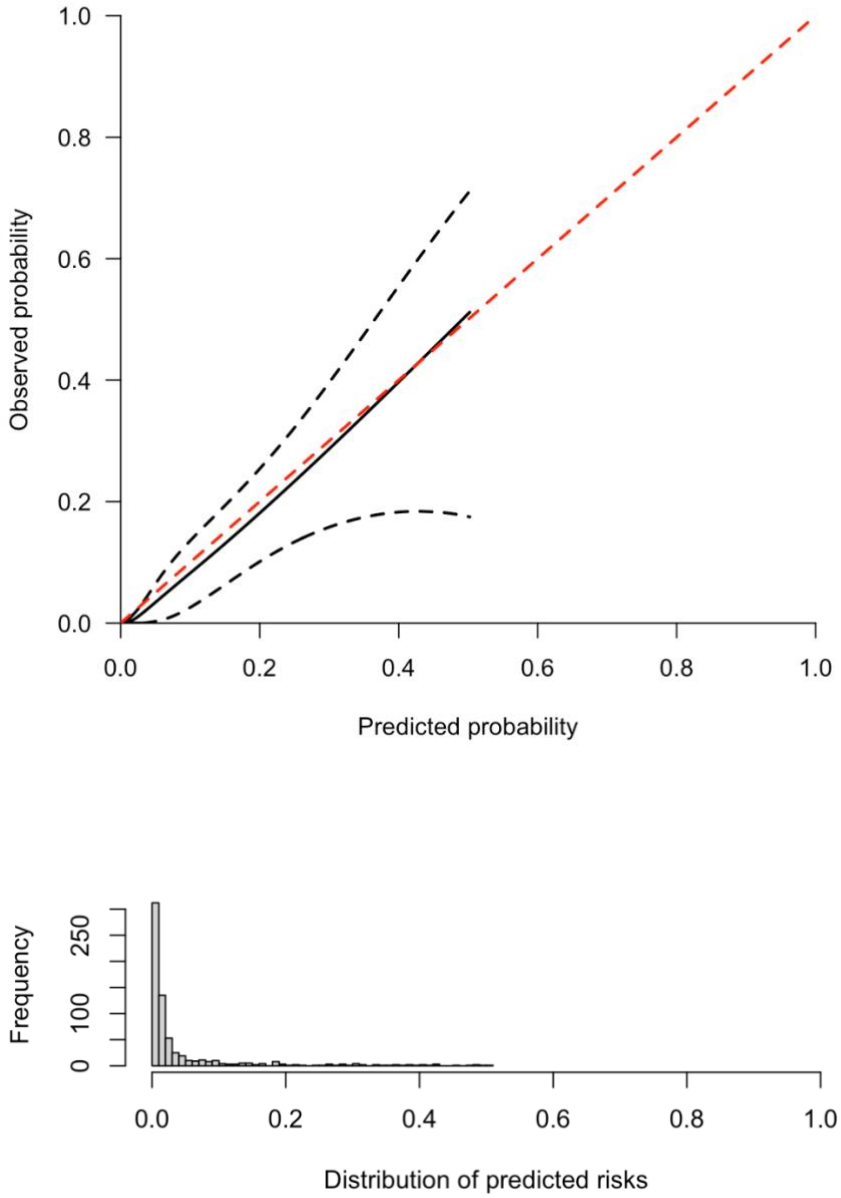


Figure 3- Subgroup analysis of the discrimination of the ANTICIPATE-NASH-LRE model for predicting liver-related events (LRE) at 3 years of follow-up in patients with BMI ≥35 kg/m2 from cohort 1. In cohort 1, 163 of the 699 patients presented a BMI ≥35 kg/m2 and presented only 4 LRE during a 3-year follow-up. The performance of the ANTICIPATE-NASH-LRE was very good, with a c-statistic of 0.994. Please, see below the graphic representation of the ROC curve.

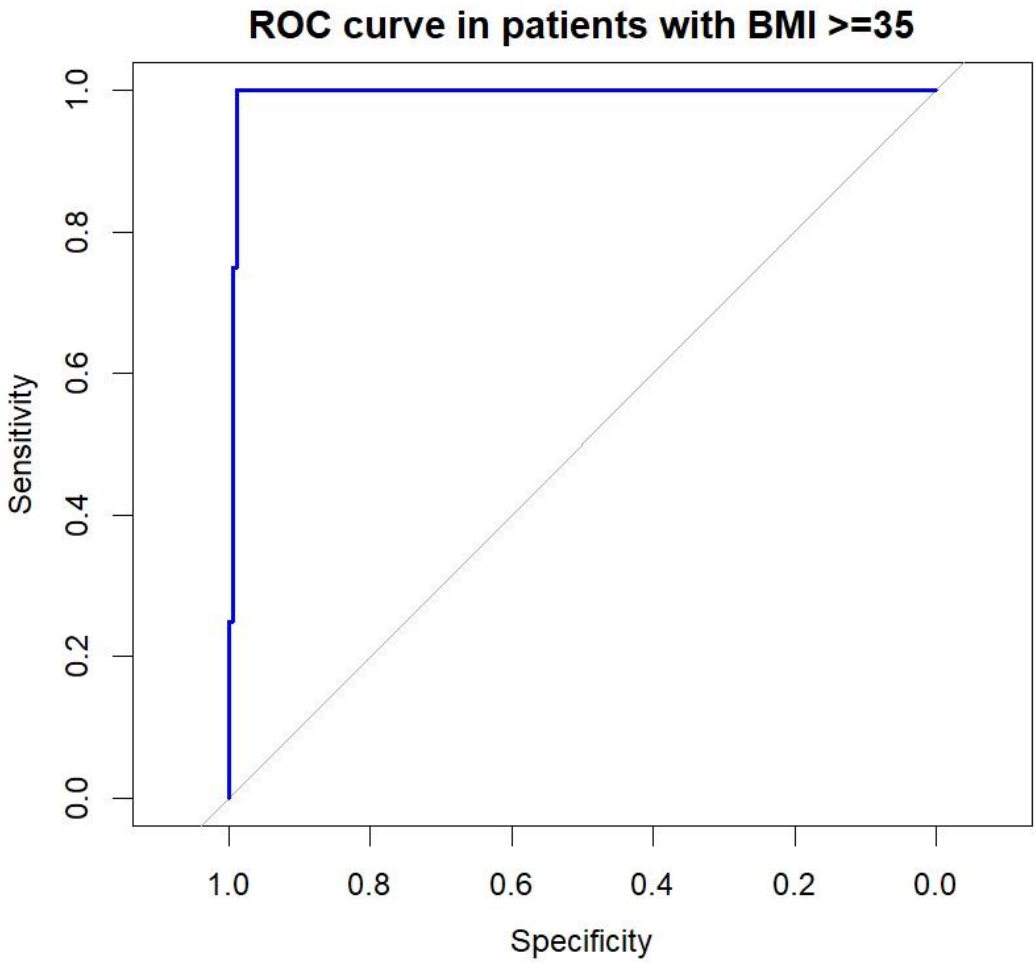
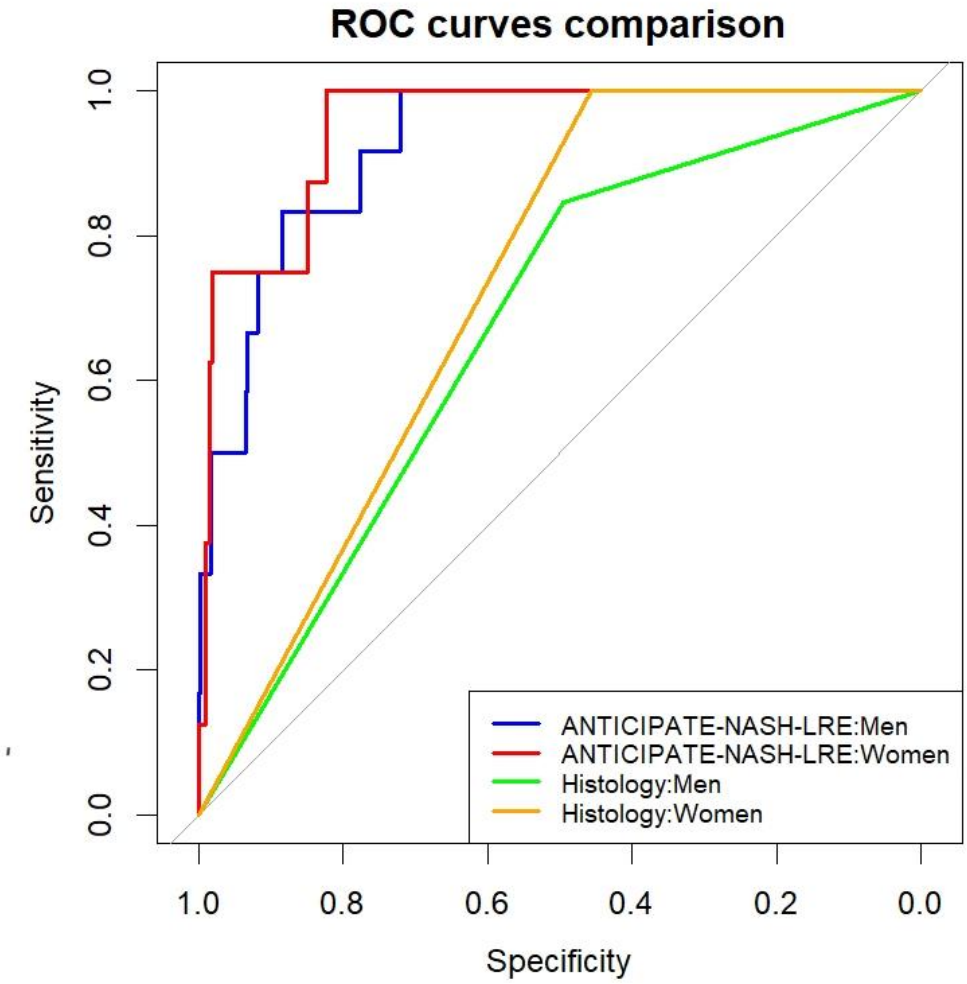


Figure 4- Discrimination analysis for the ANTICIPATE-NASH-LRE model and liver biopsy for predicting LRE at 3 years of follow-up in cohort 1, separating male from female patients. The results show that there are no differences in the performance of the model between men and women. Also, the superiority of the model to liver biopsy is maintained.



Comparison of AUC of the model according to sex (men/women)

		ROC	-Asymptotic Normal--		
Sex	**Obs**	**Area**	**Std. Err.**	**[95% Conf.**	**Interval]**
Men	348	0.9268	0.0279	0.87211	0.98156
Women	323	0.9500	0.0263	0.89841	1.00000

chi2(1) = 0.36 Prob>chi2 = **0.5461**

Comparison of AUC of histology according to sex (men/women)

		ROC	-Asymptotic Normal--	
Sex	**Obs**	**Area**	**Std. Err.**	**[95% Conf. Interval]**
Men	365	0.6702	0.0538	0.56487 0.77560
Women	334	0.7277	0.0138	0.70058 0.75481

chi2(1) = 1.07 Prob>chi2 = **0.3006**

Comparison of AUC of the model (ANTICIPATE-NASH-LRE) and histology in men

	ROC	-Asymptotic Normal--	
Obs	**Area**	**Std. Err.**	**[95% Conf. Interval]**
Model	348	0.9268	0.0279	0.87211 0.98156
Histology	348	0.6577	0.0578	0.54442 0.77106

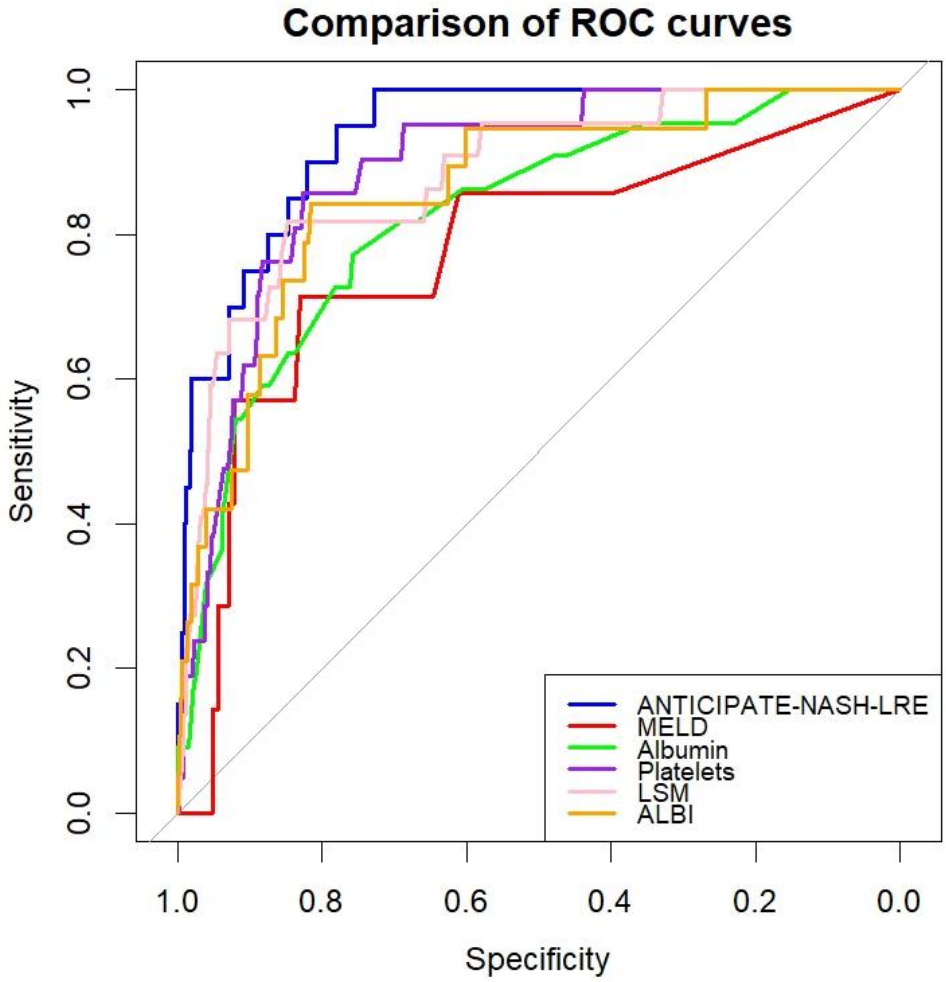
chi2(1) = 17.42 Prob>chi2 = **0.0000**

Comparison of AUC of the model (ANTICIPATE-NASH-LRE) and histology in women.

	ROC	-Asymptotic Normal--	
Obs	**Area**	**Std. Err.**	**[95% Conf. Interval]**
Model	323	0.9500	0.0263	0.89841 1.00000
Histology	323	0.7238	0.0140	0.69631 0.75131

chi2(1) = 60.18 Prob>chi2 = **0.0000**

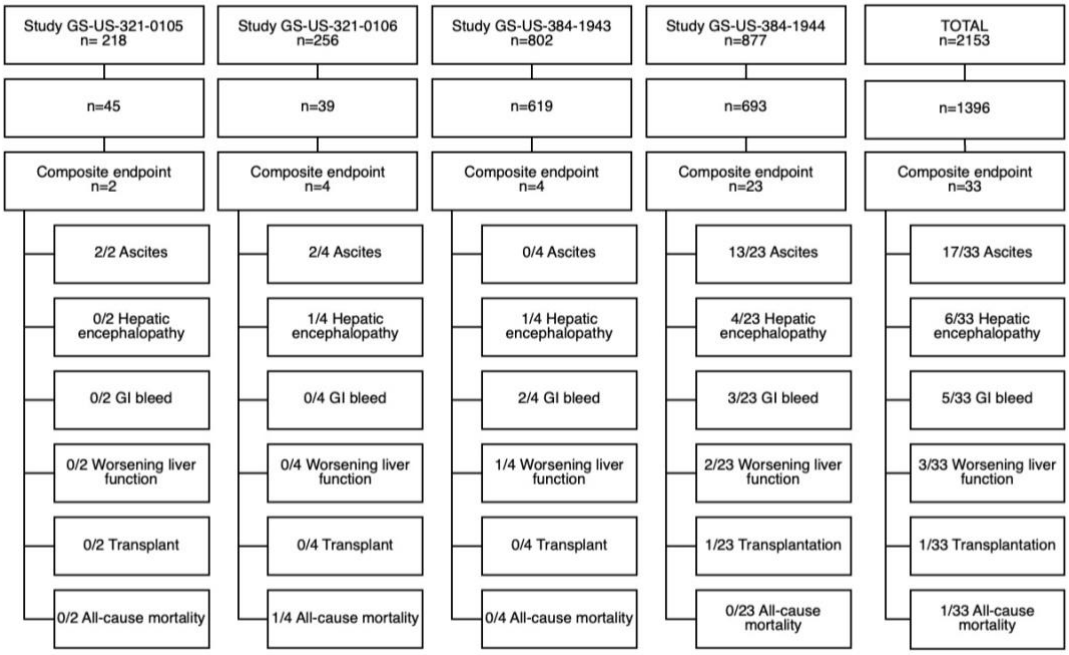
Figure 5- Discrimination analysis of the performance of the ANTICIPATE-NASH-LRE model for predicting LRE at 3 years of follow-up in patients from cohort 1 compared to other non-invasive parameters. Unfortunately, we did not have all the parameters to calculate MELD score in the 699 patients from cohort 1. We are showing below the performance of baseline MELD score compared to the ANTICIPATE-NASH-LRE score in the 149 patients from Barcelona, and albumin, LSM alone, platelet count and ALBI score in the 699 from the whole cohort 1. ANTICIPATE-NASH-LRE AUC is significantly better than all other measures.



	AUC	CI 95%
ANTICIPATE-NASH-LRE	0.94	0.89-0.98
MELD	0.77	0.56-0.98
Platelet count	0.88	0.81-0.95
LSM	0.88	0.79-0.97
Albumin	0.79	0.68-0.90
ALBI score	0.85	0.76-0.95

AUC= area under the curve, CI= confidence interval, LSM= Liver stiffness measurement, ALBI= Albumin-Bilirubin score.

Figure 6 – Flowchart and clinical events (composite endpoint) for validation cohort 2.



The first row indicates the total number of patients included in the 4 original clinical trials. The second row denotes the number of patients in which the ANTICIPATE-NASH could be calculated and were included in the study. The composite endpoint includes hepatic decompensation (ascites, portal hypertensive bleeding and hepatic encephalopathy), liver transplantation, worsening liver function with qualification for transplantation (MELD ≥15), or all-cause mortality.