



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Comparison of screening strategies with two new tests to score and diagnose varices needing treatment

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Paul Cales, F.R. (2022). Comparison of screening strategies with two new tests to score and diagnose varices needing treatment. CLINICS AND RESEARCH IN HEPATOLOGY AND GASTROENTEROLOGY, 46(7), 1-13 [10.1016/j.clinre.2022.101925].

Availability:

This version is available at: <https://hdl.handle.net/11585/894691> since: 2023-05-26

Published:

DOI: <http://doi.org/10.1016/j.clinre.2022.101925>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Calès P, Ravaioli F, Berger A, Farcau O, Festi D, Stefanescu H, Vitellius C, Nahon P, Bureau C, Ganne-Carriè N, Berzigotti A, de Ledinghen V, Petta S.

Comparison of screening strategies with two new tests to score and diagnose varices needing treatment.

Clin Res Hepatol Gastroenterol. 2022 Aug-Sep; 46(7): 101925.

The final published version is available online at: [10.1016/j.clinre.2022.101925](https://doi.org/10.1016/j.clinre.2022.101925)

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

TITLE PAGE

**Comparison of screening strategies with two new tests to score and
diagnose varices needing treatment**

Short Title: Strategies for varice screening

Paul Calès,¹ Federico Ravaioli,² Arthur Berger,¹ Oana Farcau,³ Davide Festi,² Horia
Stefanescu,³ Carole Vitellius,¹ Pierre Nahon,⁴ Christophe Bureau,⁵ Nathalie Ganne-Carriè,⁴
Annalisa Berzigotti,⁶ Victor de Ledinghen,⁷ Salvatore Petta⁸

Authors' institutions:

¹ Hepato-Gastroenterology Department, University Hospital, Angers, France; HIFIH
laboratory, UPRES 3859, UNIV Angers, France

² Gastroenterology and Hepatology Unit, Department of Medical and Surgical Sciences
(DIMEC), University of Bologna, Bologna, Italy

³ Liver Unit, Regional Institute of Gastroenterology and Hepatology, Cluj-Napoca, Romania

⁴ Hepatology Department, Avicenne Hospital, Bobigny, Paris hospitals; INSERM UMR-1162,
Paris Sorbonne University, France

⁵ Hepato-Gastroenterology Department, Purpan University Hospital, Toulouse, France

⁶ Hepatology, Swiss Liver Center, Visceral Surgery and Medicine Clinic (UVCN), Inselspital,
University of Bern, Switzerland

⁷ Hepatology Department, Haut-Lévêque Hospital, Bordeaux, France

⁸ Department of Gastroenterology and Hepatology, PROMISE, University of Palermo,
Palermo, Italy

Funding Support : institutional

Abbreviations:

ALD: alcohol related liver disease

AUROC: area under the receiver operating characteristic

cACLD: compensated advanced chronic liver disease

CI: confidence interval

CLD: chronic liver disease

EV: oesophageal varice

INR: international normalised ratio

LR-: negative likelihood ratio

LSM: liver stiffness measurement

MELD: model for end-stage liver disease

NAFLD: non-alcoholic fatty liver disease

NSBB: non-selective beta-blocker

NPV: negative predictive value

PHT: portal hypertension

PPV: positive predictive value

VANT: VNT virus alcohol NAFLD test

VARS: varice risk score

VCTE: vibration-controlled transient elastography

VNT: varices needing treatment

Correspondence:

Paul Calès, Hepatology, CHU, 49933 Angers Cedex 09, France. Tel: (33) 2 41 35 34 10, Fax:
(33) 2 41 35 41 19, E-mail: paul.cales@univ-angers.fr

<http://orcid.org/0000-0003-4866-5274>

Author contributions:

Paul Calès: study concept and design, analysis and interpretation of data, drafting of the manuscript, data collection, statistics, guarantor of the article

Federico Ravaioli: study supervision, critical revision of the manuscript, data collection

Arthur Berger: study supervision, critical revision of the manuscript, data collection, verification of statistics

Oana Farcau: critical revision of the manuscript, data collection

Davide Festi: study supervision, critical revision of the manuscript, data collection

Horia Stefanescu: critical revision of the manuscript, data collection

Carole Vitellius: critical revision of the manuscript, data collection, literature review

Pierre Nahon: critical revision of the manuscript, data collection

Christophe Bureau: critical revision of the manuscript, data collection

Nathalie Ganne-Carriè: critical revision of the manuscript, data collection

Annalisa Berzigotti: critical revision of the manuscript, data collection

Victor de Ledinghen: critical revision of the manuscript, data collection

Salvatore Petta: critical revision of the manuscript, data collection

VEB6 group: data collection

All authors approved the final version of the article, including the authorship list.

Item count:

Main text: 4390 words (4,000 words maximum, including figure and table legends and main text), number of figures and tables: 4 + 4, number of references: 31 (≤ 100), number of supplementary files for online publication: 1 including 16 tables and 5 figures. Title: 102 characters (< 130 characters including spaces). Short title: 31 characters.

Declaration of Conflict of Interest: none related to the topic

Authors' e-mails:

paul.cales@univ-angers.fr

f.ravaioli@unibo.it

arthur.berger@chu-bordeaux.fr

oana.farcu@gmail.com

davide.festi@unibo.it

horia.stefanescu@irgh.ro

Carole.Vitellius@chu-angers.fr

pierre.nahon@aphp.fr

bureau.c@chu-toulouse.fr

nathalie.ganne@aphp.fr

Annalisa.Berzigotti@insel.ch

victor.deledinghen@chu-bordeaux.fr

salvatore.petta@unipa.it

Writing assistance:

Kevin L. Erwin for English proofreading (institutional support)

Highlights

- Three screening strategies, based on two new tests to rule out or in (with 100% specificity) varices needing treatment, were developed in the main liver disease etiologies irrespective of liver severity.
- VANT test spared 40% of endoscopies in population screening based on 95% sensitivity.
- The sparing rate of VARS test was 62% in individual screening based on 95% negative predictive value, and 12% in 100% safe screening based on 100% sensitivity and specificity.

ABSTRACT

Background and Aims: We aimed to improve non-invasive screening of varices needing treatment (VNT) and compare different screening strategies. **Methods:** 2,290 patients with chronic liver disease were included in a retrospective study. Etiologies were: virus: 50.0%, NAFLD: 29.5%, alcohol: 20.5%, VNT: 14.9%. Test descriptors were performance (spared endoscopy) and safety (missed VNT). VNT tests were evaluated according to their safety levels either for individual screening (95% negative predictive value (NPV)), population screening (95% sensitivity) or undifferentiated screening (100% sensitivity/NPV) without missed VNT. The tests provided three categories: missed VNT <5%, VNT 100% specificity (new category), both sparing endoscopies, and intermediate (endoscopy required). **Results:**

Independent VNT predictors (etiology, sex, age, platelets, prothrombin index, albumin, ALT, liver stiffness) were included in two tests: VNT virus alcohol NAFLD test (VANT) and varice risk score (VARS). We report results of the whole population. Considering population screening, performances were, Baveno VI criteria: 24.1%, Anticipate: 24.7%, VariScreen: 35.3%, VANT: 40.2% ($p < 0.001$ vs other tests). VANT spared 58.0% more endoscopies in the whole population than Baveno criteria in compensated advanced chronic liver diseases. Considering individual screening, VARS performance was, in all patients: 62.0% vs 42.9% for the expanded Baveno VI criteria ($p < 0.001$), and, in NAFLD: 72.8% vs 65.1% for the NAFLD cirrhosis criteria ($p < 0.001$). Considering undifferentiated screening, VARS performance was 12%. The VARS score estimated VNT probability from 0 to 100% (AUROC: 0.826). **Conclusion:** VANT and VARS spared from 12% (undifferentiated screening) to 40% (population screening) or 62% (individual screening) of endoscopies in main-etiology patients without ascites.

Words: 250 (≤ 250)

Keywords: Portal hypertension; oesophageal varices; non-invasive test; elastometry; screening

INTRODUCTION

The Baveno VI criteria, based on platelets and liver stiffness measurement (LSM), were aimed at ruling out varices needing treatment (VNT) and avoiding unnecessary endoscopies [1]. They enabled the wide clinical acceptance of non-invasive tests [2]. However, they do have some limitations. First, their spared endoscopy rate has been shown to be only about 20% [3]. Second, the safety criterion of non-invasive VNT screening is fixed at a missed VNT rate of <5%. However, missed VNT may be defined in two different ways (primarily), which makes the interpretation of the tests confusing [4]. One of those definitions is based on the probability of missing <5% VNT in patients who do not undergo endoscopy. This is a negative predictive value (NPV) adapted to an individual screening strategy. The other is based on the probability of missing <5% VNT in patients with VNT. This corresponds to VNT sensitivity and thus is adapted to a population screening strategy. Screening categories were not explicitly stated for previous tests and no study has adequately compared the two strategies. This has resulted in a confusing situation for test comparisons [5]. Third, the Baveno VI criteria were originally applied to compensated advanced chronic liver disease (cACLD). However, we have recently shown that this strategy restricted to cACLD spared less endoscopies than a global strategy applied to a larger spectrum of liver severity [6]. Moreover, the lower limit of cACLD (LSM ≥ 10 kPa) penalises the safety criterion since VNT are encountered below that cut-off and thus more than 5% of VNT can be missed [6]. Concomitantly, decreasing the LSM cut-off of cACLD below 10 kPa was recently suggested [7, 8]. Also, decreasing the lower limit for cACLD introduces other advantages. Thus, thanks to that change, we can more confidently analyse two new issues. First, is it possible and worthwhile to determine a test cut-off for 100% sensitivity for VNT? Second, a lesser degree of liver dysfunction in the 5% of missed VNT would legitimise this accepted missed VNT

risk in the Baveno VI statements. Indeed, variceal rupture risk and mortality increase as a function of liver dysfunction [9]. The change of the cACLD upper limit (complications) to LSM availability (i.e. ascites absent) enabled more endoscopies to be spared [6]. Moreover, we introduce here a new paradigm by hypothesising that if tests could reach 100% specificity for VNT, that too would rule out the need for endoscopy. Moreover, a test combining 100% specificity and 100% sensitivity would confer 100% safety, which is an alternative screening strategy worthy of exploring. Indeed, the test cut-offs for 100% specificity and sensitivity correspond to those of predictive values. Therefore, this 100% safety strategy would encompass individual and population screenings and thus prevent any misuse of tests at the individual level. Fourth, the Baveno VI criteria were originally not suitable for non-alcoholic fatty liver disease (NAFLD). Their cut-offs were finally modified [10] resulting in the *NAFLD cirrhosis criteria*, but these latter were only designed for individual screening. Fifth, the Baveno VI criteria and their derivatives do not estimate VNT probability. Finally, Baveno VII statements do not delete the interest of VNT tests (see discussion) [11].

Thus, our primary objective was to develop tests that improve the rate of spared endoscopy. Our main secondary objective was to compare different screening strategies according either to their epidemiological characteristics (screenings of individuals, populations, or both) or to their clinical spectrum, i.e. restricted or not to cACLD. Other secondary objectives were to quantify VNT probability and refine safety criteria.

PATIENTS AND METHODS

Patients

The clinical information of patients with CLD was collected from a number of centres participating in several studies wherein VNT was usually the main outcome and vibration-controlled transient elastography (VCTE) the main measurement. The protocol conformed to the Declaration of Helsinki and received approval from the ethics review boards of all participating centres. All study participants gave informed consent. Patients included in previously recorded CLD subpopulations of any main etiology (alcohol related CLD (ALD), NAFLD, hepatitis B or C virus) were eligible for inclusion if they had undergone an endoscopy to determine oesophageal varice (EV) size. The four minimum inclusion criteria were a platelet count, successful LSM by VCTE (using the M probe), known EV stage and a maximum delay of six months between endoscopy and LSM or platelets. The exclusion criteria were ascites, gastro-intestinal bleeding and treatment of portal hypertension (PHT) (TIPS, band ligation or sclerotherapy of EV, and non-selective beta-blockers (NSBBs)) and incomplete data. Also, patients were included irrespective of LSM values and liver severity (i.e. non limited to cACLD) to enable a less biased analysis of the VNT subset. The 2290 included patients (flowchart in Figure S1 in the Supplemental Material) were randomised in derivation (2/3) and validation (1/3) populations with stratification on VNT and etiology.

Methods

Data collection

Clinical data - The main clinical data were age, sex, height, body mass index, and CLD etiology. The main laboratory data were liver function tests, blood cell count and serum creatinine (measured in each centre). The model for end-stage liver disease (MELD) score included bilirubin, the international normalised ratio (INR) and creatinine [12].

Endoscopy - A standard endoscopy was performed by experienced operators, and EV grades were recorded.

LSM - All LSMs were performed by experienced operators using VCTE, specifically M probe-equipped Fibroscan devices (Echosens, Paris, France). Patients with LSM using the XL probe only were not included because they were not sufficiently numerous to be separately evaluated (Figure S1). Technical characteristics are detailed elsewhere [13].

Definitions

Study

This was a post-hoc analysis (retrospective study) of prospectively collected data and a TRIPOD 2a study [14].

Outcome

The main outcome was VNT, defined as large EV (a diameter ≥ 5 mm [15]).

VNT estimators

A diagnostic estimator was called a *score* when it provided a numerical variable quantifying precisely the VNT probability (from 0 to 1). An estimator was called a *test* when it was categorised by cut-off(s), resulting in a qualitative variable indicating VNT categories (ruled out, indeterminate, ruled in).

Outcome measurements

Tests - Performance was the spared endoscopy rate and *safety* was reflected by the missed VNT rate by tests. The spared endoscopy rate was calculated as the ratio between the number of patients with missed VNT by the test and the total number of patients. Safety was the ratio missed VNT/reference set. For population screening, the reference set was patients with VNT; thus, safety corresponded to VNT sensitivity. For individual screening, the reference set was patients spared of endoscopies; thus, safety corresponded to VNT NPV [4]. The rate of *unnecessary endoscopies* was calculated as the number of endoscopies needed to diagnose one patient with VNT. It expresses the number of times endoscopies were unnecessary. For example, in a group of five patients where endoscopy found VNT in one, the rate would be five. The negative likelihood ratio (LR-) was another outcome measurement (details in the Supplemental Material).

Scores - The scores estimating VNT probability were mainly evaluated by their calibration and discrimination. Calibration was measured by the correlation between score percentiles and VNT prevalence. VNT discrimination was measured by the area under the receiver operating characteristic (AUROC).

Safety criteria

Classical strategies - We evaluated three criteria. First, the classical *quantitative* criterion is a missed VNT rate <5% according to Baveno VI statements [1]. Second, we evaluated the *qualitative* safety by the level of liver dysfunction as a function of VNT status. The principle of *qualitative* safety is to, among tests with similar performance, privilege the one having the lowest liver dysfunction in missed VNT and discard (or limit) any test inducing missed VNT in severe CLD. Indeed, the incidence and mortality of variceal bleeding grows with liver

severity [9]. Thus, *secureness* describes here the absence of missed VNT in CLD with poor liver function (MELD score ≥ 10 or INR ≥ 1.24) called *severe* CLD hereafter [6]. The level of liver dysfunction in missed VNT was called *functional* safety (new criterion).

100% safety - The aim was to design a test with neither missed VNT nor false VNT diagnosis thanks to cut-offs for 100% sensitivity (=100% NPV) and 100% specificity (=100% positive predictive value (PPV)), respectively. Therefore, this strategy was called *undifferentiated*.

VNT ruled in - This new category was made possible by high estimator accuracy. The 100% specificity cut-off was chosen for all strategies since this should be considered a firm VNT diagnosis due to the therapeutic consequences.

Comparators

The new tests were compared to published tests. These were, for population screening:

Baveno VI criteria [1], Anticipate [16], PLER and PLEASE included in the VariScreen algorithm [6]; and, for individual screening: expanded Baveno VI criteria [17] and NAFLD cirrhosis criteria [10].

VNT screening strategies

The main characteristics of the three strategies (population, individual, undifferentiated) are summarised in Table 1.

Estimator construction in the derivation set

Overview - Details on the predefined strategy resulting from our previous studies [4, 6] and estimator construction are provided in the Complementary Data. Briefly, we first determined the eight independent VNT markers included in the two estimators by logistic regression: etiology, sex, age, platelets, prothrombin index, liver stiffness, albumin and ALT. Then, we constructed a logit score (range: 0 to 1) estimating the VNT probability called *varice risk score* (VARS). We also developed VARS as a test to spare endoscopy since this score could be easily categorised by cut-offs (contrary to the next test). Thus, the VARS score was categorised for the three screening strategies using different cut-offs. Thereafter, we constructed a qualitative binary (yes/no) algorithm called *VNT virus alcohol NAFLD test* (VANT) to spare endoscopies in population screening. Although it included the same markers as VARS, VANT was expected to provide better performance since its target was spared endoscopies (and not VNT). Finally, the VANT and VARS tests included three categories: VNT ruled out, indeterminate VNT status and VNT ruled in. The respective use of VARS et VANT as a function of screening strategies is summarised in Table S1.

Design - VARS was the combination of six logit scores, each designed for the subsets sex per etiologies and the target being VNT. So, VARS was a composite score estimating VNT probability. The VNT ruled out category for VANT resulted from the selection of the two logit scores providing the largest subset with 95% sensitivity for VNT in each etiology. The VNT ruled in category for VANT resulted from the selection of the two logit scores providing the largest subset with 100% specificity for VNT (one of them being VARS). So, VANT was targeted for spared endoscopies. Formulae are detailed in the Supplemental Material.

Statistics

Quantitative variables were expressed as mean \pm standard deviation and compared using the Student test or ANOVA. Qualitative variables were expressed as proportions and compared using the Chi² or Fisher test when unpaired and the Cochran or McNemar test when paired. Paired AUROCs were compared by the Delong test. 95% confidence intervals (CI) of test performance and safety were determined by bootstrap on 1000 samples stratified on sex and etiology. Independent VNT predictors were determined by binary logistic regression; variables responsible for collinearity ($r > 0.8$) were always excluded. Data were reported according to Liver FibroSTARD statements [18]; however, classical contingency tables were not reported since the main outcome measurement was not a binary diagnostic test for VNT. Data were analysed on a partial intention-to-diagnose basis. Thus, all patients were included irrespective of LSM reliability criteria [19] (except in one NAFLD subpopulation [10]) but patients with unsuccessful examinations (LSM and endoscopy) were not included. The main statistical analyses were performed using SPSS version 18.0 (IBM, Armonk, NY, USA).

RESULTS

Patient characteristics

Because there were no significant differences between the derivation and validation populations (Table S2 in the Supplemental Material), patient characteristics are described here in the whole population (2290 patients). Viral CLD was the most frequent etiology (50%); other etiologies included NAFLD (29.5%) or ALD (20.5%). The VNT prevalence was 14.9%. Also, 93% of patients and 98.5% of VNT were observed in $LSM \geq 10$ kPa; 7.6% of VNT were observed in $LSM < 15$ kPa. Patients with VNT were more frequently male or affected by ALD and had more severe CLD (Table S3). The prevalence of cACLD was calculated at 43.8%. Most patient characteristics were significantly different as a function of etiology (Table S4).

Non-invasive diagnosis

Scores estimating VNT probability

In the derivation population, the AUROC of VARS for VNT (0.832) was higher ($p < 0.001$) than those of published scores: PLER (0.758), Anticipate (0.767) and PLEASE (0.797) (Table S5). In the validation population, the VARS AUROC (0.816) remained significantly higher than that of other scores ($p \leq 0.027$) and was not significantly different from that of the derivation population ($p = 0.512$). Therefore, the following characteristics are described in the whole population. VARS was able to reach 100% sensitivity in one patient subset (11.4%), but it also reached 100% specificity in another (0.6%) (Figure 1A). VARS was well calibrated for VNT probability with $r_s = 0.953$ and $r_p = 0.981$ (Figure 1B). The PPV for VNT was well

calibrated since it linearly increased ($r_p=0.996$) from 14.9% to 100% as a function of VARS score (Figure 2A).

Tests to spare endoscopy

New tests

We evaluated the VANT test and the VARS test according to screening strategies (Table 2). Considering population screening, the performances in the derivation population were, VANT: 39.6% vs VARS: 31.2% ($p<0.001$). Considering individual screening, VARS performance was 62.1% in the derivation population, which was significantly higher ($p<0.001$) than the previous performance of VANT or VARS categorised for population screening. The missed VNT rate was $<5\%$ for all tests in both strategies in the derivation population, by construction, but also in the validation population (Table 2). Furthermore, considering undifferentiated screening (including 100% sensitivity), VARS performance dropped to 11.8% in the derivation population, which was significantly lower ($p<0.001$) than the previous performance of VARS or VANT in other strategies. Finally, all test performances were not significantly different between the derivation and validation populations. Therefore, the following results are presented in the whole population. Thus, the performance of VANT was 40.2% vs 32.0% ($p<0.001$) for VARS in population screening (Table 2). The patient set with VNT ruled in by VANT included 5.3% of VNT in 0.8% of patients (Figure 2A).

Comparison with published tests in the whole population

Population screening - The performances were, in increasing order (Table 3): Baveno VI criteria: 24.1%, Anticipate: 24.7%, VariScreen: 35.3%, VANT: 40.2% ($p<0.001$ for each pair

comparison except for Baveno VI criteria vs Anticipate: $p=0.393$). LR- were excellent (<0.1) in all tests except for Anticipate. All tests were safe (missed VNT $<5\%$).

Individual screening - VARS performance was 62.0% vs 42.9% ($p<0.001$) for the expanded Baveno VI criteria (Table 4). Their respective safeties were 4.98% and 4.0% , ($p=0.246$). In NAFLD, VARS performance was 72.8% vs 65.1% ($p<0.001$) for the NAFLD cirrhosis criteria. Their safeties were both 4.5% ($p=0.984$).

Screening comparison - This comparison was appropriate only for the VARS test (Table S6). Briefly, VARS performance was 62.0% in individual screening vs 32.0% in population screening ($p<0.001$) and 12.0% in undifferentiated screening ($p<0.001$). The epidemiological limit of individual screening was that five times more VNT were missed among VNT compared to population screening (20.5% vs 3.8% , respectively, $p<0.001$).

Sensitivity analysis

The influence of screening type on safety, liver severity, etiology, cACLD definition, LSM characteristics, VNT prevalence and transaminase level are detailed in the Supplemental Material. Briefly, the use of the VARS test must be restricted to MELD score <10 in individual screening to comply with the safety criteria.

Internal validation

The validation of tests was also internally evaluated by 95% CIs of performance and safety generated by bootstrapping (Tables 2, 3 and 4). In population screening, the Baveno VI criteria were the only test wherein the upper 95% CI limit of safety was $<5\%$ in the whole population. This limit was slightly $>5\%$ in other tests. However, it was also slightly $>5\%$ in at

least one etiology for all tests. In individual screening, this limit was slightly $>5\%$ for all tests in the whole population.

Clinical applications

All screenings - Tests are indicated in patients with LSM ≥ 9 (virus) or ≥ 10 (NAFLD, ALD) kPa (details in Supplemental Material) and without ascites, keeping in mind that test performance decreases as a function of liver severity.

Population screening - VANT is the main test and thus used first (Figure 2A). VARS is an optional score to directly estimate VNT probability. Its PPV is purely indicative as it depends on VNT prevalence. A free calculator is available at <https://gilles-hunault.leria-info.univ-angers.fr/wstat/vars-vant.php>.

Individual and undifferentiated screenings - VARS is used as a test to spare endoscopies (Figure 2B) and again as a score to estimate VNT probability. In individual screening of patients with MELD score ≥ 10 , VANT should be preferred since VARS was not secure. However, VANT performance was only 9.0% here. Finally, the respective application of the VANT and VARS tests is summarised in Figure 3 according to the three screening strategies and in Figure 4 for clinical practice.

DISCUSSION

Originalities - The originalities of our study were, A) a comparison of three screening strategies including one new undifferentiated (100% sensitivity) strategy, and one new category to spare endoscopies, i.e. VNT ruled in by 100% specificity (included in the three strategies); B) two new tests with improved performance; and C) qualitative safety criteria based on liver dysfunction in missed VNT.

New screening strategies - To our knowledge, an undifferentiated strategy has never been used to date. In practical terms, knowledge of 100% sensitivity and specificity is a major advantage because the cut-offs of predictive values and sensitivity/specificity are the same. Therefore, the missed VNT rate is null whatever the screening strategy. Also, we showed that the concept of unnecessary endoscopies can be extended to patients with VNT prevalence of 100% by developing tests offering cut-offs for 100% specificity. Therefore, 100% specificity was included in the three strategies developed here.

New tests - We developed two new estimators including eight commonly employed biomarkers, i.e. without additional cost in clinical practice. However, in exchange for the better performance they offer, a calculator is required but also freely available. The first estimator, the VARS score, was initially developed to estimate VNT probability. This was the only VNT score with an AUROC ≥ 0.8 (0.826), providing a discriminatory accuracy of value [2]. Its complete range of VNT probability from 0 to 100% had not been reached previously. This also allowed us to categorise the VARS score as a test for several applications. Thus, the VARS test spared 12% of endoscopies in undifferentiated screening, 32% in population screening and 62% in individual screening (or 73.5% in use restricted to MELD score < 10 to respect security, or 78.0% in cACLD).

The second estimator, the VANT test, was directly developed to spare endoscopies in population screening. This was the best performing and most secure test in this strategy. VANT had three advantages over published tests. The first was a significantly increased spared endoscopy rate which was especially marked in NAFLD and ALD. Those are the two major etiologies where the Baveno VI criteria and Anticipate had limitations in safety and performance in the present series as previously noted [5]. The second advantage of VANT, when compared to tests developed in cACLD, was its construction adapted to a larger spectrum of liver severity, resulting in the absence of missed VNT in severe CLD. Therefore, VANT was robust against the spectrum effect [20] (details in the Supplementary Material). The third was a category where VNT are ruled in.

Which estimator? - The new diagnostic estimators are used sequentially in population screening (Figure 2A). First, VANT is calculated and patients falling in its undetermined category will require endoscopy. For that group, VNT probability can be estimated by VARS score, keeping in mind that the results of VANT to rule in or out VNT prevails over those of VARS. Considering individual screening (Figure 2B), the VARS test offered the best performance but its use is restricted to MELD scores <10. Furthermore, VARS can be used optionally for its 100% safety in 12.0% of patients. These estimators should be repeated every six months, because performance decreased when there was a delay >6 months between the test and endoscopy [6].

Which strategy? - The three possible strategies offer different options to physicians and patients (Figures 3 and 4). Performance was two-fold better in individual screening vs population screening. However, three limits of individual screening, compared to population screening, must be weighed against its high performance. First, its number of patients with missed VNT was five times higher, representing the most important limit for public health. Second, those patients with missed VNT had more altered liver function which is partially

alleviated by restricting use to MELD score <10 (or INR <1.24). Third, VARS use is limited to VNT prevalences $\leq 15\%$. That however should often be the case considering a global screening strategy applied to CLD without ascites. Finally, a sensitivity $\geq 95\%$ is a classical criterion of effective screening [21]. Therefore, we privilege population screening as the first-line option [4]. Individual screening is an option restricted to certain patients, e.g. those particularly reluctant to screening. The undifferentiated strategy is an option for circumstances requiring a firm diagnosis (e.g. an anxious patient). VNT screening is a part of the primary prophylaxis of VNT rupture. The ruled in category might eliminate the need for endoscopy when there is no contra-indication to NSBBs. Otherwise, a contra-indication to NSBBs might spare one endoscopy since the diagnostic and therapeutic endoscopies are often performed at different times.

Which patients should be tested? - The biomarker that most spared the recourse to VNT tests was prothrombin index in NAFLD and ALD, and LSM in viral CLD in the present population. However, considering normal values, LSM was the most useful biomarker with a cut-off of 10 kPa in NAFLD and ALD, and 9.0 kPa in viral CLD (details in Supplemental Material). The use of a VNT test is justified in patients with LSM between 9 and 10 kPa (i.e. below the cACLD definition) since the performance there is very high and the additional endoscopy rate is only 0.5%. Moreover, ignoring these patients with LSM <10 kPa would significantly impact the safety definition in population screening [6]. The use of a VNT test is also justified in patients beyond the cACLD definition (but without ascites) since VANT performance was 15% without missed VNT in severe CLD. Finally, the comparison of endoscopies spared by VANT performed in all patients without ascites showed a gain of 30.5% vs VANT or 58.0% vs the Baveno VI criteria performed only in cACLD (details in Supplemental Material).

Limits - The first category of limits, which have been discussed in depth elsewhere [6], comprises those inherent to the population and are recalled in the Supplemental Material. The second category of limits is particular to the present study, i.e. new tests. We observed no significant inter-centre variability for performance or safety after adjustment on patient characteristics (details in Complementary Data). Our new tests were based on algorithms that are more complex than those of previous tests. Nevertheless, the multiple adjustments in algorithms, especially on sex, etiology and liver severity explained VANT's robustness against the spectrum effect [20]. Also, we think that several factors decreased the overfitting risk, including the large population size, low VNT prevalence [22] and models including multiple predictors with potentially opposite influences on VNT prediction. A significant optimism bias was discarded by reproducible results in the validation population. However, optimism bias was more precisely evaluated by 95% CIs of performance and safety obtained by bootstrapping in 1000 samples (i.e. internal validation). The narrow safety 95% CIs (VARs: 3.8-6.1%, VANT: 2.5-7.1%) indicated good robustness. Moreover, VANT had the narrowest safety 95% CI in NAFLD. However, the new tests should be validated in independent populations. Of note, our previous VariScreen algorithm, constructed in the same derivation population, was externally validated [23].

In cACLD, VARs and VANT performed better than other tests but VANT was slightly unsafe; this is due to the safety objective being targeted to the global strategy, i.e. the whole population, rather than to the cACLD subgroup [6]. However, our present results and recent data [7] show that non-invasive VNT screening should not be limited to the classical cACLD definition. The new definition included LSM cut-offs very close to those observed in the present study. Moreover, our patients did not have ascites due to the LSM requirement. Thus, LSM availability is the main limitation to the application of our strategies. There were some limits in patients with LSM <10 kPa for population screening. The Baveno VI criteria were

not suitable for these patients due to a high rate of unnecessary endoscopies. VariScreen should be preferred to VANT since it offers better safety and less unnecessary endoscopies. For several reasons, the use of VNT tests should continue even if primary prevention by NSBBs is extended to clinically significant PHT [24, 25] according to Baveno VII statements [11]. These reasons include adapting motivation for drug adherence, which is an important clinical challenge in asymptomatic patients, and managing the contraindications and side effects of those NSBBs [11, 25]. Indeed, in real life, up to 71% of cirrhotic patients might not receive NSBBs [26] and drug adherence is poor in cirrhosis [27] especially in primary prevention [28]. Moreover, the non-invasive criteria of clinically significant PHT [29] are not validated [2] especially outside the original Baveno VI definition of cACLD, which was recently redefined [7]. Also, 8% of VNT were observed in LSM <15 kPa, i.e. below the cut-off of Baveno VII criterion for prescribing NSBB. Beyond those considerations, adherence to screening by patients and primary care providers is improved by precise information [30, 31] on and negative perception of the disease [28]. With that respect, knowing precise VNT status and probability would be more convincing than an abstract diagnosis of clinically significant PHT. Finally, VNT tests will be also used by centers performing VNT banding in primary prevention.

Conclusion - Two new VNT estimators (one for individual or undifferentiated screening and VNT probability, and the other for population screening) improved and secured performance, especially in current major etiologies like NAFLD and ALD where VNT tests should be performed by LSM ≥ 10 kPa. These three strategies had respective advantages and limits, justifying population screening as the preferred option and restrictions to certain clinical settings for individual or undifferentiated screening. Their respective applications should improve adherence to VNT screening.

Acknowledgments

We thank the co-investigators of the VEB6 (validation of expanded Baveno VI criteria) group as a function of subpopulations:

Monocentre subpopulations:

- Angers (SNIFF 95 study, France): François Buisson, Arthur Berger, Carlotta Carboni (and Bologna), Frédéric Oberti, Isabelle Fouchard, Jérôme Boursier, Adrien Lannes, Gilles Hunault, Paul Calès.
- Bologna (Italy): Elton Dajti, Vanessa Luigina Alemanni, Giovanni Marasco, Federico Ravaioli, Antonio Colecchia, Davide Festi.
- Cluj (Romania): Horia Stefanescu, Oana Farcau, Anca Maniu, Andreea Ardelean, Petra Fischer, Crina Grigoras, Bogdan Procopet.
- Bern (Swiss): Annalisa Berzigotti.
- Toulouse (France): Christophe Bureau, Marie Angèle Robic, Maeva Guillaume.
- Bobigny (previously Bondy) (France): Nathalie Ganne-Carriè, Nathalie Barget.

Multicentre subpopulations:

- VO-VCO (PHRC, France): Sylvie Sacher Huvelin, Paul Calès, Dominique Valla, Christophe Bureau, Anne Olivier, Frédéric Oberti, Jérôme Boursier, Jean Paul Galmiche, Jean Pierre Vinel, Clotilde Duburque, Alain Attar, Isabelle Archambeaud, Robert Benamouzig, Marianne Gaudric, Dominique Luet, Patrice Couzigou, Lucie Planche, Emmanuel Coron, Jean-Baptiste Hiriart, Faiza Chermak, Maude Charbonnier.
- M116 study (Echosens, Italy, Romania, France, United Kingdom): Horia Stefanescu, Giovanni Marasco, Paul Calès, Mirella Fraquelli, Matteo Rosselli, Nathalie Ganne-Carriè,

Victor de Ledinghen, Federico Ravaioli, Antonio Colecchia, Corina Rusu, Pietro Andreone, Giuseppe Mazzella, Davide Festi.

- B6C NAFLD group: Salvatore Petta¹, Giada Sebastiani², Elisabetta Bugianesi³, Mauro Vigano⁴, Vincent Wai-Sun Wong⁵, Annalisa Berzigotti⁶, Anna Ludovica Fracanzani⁷, Quentin M. Anstee^{8,9}, Fabio Marra^{10,11}, Marco Barbara¹, Vincenza Calvaruso¹, Calogero Camma¹,

¹Sezione di Gastroenterologia e Epatologia, Di.Bi.M.I.S, Università di Palermo, Italy;

²Division of Gastroenterology and Hepatology, McGill University Health Centre, Montreal, QC, Canada;

³Division of Gastroenterology, Department of Medical Sciences, University of Torino, Torino, Italy;

⁴Hepatology Unit, Ospedale San Giuseppe, University of Milan, Milan, Italy;

⁵Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong;

⁶Swiss Liver Center, Hepatology, University Clinic for Visceral Surgery and Medicine, Inselspital, University of Bern, Switzerland;

⁷Department of Pathophysiology and Transplantation, Ca' Granda IRCCS Foundation, Policlinico Hospital, University of Milan, Italy;

⁸Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, United Kingdom; ⁹Liver Unit, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Freeman Hospital, Newcastle upon Tyne, United Kingdom;

¹⁰Dipartimento di Medicina Sperimentale e Clinica, University of Florence, Italy;

¹¹Research Centre DENOTHE, University of Florence, Italy;

¹²Centre d'Investigation de la Fibrose Hépatique, INSERM U1053, Hôpital Haut-Lévêque, Bordeaux University Hospital, Pessac, France

- ANRS CO12 CirVir group: Pierre Nahon¹, Patrick Marcellin², Dominique Guyader³, Stanislas Pol⁴, H el ene Fontaine⁴, Dominique Larrey⁵, Victor De L edinghen⁶, Denis Ouzan⁷, Fabien Zoulim⁸, Dominique Roulot⁹, Albert Tran¹⁰, Jean-Pierre Bronowicki¹¹, Jean-Pierre Zarski¹², Vincent Leroy¹², Ghassan Riachi¹³, Paul Cal es¹⁴, Jean-Marie P eron¹⁵, Laurent Alric¹⁶, Marc Bourli ere¹⁷, Philippe Mathurin¹⁸, Sebastien Dharancy¹⁸, Jean-Fr ed eric Blanc¹⁹, Armand Abergel²⁰, Lawrence Serfaty²¹, Ariane Mallat²², Jean-Didier Grang e²³, Pierre Attali²⁴, Yannick Bacq²⁵, Claire Wartelle²⁶, Th ong Dao²⁷, Yves Benhamou²⁸, Christophe Pilette²⁹, Christine Silvain³⁰, Christos Christidis³¹, Dominique Capron³², G erard Thieffin³³, Sophie Hillaire³⁴, Vincent Di Martino³⁵.

¹AP-HP, H opital Jean Verdier, Service d'H epatologie, Bondy, Universit  Paris 13, Bobigny et INSERM U1162, Universit  Paris 5, Paris; ²AP-HP, H opital Beaujon, Service d'H epatologie, Clichy; ³CHU Pontchaillou, Service d'H epatologie, Rennes; ⁴AP-HP, H opital Cochin, D epartement d'H epatologie et INSERM UMS20 et U1223, Institut Pasteur, Universit  Paris Descartes, Paris; ⁵H opital Saint Eloi, Service d'H epatologie, Montpellier; ⁶H opital Haut-L ev eque, Service d'H epatologie, Bordeaux; ⁷Institut Arnaud Tzanck, Service d'H epatologie, St Laurent du Var; ⁸H opital H otel Dieu, Service d'H epatologie, Lyon; ⁹AP-HP, H opital Avicenne, Service d'H epatologie, Bobigny; ¹⁰CHU de Nice, Service d'H epatologie, et INSERM U1065, Universit  de Nice-Sophia-Antipolis, Nice; ¹¹H opital Brabois, Service d'H epatologie, Vandoeuvre-les-Nancy; ¹²H opital Michallon, Service d'H epatologie, Grenoble; ¹³H opital Charles-Nicolle, Service d'H epatologie, Rouen; ¹⁴CHU d'Angers, Service d'H epatologie, Angers; ¹⁵H opital Purpan, Service d'H epatologie, Toulouse; ¹⁶CHU Toulouse, Service de M edecine Interne-P ole Digestif UMR 152, Toulouse; ¹⁷H opital Saint Joseph, Service d'H epatologie, Marseille; ¹⁸H opital Claude Huriez, Service d'H epatologie, Lille; ¹⁹H opital St Andr e, Service d'H epatologie, Bordeaux; ²⁰H opital H otel Dieu, Service d'H epatologie, Clermont-

Ferrand; ²¹AP-HP, Hôpital Saint-Antoine, Service d'Hépatologie, Paris; ²²AP-HP, Hôpital Henri Mondor, Service d'Hépatologie, Créteil; ²³AP-HP, Hôpital Tenon, Service d'Hépatologie, Paris; ²⁴AP-HP, Hôpital Paul Brousse, Service d'Hépatologie, Villejuif; ²⁵Hôpital Trousseau, Unité d'Hépatologie, CHRU de Tours; ²⁶Hôpital d'Aix-En-Provence, Service d'Hépatologie, Aix-En-Provence; ²⁷Hôpital de la Côte de Nacre, Service d'Hépatologie, Caen; ²⁸AP-HP, Groupe Hospitalier de La Pitié-Salpêtrière, Service d'Hépatologie, Paris; ²⁹CHU Le Mans, Service d'Hépatologie, Le Mans; ³⁰CHU de Poitiers, Service d'Hépatologie, Poitiers; ³¹Institut Mutualiste Montsouris, Service d'Hépatologie, Paris; ³²Hôpital Amiens Nord, Service d'Hépatologie, Amiens; ³³Hôpital Robert Debré, Service d'Hépatologie, Reims; ³⁴Hôpital Foch, Service d'Hépatologie, Suresnes; ³⁵Hôpital Jean Minjoz, Service d'Hépatologie, Besançon. France.

We also thank Kevin L. Erwin for writing assistance (English proofreading) and the sponsors of the subpopulations (who had no other role in the present study): VO-VCO study: the PHRC (Health Ministry, France); M116 study: Echosens, Paris, France; CO12 CirVir cohort: ANRS, France.

REFERENCES

- [1] De Franchis R, Faculty B V. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. *J Hepatol* 2015; 63: 743-752.
- [2] Easl. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol* 2021; 75: 659-689.

- [3] Ravaioli F, Montagnani M, Lisotti A, Festi D, Mazzella G, Azzaroli F. Noninvasive Assessment of Portal Hypertension in Advanced Chronic Liver Disease: An Update. *Gastroenterol Res Pract* 2018; 2018: 4202091.
- [4] Cales P, Buisson F, Ravaioli F, Berger A, Carboni C, Marasco G, et al. How to clarify the Baveno VI criteria for ruling out varices needing treatment by non-invasive tests. *Liver Int* 2019; 39: 49-53.
- [5] Stafylidou M, Paschos P, Katsoula A, Malandris K, Ioakim K, Bekiari E, et al. Performance of Baveno VI and Expanded Baveno VI Criteria for Excluding High-Risk Varices in Patients With Chronic Liver Diseases: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol* 2019; 17: 1744-1755 e1711.
- [6] Berger A, Ravaioli F, Farcau O, Festi D, Stefanescu H, Buisson F, et al. Including Ratio of Platelets to Liver Stiffness Improves Accuracy of Screening for Esophageal Varices That Require Treatment. *Clin Gastroenterol Hepatol* 2021; 19: 777-787.
- [7] Papatheodoridi M, Hiriart J B, Lupsor-Platon M, Bronte F, Boursier J, Elshaarawy, et al. Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease. *J Hepatol* 2021; 74: 1109-1116.
- [8] Zhou Y J, Gao F, Liu W Y, Wong G L, Mahadeva S, Raihan Nik Mustapha N, et al. Screening for compensated advanced chronic liver disease using refined Baveno VI elastography cutoffs in Asian patients with nonalcoholic fatty liver disease. *Aliment Pharmacol Ther* 2021; 54: 470-480.
- [9] Singal A K, Kamath P S. Model for End-stage Liver Disease. *J Clin Exp Hepatol* 2013; 3: 50-60.
- [10] Petta S, Sebastiani G, Bugianesi E, Vigano M, Wong V W, Berzigotti A, et al. Noninvasive Prediction of Esophageal Varices by Stiffness and Platelet in Nonalcoholic Fatty Liver Disease Cirrhosis. *J Hepatol* 2018; 69: 878-885.

- [11] De Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Baveno V I I F. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol* 2022; 76: 959-974.
- [12] Kamath P S, Kim W R, Advanced Liver Disease Study G. The model for end-stage liver disease (MELD). *Hepatology* 2007; 45: 797-805.
- [13] Boursier J, Konate A, Gorea G, Reaud S, Quemener E, Oberti F, et al. Reproducibility of liver stiffness measurement by ultrasonographic elastometry. *Clin Gastroenterol Hepatol* 2008; 6: 1263-1269.
- [14] Collins G S, Reitsma J B, Altman D G, Moons K G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; 350: g7594.
- [15] Cales P, Oberti F, Bernard-Chabert B, Payen J L. Evaluation of Baveno recommendations for grading esophageal varices. *J Hepatol* 2003; 39: 657-659.
- [16] Abraldes J G, Bureau C, Stefanescu H, Augustin S, Ney M, Blasco 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.
- [17] Augustin S, Pons M, Maurice J B, Bureau C, Stefanescu H, Ney M, 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.
- [18] Boursier J, De Ledinghen V, Poynard T, Guechot J, Carrat F, Leroy V, et al. An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. *J Hepatol* 2015; 62: 807-815.
- [19] Boursier J, Zarski J P, De Ledinghen V, Rousselet M C, Sturm N, Lebaill B, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013; 57: 1182-1191.

- [20] Usher-Smith J A, Sharp S J, Griffin S J. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* 2016; 353: i3139.
- [21] Obuchowski N A, Graham R J, Baker M E, Powell K A. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. *AJR Am J Roentgenol* 2001; 176: 1357-1362.
- [22] Pateu E, Oberti F, Cales P. The noninvasive diagnosis of esophageal varices and its application in clinical practice. *Clin Res Hepatol Gastroenterol* 2018; 42: 6-16.
- [23] Hu Y, Wen Z. Validation and comparison of non-invasive prediction models based on liver stiffness measurement to identify patients who could avoid gastroscopy. *Sci Rep* 2021; 11: 150.
- [24] Villanueva C, Albillos A, Genesca J, Garcia-Pagan J C, Calleja J L, Aracil C, et al. beta blockers to prevent decompensation of cirrhosis in patients with clinically significant portal hypertension (PREDESCI): a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* 2019; 393: 1597-1608.
- [25] Garcia-Tsao G, Abraldes J G. Nonselective Beta-Blockers in Compensated Cirrhosis: Preventing Variceal Hemorrhage or Preventing Decompensation? *Gastroenterology* 2021; 161: 770-773.
- [26] Silvain C, Chauvin C, Verneau A, Carretier M, Beauchant M. [How many cirrhotic patients may receive propranolol after digestive hemorrhage?]. *Gastroenterol Clin Biol* 1985; 9: 670-673.
- [27] Hayward K L, Valery P C, Cottrell W N, Irvine K M, Horsfall L U, Tallis C J, et al. Prevalence of medication discrepancies in patients with cirrhosis: a pilot study. *BMC Gastroenterol* 2016; 16: 114.

- [28] Sovaila S, Purcarea A, Gheonea D, Ciurea T. Specific Factors That Influence Adherence to Beta Blocker Treatment in Primary Prevention of Variceal Bleeding in Cirrhotic Romanian Patients. a Proof of Concept Qualitative Study. *J Med Life* 2018; 11: 355-358.
- [29] Pons M, Augustin S, Scheiner B, Guillaume M, Rosselli M, Rodrigues S G, et al. Noninvasive Diagnosis of Portal Hypertension in Patients With Compensated Advanced Chronic Liver Disease. *Am J Gastroenterol* 2021; 116: 723-732.
- [30] Yakovchenko V, Bolton R E, Drainoni M L, Gifford A L. Primary care provider perceptions and experiences of implementing hepatitis C virus birth cohort testing: a qualitative formative evaluation. *BMC Health Serv Res* 2019; 19: 236.
- [31] Ispas S, So S, Toy M. Barriers to Disease Monitoring and Liver Cancer Surveillance Among Patients with Chronic Hepatitis B in the United States. *J Community Health* 2019; 44: 610-625.

Supplemental Material

The Supplemental Material includes 16 tables and 5 figures.

FIGURE LEGENDS

Figure 1. Characteristics of the VARS score. Panel (A): AUROCs for VNT; magnifications on right show that VARS reached 100% sensitivity and specificity. Panel (B): VARS calibration for VNT probability via VARS percentiles.

Figure 2. New test results as a function of screening strategy. Panel (A): population screening. Panel (B): individual and undifferentiated screenings. Figures within bars from the whole population indicate category prevalence (top) and VNT prevalence and, in brackets, category proportion among VNT (bottom).

Figure 3. Main characteristics of the three screening strategies.

Figure 4. Clinical use of new tests according to population and individual screenings.

Table 1. Main characteristics of the three strategies for VNT screening.

Strategy	Cut-offs for VNT ruled:		Advantages	Limits
	out	in		
Individual patient	95% NPV	100% PPV / specificity	High performance. Easier cut-off determination. The rate of unnecessary endoscopies is not sensitive to liver dysfunction.	Increased number and liver dysfunction of missed VNT. High VNT prevalence dependence. Restricted to MELD <10. Comparison of missed VNT rate is less powerful ^a . LR- is not applicable ^b .
Population	95% sensitivity	100% PPV / specificity	Optimal safety / performance ratio. The lowest liver dysfunction in missed VNT. LR- is a unique diagnostic descriptor.	Weak VNT prevalence dependence. Cut-offs are less optimistic since the reference population is smaller.
100% safety ^c	100% NPV/ sensitivity	100% PPV / specificity	Maximum safety. Weak VNT prevalence dependence.	Poor performance.

VNT: varices needing treatment, NPV: negative predictive value, PPV: positive predictive value, LR-: negative likelihood ratio.

New strategy characteristics developed in the present study are in bold.

^a Because of using an unpaired statistical test.

^b Because LR are based on sensitivity and specificity.

^c Adapted to individual and population screening.

Table 2. Safety and performance (%) of new tests as a function of populations and screening strategies.

	All patients	Derivation	Validation	<i>p</i> ^a
Population screening:				
VANT:				
Missed VNT ^b	4.7 (2.5-7.1)	4.7 (2.2-7.4)	4.5 (0.9-9.3)	1
Spared endoscopy	40.2 (38.0-42.1)	39.6 (37.2-41.9)	41.4 (37.7-44.7)	0.416
VARs:				
Missed VNT ^b	3.8 (2.0-5.9)	4.7 (2.2-7.6)	1.8 (0.0-4.9)	0.237
Spared endoscopy	32.0 (30.2-33.9)	31.2 (29.0-33.3)	33.5 (30.1-36.6)	0.274
Undifferentiated screening by VARs:				
Missed VNT	0	0	0	1
Spared endoscopy	12.0 (10.8-13.3)	11.8 (10.2-13.4)	12.3 (10.1-14.8)	0.733
Individual screening by VARs:				
Missed VNT ^c	4.96 (3.8-6.1)	4.98 (3.7-6.5)	4.90 (3.1-6.9)	1
Spared endoscopy	62.0 (60.1-64.2)	62.1 (59.6-64.4)	61.9 (58.3-65.2)	0.964

VNT: varices needing treatment. Results in brackets are 95% CI obtained by bootstrapping based on 1000 samples stratified on etiology and sex.

^a Fisher test between derivation and validation sets.

^b In patients with VNT.

^c In patients with spared endoscopy.

Table 3. Safety and performance (%) of tests for population screening in the whole population and as a function of etiology.

	B6C	Anticipate	VariScreen	VANT	p^a
Whole population:					
Missed VNT	2.6 (0.9-4.5)	3.5 (1.5-5.7)	3.8 (1.9-6.1)	4.7 (2.5-7.1)	0.322
Spared endoscopy	24.1 (22.3-25.8)	24.7 (22.8-26.4)	35.3 (34.1-38.2)	40.2 (38.0-42.1)	<0.001 ^b
LR-	0.094	0.123	0.093	0.103	-
Virus:					
Missed VNT	1.2 (0.0-3.1)	2.4 (0.5-5.0)	3.6 (1.1-6.7)	4.8 (1.8-8.1)	0.083
Spared endoscopy	21.8 (19.2-24.3)	25.2 (22.8-27.6)	38.1 (35.2-40.9)	41.1 (38.4-43.8)	<0.001
NAFLD:					
Missed VNT	7.4 (2.4-14.0)	7.4 (2.5-14.1)	4.9 (1.0-10.3)	4.9 (1.2-10.0)	0.721
Spared endoscopy	33.6 (30.3-37.2)	29.0 (25.4-32.7)	37.7 (33.9-41.6)	44.2 (40.4-48.0)	<0.001
ALD:					
Missed VNT	1.1 (0.0-3.5)	2.1 (0.0-5.6)	3.2 (0.0-7.4)	4.3 (1.0-9.1)	0.232
Spared endoscopy	16.0 (12.6-19.5)	17.3 (13.6-20.8)	24.9 (20.6-28.9)	32.0 (27.6-36.0)	<0.001
Comparison between etiologies (p ^c):					
Missed VNT	0.021	0.130	0.825	0.973	-
Spared endoscopy	<0.001	<0.001	<0.001	<0.001	-

B6C: Baveno VI criteria, VNT: varices needing treatment, LR-: negative likelihood ratio. Results in brackets are 95% CI obtained by bootstrapping based on 1000 samples stratified on etiology and sex. A colour version of this table is available in the Complementary Data.

^a Paired Cochran test.

^b Each pair comparison: p<0.001 except for B6C vs Anticipate: p=0.393 by McNemar test.

^c Unpaired Chi² test for spared endoscopy and likelihood ratio test for missed VNT.

Table 4. Safety and performance (%) of tests for individual screening in the whole population and as a function of etiology.

	EB6C	<i>p</i> ^a	VARs	<i>p</i> ^a	NCC
Whole population:					
Missed VNT	4.0 (2.8-5.1)	0.246	4.98 (3.8-6.1)	-	-
Spared endoscopy	42.9 (40.9-45.1)	<0.001	62.0 (60.0-64.0)	-	-
LR-	0.236	-	- ^b	-	-
Virus:					
Missed VNT	3.5 (1.9-5.2)	0.392	4.5 (3.1-6.2)	-	-
Spared endoscopy	42.3 (39.5-45.2)	<0.001	62.5 (59.7-65.4)	-	-
LR-	0.213	-	- ^b	-	-
NAFLD:					
Missed VNT	4.2 (2.3-6.3)	0.846	4.5 (2.8-6.6)	0.984	4.5 (2.7-6.5)
Spared endoscopy	55.8 (51.8-59.3)	<0.001	72.8 (69.5-76.2)	<0.001	65.1 (61.3-68.5)
LR-	0.326	-	- ^b	-	0.350
ALD:					
Missed VNT	4.96 (1.5-9.1)	0.349	7.6 (4.4-11.3)	-	-
Spared endoscopy	25.8 (21.9-29.7)	<0.001	45.2 (40.9-49.9)	-	-
LR-	0.208	-	- ^b	-	-
Comparison between etiologies (<i>p</i> ^c):					
Missed VNT	0.722	-	0.162	-	-
Spared endoscopy	<0.001	-	<0.001	-	-

EB6C: expanded Baveno VI criteria, NCC: NAFLD cirrhosis criteria, VNT: varices needing treatment, LR-: negative likelihood ratio. Results in brackets are 95% CI obtained by bootstrapping based on 1000 samples stratified on etiology and sex.

^a Paired McNemar test for spared endoscopy and unpaired Chi² test for missed VNT.

^b Not evaluable as there are three VNT categories.

^c Unpaired Chi² test for spared endoscopy and likelihood ratio test for missed VNT.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

**Data sharing*****Patient data***

Data entry sheet is available on simple request.

Individual data are available for validation studies or meta-analyses with predefined strategy.

Contact the corresponding author.

Test algorithms

They are fully detailed in the Supplemental Material.

Their calculation is available for individual patient and data base at an URL stated in the manuscript.

The methodological requirements for this use are detailed in the manuscript.

Additional data

Several additional data are detailed in the Supplemental material and in the Complementary material whose URL is indicated in the Supplemental material.

Figure 1

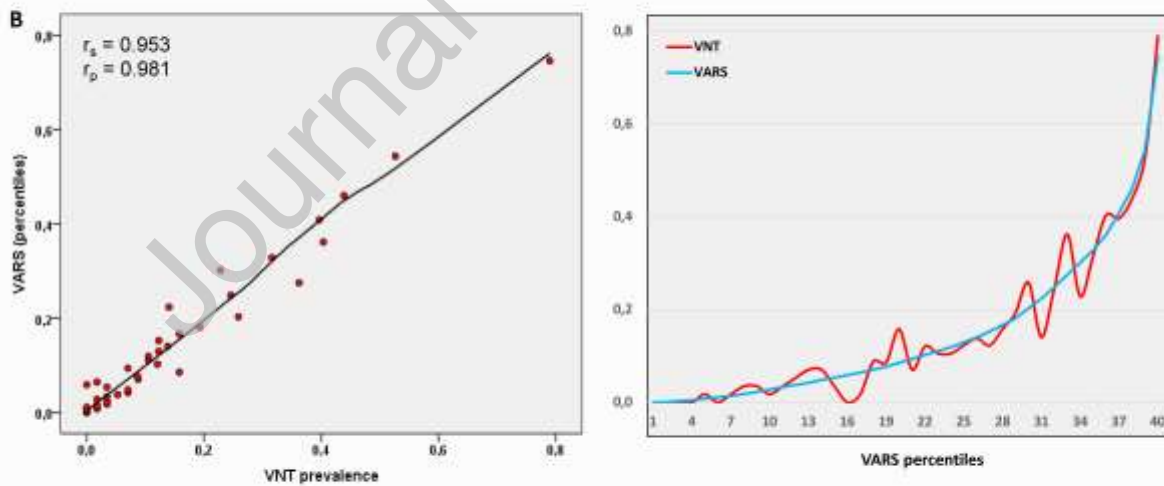
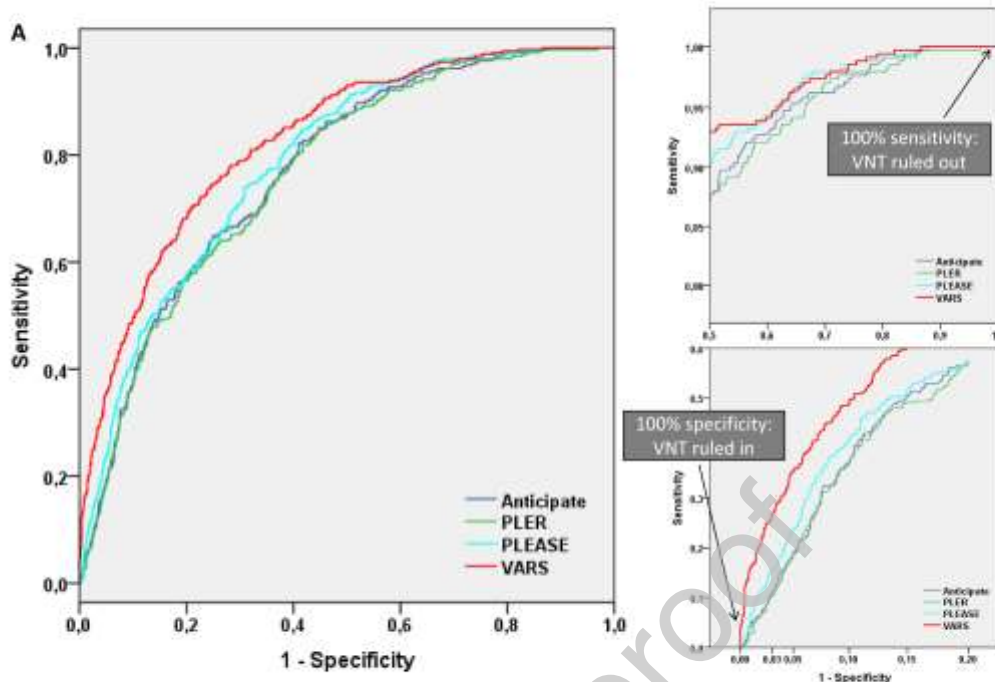


Figure 2

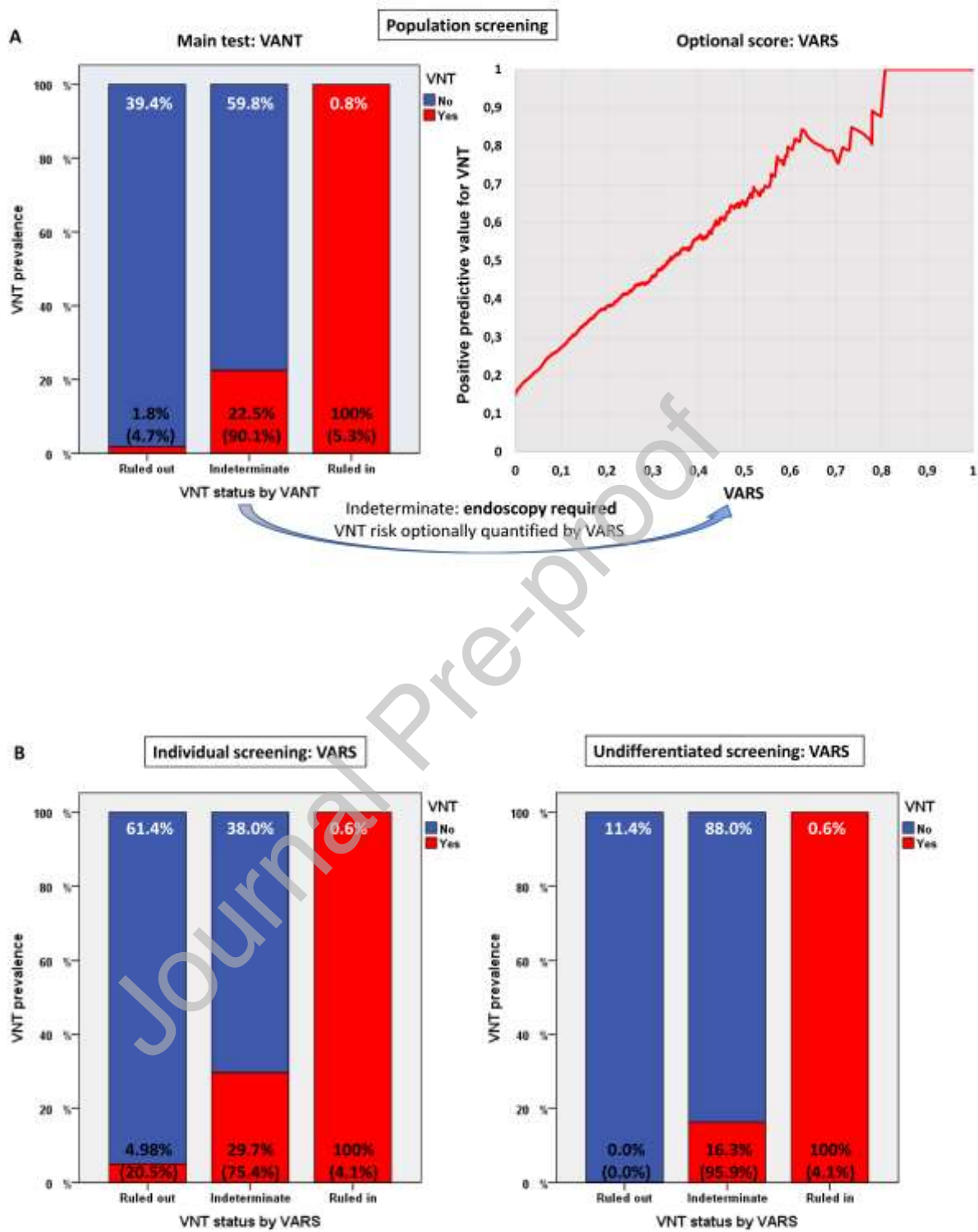


Figure 3

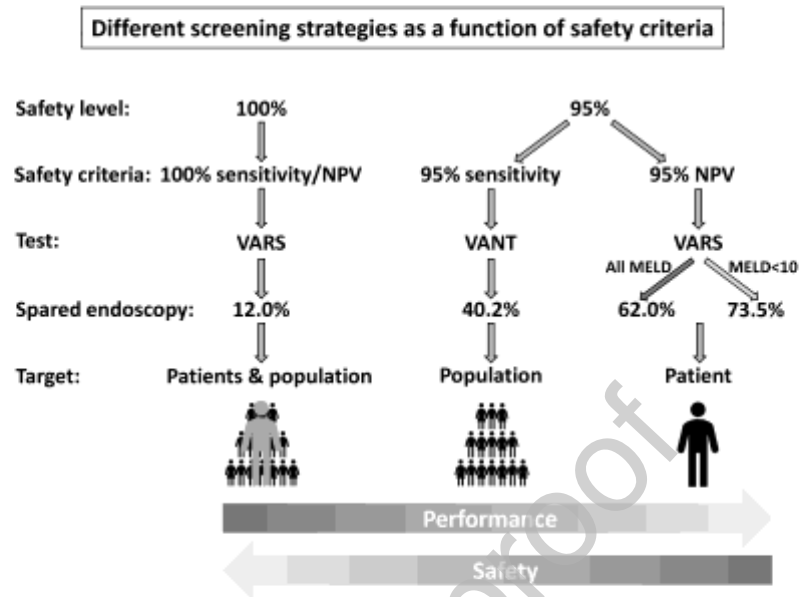


Figure 4

