

# Impact of steatosis and inflammation definitions on the performance of NASH tests

Thierry Poynard<sup>a,b,c</sup>, Mona Munteanu<sup>d</sup>, Frederic Charlotte<sup>a,b</sup>, Hugo Perazzo<sup>a,b,c</sup>, Yen Ngo<sup>d</sup>, Olivier Deckmyn<sup>d</sup>, Raluca Pais<sup>a,b,c</sup>, Philippe Mathurin<sup>e</sup> and Vlad Ratziu<sup>a,b,c</sup>; for the FLIP consortium, the FibroFrance-CPAM group; and the FibroFrance-Obese group

**Background and aim** One of the unmet needs in subjects with metabolic risks is the prediction of metabolic liver disease by noninvasive tests. The construction of performant tests is dependent on the appropriateness of the histological reference definition. The aim of this study was to analyze the limitations of similar European (Fatty Liver Inhibition of Progression) and USA (Clinical-Research-Network) standard definitions and their impact on the construction of tests.

**Methods** We hypothesized that a simpler histological definition of non-alcoholic steato-hepatitis (NASH), which does not require the presence of steatosis and the presence of both lobular inflammation and ballooning, should improve the concordance rates with previously validated blood tests. We reviewed the landmark studies in metabolic liver disease, sources of the standard definitions, and we compared the adequacy of these standards to other possible definitions in 1081 subjects with biopsies, by concordance and accuracy rates.

**Results** The limitations of standard definitions included the presence of appropriate controls in only 6.6% of landmark studies, an arbitrary definition of steatosis and NASH covering only four (15%) out of 27 possible combinations of features, compared with 18 (67%) for a simplified NASH definition, which did not require steatosis. A total of 39/1081 (3.6%) cases were not identified by standard definition, but were identified by the simplified definition as significant active disease, including 15 cases with significant fibrosis. The simplified definition increased the  $\kappa$  concordance ( $P < 0.0001$ ) between test prediction and histological reference.

**Conclusion** A simplified definition of NASH could help in the construction of biomarkers with higher performances. Eur J Gastroenterol Hepatol 30:384–391

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc.

## Introduction

One of the unmet needs in subjects with metabolic risk factors is the availability of noninvasive tests (NITs) for the prediction of NASH and metabolic liver disease (MLD), combining steatosis and the features of necroinflammatory activity [1–4]. The nonalcoholic fatty liver disease (NAFLD) activity score, developed by the NASH Clinical Research Network (NASH-CRN) [4] and used in the FLIP-algorithm [5,6], is the current reference for the diagnosis of NASH [4]

(Table 1). However, this reference has several limitations that have been discussed in the literature [7–10], including sampling error [7] and interobserver variations among pathologists [9].

Several other limitations should be also fully discussed, such as the exclusion of patients with metabolic risk factors and histological fibrosis or inflammation and without other causes of liver disease. These patients can have type-2 diabetes with significant histological fibrosis but with steatosis less than 5%, or grade 2 lobular inflammation without ballooning. Another limitation is the current definition of the disease of interest NAFLD: ‘hepatic steatosis in the absence of other known liver disease’. This definition seems better adapted to the pathologist who is scoring histological features in ‘patients with steatosis’, than to the clinician who is seeking to identify the presence of significant MLD (fibrosis or activity) in patients with metabolic factors, the real context of use, and whether steatosis is present or not. The absence of steatosis, or of one type of inflammation, can be attributable to the sampling error of biopsy, or to the temporal variability of these features, such as burn-out cirrhosis [10].

We hypothesized that a simpler histological definition of NASH, which does not require the presence of steatosis and the presence of both lobular inflammation and ballooning, should improve the concordance rates with previously validated blood tests (Fig. 1). The first step was to underline that, in the landmark studies (population-1) [4,5,9,11,12] leading to NASH definitions, only a very small number of true controls were analyzed, despite the fact that they are

European Journal of Gastroenterology & Hepatology 2018, 30:384–391

**Keywords:** FibroTest, FLIP algorithm, metabolic liver disease, NAFLD score, NASH quantitative test, NASH simplified definition, NASH Test, spectrum effect, steatosis definition, SteatoTest

<sup>a</sup>Groupe Hospitalier Pitié Salpêtrière APHP, Sorbonne Universités, UPMC Univ Paris 06, UMR\_S 938, <sup>b</sup>Institute of Cardiometabolism and Nutrition (ICAN), <sup>c</sup>INSERM, <sup>d</sup>BioPredictive, Paris and <sup>e</sup>Service Maladie de l’Appareil Digestif, Hôpital Huriez, CHRU Lille, Lille, France

Correspondence to: Thierry Poynard, Hepato-Gastroenterology Department, Groupe Hospitalier Pitié Salpêtrière, 75013 Paris, France

E-mail: thierry@poynard.com

**Received** 15 September 2017 **Accepted** 8 November 2017

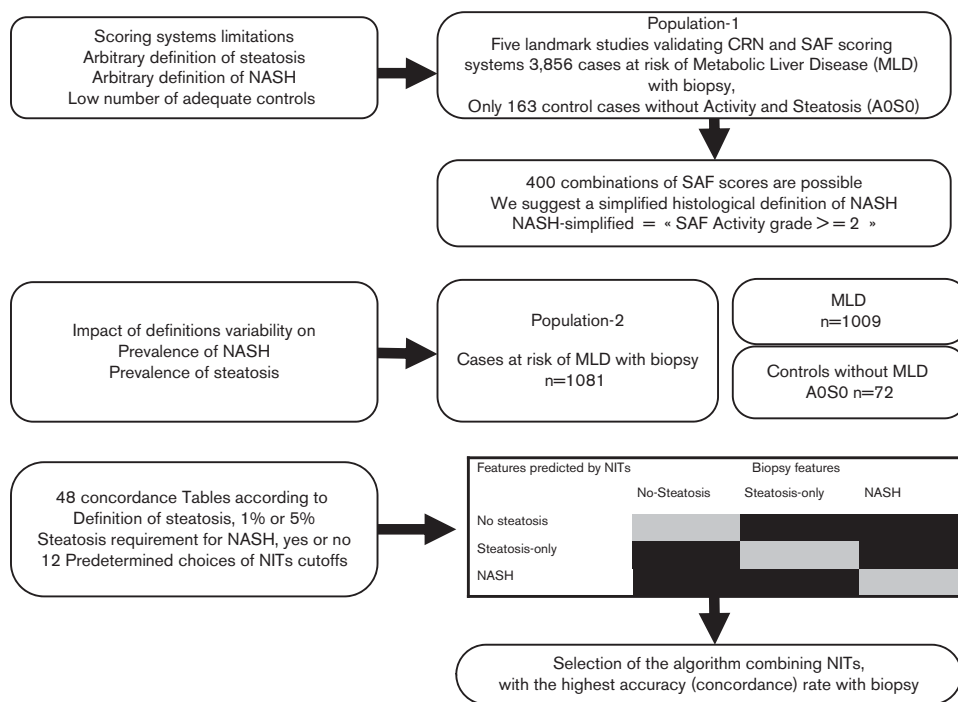
Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s website, [www.eurojgh.com](http://www.eurojgh.com).

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

**Table 1.** Comparison between histological scoring systems published and suggested in metabolic liver diseases

	SAF-FLIP	NASH-CRN	SAF-simplified MLD score
Fibrosis			
F0	None	None	None
F1	Perisinusoidal or portal	Perisinusoidal or portal	Perisinusoidal or portal
F2	Perisinusoidal and portal without bridging	Perisinusoidal and portal without bridging	Perisinusoidal and portal without bridging
F3	Bridging	Bridging	Bridging
F4	Cirrhosis	Cirrhosis	Cirrhosis
Activity <sup>a</sup>			
A0	Ballooning + lobular inflammation = 0	Ballooning + lobular inflammation = 0	Ballooning + lobular inflammation = 0
A1	Ballooning + lobular inflammation = 1	Ballooning + lobular inflammation = 1	Ballooning + lobular inflammation = 1
A2	Ballooning + lobular inflammation = 2	Ballooning + lobular inflammation = 2	Ballooning + lobular inflammation = 2
A3	Ballooning + lobular inflammation = 3	Ballooning + lobular inflammation = 3	Ballooning + lobular inflammation = 3
A4	Ballooning + lobular inflammation = 4	Ballooning + lobular inflammation = 4	Ballooning + lobular inflammation = 4
A5	No A5 category: lobular inflammation with three grades	Ballooning + lobular inflammation = 5	Ballooning + lobular inflammation = 4
Steatosis (%)			
S0	< 5	< 5	0
S1	5–33	5–33	> 0–5
S2	> 33–66	> 33–66	> 5–33
S3	> 66	> 66	> 33
NASH definition	FLIP-algorithm: S1–S3 and A2–A4 with at least ballooning B1 and lobular inflammation L1	S1–S3 and A2–A5 with at least ballooning B1 and lobular inflammation L1	H-NASHs: A2–A4, steatosis not required

CRN, Clinical-Research-Network; FLIP, Fatty Liver Inhibition of Progression; MLD, metabolic liver disease; SAF, steatosis, activity and fibrosis.



**Fig. 1.** Flow charts of methodology and the corresponding populations used in the analyses. The context of use population is cases at risk of metabolic liver disease (MLD), preferred to the acronym nonalcoholic fatty liver disease (NAFLD), as requirement of steatosis is not mandatory in the absence of other causes of liver disease. Steatosis could be absent because of sampling error or temporal variability. SAF referred to steatosis, activity and fibrosis elementary histological features. Scores referred to NASH Clinical Research Network (NASH-CRN) and fatty liver inhibition of progression (FLIP) scoring systems. NASHs referred to the proposed simplified histological definition of NASH, being at least a grade 2 in the SAF scoring system not requiring both ballooning and lobular inflammation, and not requiring presence of steatosis (1 or 5%), in the presence of metabolic risk factors and in the absence of all other causes of liver disease. NITs is the acronym of noninvasive tests.

mandatory for assessing the specificity of NITs. These controls must be defined as cases at risk of MLD but without histological activity and steatosis (A0S0). The second step was to demonstrate that the definition of NASH currently recommended both by CRN and steatosis, activity and fibrosis (SAF) scoring systems, represented only one out of 400 possible combinations, according to the score of each elementary feature (from S0A0F0 to S3A4F4) and the

requirement of steatosis (yes or no) as well as the requirement of a cutoff of 1 or 5% for defining steatosis.

Thus, we first systematically assessed the variability of NASH and steatosis definitions in the landmark validation studies of existing scoring systems. Thereafter, we evaluated the number of possible combinations that could be used for defining NASH. The next step was to demonstrate the significant impact of choosing different NASH definitions

on NITs' performance. For this purpose, we assessed the strength of agreement between NITs' predictions and the various definitions of NASH, steatosis-only and no-steatosis. We observed very significant differences permitting to identify better algorithms for the prediction of simplified NASHs.

## Patients and methods

The aim of this proof of concept study was to improve the construction of NITs for the diagnosis of inflammatory activity in subjects at risk of MLD. From the analyses of the standard definitions (FLIP-CRN), the different other possible definitions of inflammatory activity, and the appropriate choices of NITs' cutoffs, we identified in a large population with biopsy, algorithms with higher accuracy than ActiTest (BioPredictive, Paris, France), a previously validated inflammatory NIT. Details on patients were given elsewhere in separate studies [13,14], focusing on the validation of NITs using FibroSTARD recommended guidelines [15].

In the FLIP cohort, written patient consent for routine liver biopsy and data collection was obtained from each subject before inclusion. This epidemiological, noninterventional study was exempt from IRB review (Ethical Committee of 'Comite de Protection des Personnes of Paris-Ile-de-France', FIBROFRANCE project. CPP-IDF-VI, 10-1996-DR-964, DR-2012-222 and USA-NCT01927133).

## Review of the landmark studies of metabolic liver diseases

We reviewed the landmark studies in MLD, sources of the FLIP-CRN definitions (population-1), to clarify the main definitions of the population of interest [4,5,9,11,12]. For NITs' construction, the appropriate context of use was defined as carriers of metabolic risk factor, who are at risk of MLD, the disease of interest, in the absence of other known liver disease. The definition of MLD included three main features, SAF. PUBMED was screened with the following tags: 'NAFLD, metabolic liver disease, biopsy, human' (5 January 2017). Inclusion criteria for studies were adult population with 500 or more biopsies (population-1, Fig. 1) (see Supplementary File S2, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

## Current definitions versus simplified definitions applied to population-2

The current definitions of fibrosis stages, steatosis and activity grades and NASH, according to CRN or FLIP, and three simplified definitions are detailed in Table 1 [5,6]. The definitions of fibrosis stages were the same. The definitions of steatosis grades were only different for simplified definitions with a more sensitive score including 0% grade (S0), a minimal grade (1 to <5%; S1), moderate grade (5–33%; S2) and severe grade (>33%; S3). For activity grade, CRN had six grades, one more than SAF-FLIP, limited to three grades for both ballooning and lobular inflammation. For NASH, the definitions were the same for CRN and SAF-FLIP scoring systems, and not requiring steatosis and both ballooning and lobular inflammation for the simplified H-NASHs' definition. To reduce inter-observer variability and homogenize the reading using the SAF-FLIP histological classification, we used only reports reviewed by members of the FLIP Pathology Consortium

(DT and PB for the FLIP subpopulation and FC for the FibroFrance subpopulation) [5,6].

## Number of possible combinations for defining metabolic liver disease and NASH

Ideally, all cases of the population of interest should be classified according to the SAF scoring systems from S0A0F0 to S3A4F4 [4–6]. This represents a total of 100 possible combinations of two features with five levels and one with four levels ( $5 \times 5 \times 4 = 100$ ). This number of combinations is minimal because it should be multiplied by two if subjects without steatosis were included, and also if the cutoff for defining steatosis was 1 or 5%, for a total of 400 combinations. For NASH-CRN, the number of combinations could be even greater because the activity score uses a supplementary grade of A5 (Table 1).

The NAFLD-CRN and the FLIP-algorithm proposed a seemingly straightforward definition of NASH: 'A case presenting with at least grade 1 of each of the three features (steatosis, ballooning, lobular inflammation) is classified as NASH'. Other cases without NASH are diagnosed as 'steatosis-only'. However, these definitions only represent four (15%) out of 27 possible combinations of steatosis, ballooning and lobular inflammation (three levels each) (Table 2).

## New simplified definitions of NASH

We analyzed three simplified definitions of NASH versus the standard one. One that did not exclude subjects with 1–4% of steatosis (NASH-AlgoV2), one that did not require the presence of steatosis, whatever its percentage (NASH-AlgoV3), and the most simplified H-NASHs, which simply used the SAF-activity grade of at least 2 as a minimum for NASH (clinically significant activity) without requiring steatosis or the simultaneous presence of ballooning and lobular inflammation. The H-NASH's definition represented 18 out of 27 (66.7%) possible combinations (Table 2). Therefore, this definition was taken as the histological reference for identifying the best NIT for the diagnosis of NASHs.

We compared the prevalence of MLD in the disease of interest group (patients with metabolic risk factors and MLD) to assess the sensitivity of NITs, as well as in the control group to assess the specificity, according to the chosen histological definitions. The prevalence of NASH was calculated with the FLIP-algorithm [6] and three simplified definitions.

## Impact of histological definitions on NITs' performances

We focused on the following four factors of variability possibly associated with NIT's performances, two related to the choices of histological definitions of NASH and steatosis, and two to the choices of cutoffs of validated NITs for the prediction of steatosis (SteatoTest, BioPredictive, Paris, France) and activity (ActiTest), respectively [14,16].

The choice for NASH definition was the requirement of steatosis for NASH diagnosis or not (two levels: yes, or no). The choice for steatosis definition was of 1 or 5% of hepatocytes (two levels). The choices for SteatoTest were among the four grades (four levels), and for ActiTest three grades (three levels), as previously validated [14,16] (see Supplementary Table S3, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>). These combinations of

**Table 2.** Prevalence of histological NASH according to the 27 possible definitions of steatosis and activity features in the Fatty Liver Inhibition of Progression population (*n* = 1081)

Features of SAF scoring system			NASH histological definitions according to 27 combinations				Prevalence of each combination	
Steatosis grade	Activity grade		FLIP-CRN		Simplified		Steatosis, ballooning and lobular inflammation	95% CI
	Ballooning (B)	Lobular inflammation (L)	V1	V2	V3	H-NASHs		
3 levels	3 levels	3 levels	0 = absence 1 = presence				<i>n</i> (%)	95% CI
0%	0	0	0	0	0	0	38 (3.5)	2.5–4.8
0%	0	1	0	0	0	0	5 (0.5)	0.1–1.1
0%	0	2	0	0	0	1	0 (0)	0.0–0.3
0%	1	0	0	0	0	0	3 (0.3)	0.1–0.8
0%	1	1	0	0	1	1	3 (0.3) <sup>a</sup>	0.1–0.8
0%	1	2	0	0	1	1	0 (0)	0.0–0.3
0%	2	0	0	0	0	1	1 (0.1) <sup>a</sup>	0.0–0.5
0%	2	1	0	0	1	1	0 (0)	0.0–0.3
0%	2	2	0	0	1	1	1 (0.1) <sup>a</sup>	0.0–0.5
Subtotal	Steatosis 0%						51 (4.7)	3.5–6.2
1–4%	0	0	0	0	0	0	34 (3.1)	2.2–4.4
1–4%	0	1	0	0	0	0	1 (0.1)	0.0–0.5
1–4%	0	2	0	0	0	1	0 (0)	0.0–0.3
1–4%	1	0	0	0	0	0	3 (0.3)	0.1–0.8
1–4%	1	1	0	1	1	1	1 (0.1) <sup>a</sup>	0.0–0.5
1–4%	1	2	0	1	1	1	0 (0)	0.0–0.3
1–4%	2	0	0	0	0	1	0 (0)	0.0–0.3
1–4%	2	1	0	1	1	1	0 (0)	0.0–0.3
1–4%	2	2	0	1	1	1	0 (0)	0.0–0.3
Subtotal	Steatosis 1–4%						39 (3.6) <sup>b</sup>	2.6–4.9
≥ 5%	0	0	0	0	0	0	241 (22.3)	19.8–24.9
≥ 5%	0	1	0	0	0	0	90 (8.3)	6.7–10.1
≥ 5%	0	2	0	0	0	1	12 (1.1) <sup>a</sup>	0.6–1.9
≥ 5%	1	0	0	0	0	0	78 (7.2)	5.7–8.9
≥ 5%	1	1	1	1	1	1	229 (21.2)	18.8–23.7
≥ 5%	1	2	1	1	1	1	47 (4.3)	3.2–5.7
≥ 5%	2	0	0	0	0	1	21 (1.9) <sup>a</sup>	1.2–3.0
≥ 5%	2	1	1	1	1	1	158 (14.6)	12.6–16.9
≥ 5%	2	2	1	1	1	1	115 (10.6)	8.9–12.6
Subtotal	Steatosis ≥ 5%						991 (91.7)	89.9–93.3
Total	Combinations		4	8	12	18	1081 (100)	99.7–1.00

	<i>n</i> (%)	95% CI
AOS0 prevalence by FLIP-CRN ( <i>n</i> = 72) including 38 S0% and 34 S1–S4%		
NASH prevalences		
NASH cases using V1: original FLIP-V1 required steatosis 5%, ≥ B1–L1	549 (50.8)	47.8–53.8
NASH cases using V2: steatosis expanded to 1%	550 (50.9)	47.9–53.9
NASH cases using V3: steatosis not required, but ≥ B1–L1	554 (51.2)	48.2–54.2
NASHr cases using NASHs: steatosis, B1–L1 not required (SAF-A ≥ 2)	588 (54.4)	51.4–57.4
Significant SAF activity (ballooning + lobular inflammation) ≥ 2, but not NASH by FLIP-V1 ( <i>n</i> = 39) <sup>a</sup>		
With steatosis 0% <sup>a</sup>	5 (0.5)	0.1–1.1
With steatosis 1–4% not recognized as NASH by original FLIP-algorithm <sup>a</sup>	1 (0.1)	0.0–0.5
With steatosis ≥ 5% not recognized as NASH by original FLIP-algorithm <sup>a</sup>	33 (3.1)	2.1–4.3
Among these 39 cases, 15 had a significant fibrosis, six F2, five F3 and four cirrhosis		
Among 39 steatosis 1–4%, one case had cirrhosis and 10 cases had F1 <sup>b</sup>		

CRN, Clinical-Research-Network; FLIP, Fatty Liver Inhibition of Progression; SAF, steatosis, activity and fibrosis.

four variability factors lead to 48 possible combinations (levels = 2 × 2 × 4 × 3). To identify the best NITs, we analyzed the strength of agreement, in each of the 48 ‘3 × 3 tables’ comparing the cases’ classification (NASH, steatosis, no-steatosis), by histology (columns) or by NITs (row).

We calculated the strength of agreement using the accuracy rate and the concordance  $\kappa$  index for each table. Because the risks for false positives/false negatives for NITs and for histological definitions were not the same, without perfect reference (gold standard), stronger agreement suggested better choice for the corresponding histological reference [17–19].

**Patients**

The included patients of population-2 were from the FibroFrance project (USA-NCT01927133) and the FLIP

consortium (<http://www.flip-fp7.eu/>). All clinical investigations were performed according to the principles of the Declaration of Helsinki. All authors had access to the study data and reviewed and approved the final manuscript.

**Blood tests**

As a proof of concept, we constructed a new quantitative NIT (NIT-NASHv0, patent pending) for the diagnosis of H-NASHs, using the simplified H-NASHs as the best reference for NASH, according to its accuracy rate. In this method, the originality was not only the definition of NASH as an activity SAF-grade-2 without requirement of both lobular inflammation and ballooning of at least 1, but also without the requirement of the presence of steatosis of

at least 5%. Therefore, cases of SAF-grade-2 despite steatosis 1–4% or absence of steatosis (0%) will not be excluded. NIT-NASHv0 was developed using 11 components of the SteatoTest-ActiTest without glucose and BMI, and, finally, its performance was compared with ActiTest, a previously validated NIT for inflammatory activity.

In a separate manuscript submitted, we applied the same concept, but following the recommended FibroSTARD standards for liver NITs [13,15]. The population-2 was randomized in a working group (541 cases), where another test (NIT-NASHs, patent pending) was constructed, and in a control population (540 cases) for internal validation of this new test.

The FibroTest, ActiTest and SteatoTest are patented NITs (NASH-FibroSure in USA for MLD) that have been extensively validated to assess the stages of fibrosis, activity and the grades of steatosis using the METAVIR [20–23] or SAF-scoring system [14]. The FibroTest (BioPredictive, Paris, France) includes serum  $\alpha$ 2-macroglobulin, apolipoprotein-A1, haptoglobin, total bilirubin, and  $\gamma$ -glutamyl-transpeptidase. The ActiTest includes the same components plus alanine aminotransferase (ALT). The SteatoTest includes the same six components of the FibroTest and ActiTest plus BMI, serum cholesterol, triglycerides, and glucose. Exclusion criteria were nonreliable results identified using security control algorithms [24].

### Statistical methods

The main criteria for assessing the strength of concordance was the accuracy rate. We use also the quadratic weighted  $\kappa$  coefficient ( $\kappa$ ), as the results of these two tests may be different because of departures from symmetry in the vertical and horizontal marginal totals of the concordance table, often because of a low prevalence of events [25]. The regression curves between the  $\kappa$  and accuracy according to variability factor levels were compared by the Fisher Test [26]. This allowed the clusters of the 48 algorithms to be displayed graphically, along with each variability factor level, and to identify the algorithms with the highest and lowest performances on these bidimensional figures. NCS-2013 statistical software was used [26].

## Results

### Review of the landmark metabolic liver disease studies

Only five of 1111 screened studies (0.4%) were included (see Supplementary File S2, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>). None of these studies evaluated the full spectrum of steatosis, as the number of cases without any steatosis (0% of hepatocytes) (see Supplementary File S2, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>) was not available, with only 163 (4.2%) cases with A0S0 (S0 defined as <5%), which is considered as the control group without MLD. The CRN's first validation included only 13 (2.3%) controls A0S0 and three (0.3%) in the second validation. These populations were therefore not representative of the usual context of use of NITs.

### Impact of histological definitions in population-2

A total of 1081 patients with metabolic risk factors were included, 1009 patients with MLD and 72 (6.6%) without

MLD (A0S0), according to FLIP-CRN definition (Fig. 1 and Table 2) (see Supplementary Tables S4, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

We observed a marked risk of false negatives using the original FLIP-CRN definition compared with the H-NASHs (Table 1). Thirty-nine out of 1081 cases [3.6%; 95% confidence interval (CI): 2.6–4.9] who had significant SAF-activity grade of at least 2 were not considered to have NASH, despite careful exclusion of other causes of liver disease and the presence of metabolic factors. The stage of fibrosis was significant in fifteen of these cases including five with F3 and four with cirrhosis. It is interesting to note that most of these 39 cases were not diagnosed as MLD, because of the requirement of both ballooning and lobular inflammation ( $n=34$ ), but not the requirement of steatosis of more than or equal to 5% ( $n=5$ ). One of the 39 cases with minimal steatosis (1–4%) had cirrhosis and 10 had F1.

The primary impact of the definition of steatosis ( $\geq 5$  vs.  $\geq 1$ ) was the arbitrary change of 39 (3.6%) cases considered to be 'controls' to the 'disease' category and therefore an increase from 91.7% (89.9–93.3) to 95.3% (93.8–98.5), a significant difference ( $P=0.0007$ ). The prevalence of fibrosis ( $\geq F1$ ) increased from 64.8 to 67.3%, for a significant difference of 2.9% ( $P<0.0001$ ) (see Supplementary Tables S4, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

### Impact of histological definitions on accuracy

The NITs had significant accuracies for the prediction of NASH and steatosis-alone for all the 48 combinations (see Supplementary Table S5, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

However, results varied markedly ( $P<0.0001$ ) from 0.634 (0.604–0.662) for the best NIT-algorithm (NIT-Algo#48) (Table 3) to 0.334 (0.306–0.363) for the worst algorithm (NIT-Algo#25) (see Supplementary Tables S6, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

The  $\kappa$  also varied ( $P<0.0001$ ) from 0.477 (0.378–0.482) to 0.140 (95% CI: 0.106–0.175), respectively.

### Analyses of differences between accuracy and $\kappa$ coefficient

Whatever the 48 combinations of variability factors, the accuracy rate and the  $\kappa$  coefficient were highly correlated (see Supplementary Figs S7, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

**Table 3.** Highest accuracy obtained by Algo-48 (NIT-NASHs) out of 48 combinations

Presumed by NIT	Biopsy features classified using simplified definition (NASHs)			
	No steatosis	Steatosis-only	NASH	Total
No steatosis	8	36	16	60
Steatosis-only	26	282	139	447
NASH	17	162	395	574
Prevalence [ $n$ (%)]	51 (4.7)	480 (44.4)	550 (50.9)	1081 (100)

ActiTest cutoff was 0.17 and SteatoTest cutoff 0.30.

Accuracy = 0.634 (95% CI: 0.604–0.662), with  $\kappa = 0.349$  (95% CI: 0.293–0.406). CI, confidence interval; NIT, noninvasive tests.

However, this correlation was much higher when steatosis was not required for the diagnosis of NASH ( $P=0.0004$ ) and when the most sensitive SteatoTest cutoff (0.30) was chosen ( $P<0.0001$ ). Choosing steatosis definition of at least 5% instead of 1% was associated with a small increase in correlation ( $P=0.04$ ), and choosing the most sensitive ActiTest cutoff (0.17) did not impact the correlation ( $P=0.23$ ).

### Construction of a new NIT-NASHv0

The quantitative NIT-NASHv0 was therefore constructed according to the new simplified histological reference for NASH (H-NASHs). The area under the receiver operating characteristic for the diagnosis of NASHs was 0.796 (95% CI: 0.767–0.821), significantly higher ( $P=0.0001$ ) than the standard ActiTest (0.755; 0.725–0.782;  $n=1801$ ) (see Supplementary Fig. S8, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

### Discussion

Our results confirm that the diverse histological definitions of MLD cause significant variability in the estimated performance of NITs. We have shown that a simpler definition of NASH, which does not require the presence of steatosis and the presence of both lobular inflammation and ballooning, has permitted to construct a quantitative NIT-NASHv0, with significantly higher performance than the standard ActiTest.

### Strengths of the study

First, the review of landmark studies shows that, although the existing and arbitrary definitions of MLD may be appropriate for a pathologist's 'context of use', they are not appropriate for the usual context of use of a clinician [4,5,9,11,12]. The CRN definition of NASH is based on two studies including 2.3 and 0.3%, respectively, of A0S0 cases, which are the appropriate controls to validly evaluate the specificity of the definition of NASH. The real prevalence of MLD in subjects with metabolic risk factors were unknown in these studies. We therefore suggest that 'chronic carriers of metabolic risk factors in the absence of other known liver disease' would be a more appropriate definition for use by clinicians, which does not exclude the presence or severity of MLD, steatosis, necroinflammatory activity, and fibrosis (SAF features), as ranked by severity. This should be the population of interest for the construction of NITs.

Second, our study revealed that, despite a seemingly straightforward definition of NASH, the NAFLD-CRN and the FLIP algorithm were not sensitive enough, representing four (15%) out of 27 possible combinations of steatosis, ballooning and lobular inflammation to cover the spectrum of MLD. The H-NASHr definition covered 18 (67%) out of 27 possible features' combinations, which reduced the risk of a false-negative diagnosis. A total of 39 high-risk cases (3.6%) were missed by the CRN-FLIP algorithm, including 15 cases with significant fibrosis (six F2, five F3 and four cirrhosis). Indeed, in a previous study, we observed a 20% difference in the degree of steatosis, an 18% difference in the prevalence of ballooning, and a 33%

difference in the prevalence of lobular inflammation, using simultaneous paired biopsies [7].

Third, in the absence of a perfect reference (gold standard) [7], we used a method without gold standard, on the basis of the identification of the best concordance obtained between all the combinations of the NITs and the histological reference [18]. We previously applied this methodology to identify true positive/true-negative cases in discordant cases between NITs and biopsy results [17], and between transient elastography and share-wave elastography [18,19]. Using this 'concordance method' we identified the histological definition of NASH with the lowest risk of false-positives/false-negatives on the basis of 48 possible combinations and the highest concordance with validated NITs. The best NITs' combinations used highly sensitive cutoff to predict a grade of significant activity, such as more than 0.17 for the ActiTest for SAF-activity-grade-of at least A2, compared with more than 0.52 for METAVIR-A2 in CHC (see Supplementary Table S3, Supplemental digital content 1, <http://links.lww.com/EJGH/A248>). This seems rational as grade-1 ballooning and lobular inflammation are less severe features than METAVIR grade-2, defined as moderate necrosis with inflammation. Therefore, we constructed the NIT-NASHs as a sensitive NIT for H-NASHs providing a high negative predictive value. This should reassure subjects with NIT-NASHr less than 0.50 in whom the risk of H-NASHr is very low.

Finally, this study shows that the performance of the new NIT constructed using a simpler histological definition was better for the prediction of NASH than validated NITs such as the ActiTest [14].

### Weaknesses of the study

An obvious limitation of the population in our study was the small sample size of A0S0 controls, 72 cases out of 1081 (6.6%). However, these cases represent the highest percentage of A0S0 controls ever analyzed, as compared with the CRN validations (2.3 and 0.3%) [4,9].

Moreover, the comparison of the new NITs to standard NITs was limited only to ActiTest, taken as a standard quantitative NIT for assessing necroinflammatory histological activity. However, this test was recently validated [14], and nonpatented scores, such as NAFLD-score, BARD or FIB4 scores, were not specifically constructed as quantitative NITs for histological activity, and were originally constructed for fibrosis staging.

Our integrated database was limited by median biopsy lengths ranging from 15 to 22 mm and because most of the biopsies in obese cases were performed during bariatric surgery, which may be less appropriate than intercostal liver biopsy. However, the same methodology was used as for the current definitions of MLD [7].

This study focused on NITs developed by several coauthors of the article who have an obvious conflict of interest. However, the other coauthors were totally independent, and they recruited the patients and performed the assay independently of the company as well as having full access to all data and analyses. Thus, other independent validations of new NITs are necessary.

Despite these limitations, this study has identified several possible methodological improvements to construct better NITs for the diagnosis of MLD.

### Remaining questions about metabolic liver disease definitions

The current definition of NAFLD ‘hepatic steatosis in the absence of other known liver disease’ is too restrictive because it implies, by definition, the presence of steatosis. Thus, all cases without steatosis will be classified as a non-NAFLD independently of the presence of inflammation or fibrosis. This definition is not appropriate mainly because even a 20 mm liver biopsy is not a perfect reference and has a high risk of false-positive/false-negatives scores. These evidence-based results should be taken into account to develop valid definitions of MLD. On this basis, the definition of MLD should be simplified, in particular, by not excluding steatosis between 1 and 4%, not requiring both ballooning and lobular inflammation for the diagnosis of NASH, and not requiring steatosis to define fibrosis or inflammation as features of MLD when all other causes are excluded.

Furthermore, although a requirement of steatosis and both ballooning and lobular inflammation may be appropriate for a pathologist to specifically define the paradigm of NASH, and a possible surrogate marker of severe MLD, there is no evidence that activity is a better surrogate marker of clinical severity than the progression of fibrosis, as observed in chronic viral hepatitis, independently of steatosis and activity. This notion was recently supported by studies showing that ‘steatosis-only’ can progress to NASH and to significant fibrosis [27,28]. Histologically, steatosis progressed by at least grade 1 in 19% and regressed by at least grade 1 in 33%, and ballooning progressed by at least grade 1 in 37% and regressed by the same amount in 16%. The sample population in these studies was small and also raised the question of the natural time-dependent variability of steatosis or inflammation in the same subject. This is another reason to simplify the definitions of MLD and exclude a dependence between SAF features.

The rationale of 5% of hepatocytes for the definition of steatosis seems arbitrary, as minimal activity grade and minimal fibrosis stage are already accepted as features of MLD in the current definitions. In our study only 39 (3.6%) of presumed NAFLDs had 1–4% steatosis, including only one case with significant activity (NASH), one case of cirrhosis and 10 F1s (Table 2). Therefore, it is possible that, in a larger population, the true prevalence of cases with 1–4% steatosis would be much higher.

We did not discuss the staging of fibrosis, as this main prognostic feature is not taken into account for the definition of NASH in CRN or FLIP scoring systems.

### Conclusion

This proof-of-concept study suggests that an understanding of the limitations of the standard definitions of MLD might facilitate the construction of better biomarkers. A simplified histological definition (H-NASHs) that provided the best agreement between a histological reference and validated NITs would be a safer reference for the construction of new NITs.

### Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Program (FP7/2007–2013) under grant agreement no. HEALTH-F2-2009-241762 for the project FLIP.

European Community’s Seventh Framework Program (FP7/2007–2013) grant agreement HEALTH-F2-2009-241762 for the project FLIP.

The members of the FLIP partners’ consortium, the FLIP Pathology consortium, the FibroFrance-CPAM group, and the FibroFrance-Obese group are listed in Supplementary Material S1 (Supplemental digital content 1, <http://links.lww.com/EJGH/A248>).

Trial registration number: USA-NCT01927133.

### Conflicts of interest

T.P. is the inventor of FibroTest/SteatoTest and the founder of BioPredictive, the company that markets these tests. Patents belong to the French Public Organization Assistance Publique-Hôpitaux de Paris. M.M., Y.N., O.D. are BioPredictive employees. For the remaining authors there are no conflicts of interest.

### References

- Sanyal AJ, Brunt EM, Kleiner DE, Brunt EM, Kleiner DE, Kowdley KV, *et al.* Endpoints and clinical trial design for nonalcoholic steatohepatitis. *Hepatology* 2011; 54:344–353.
- Castera L, Vilgrain V, Angulo P. Noninvasive evaluation of NAFLD. *Nat Rev Gastroenterol Hepatol* 2013; 10:666–675.
- European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016; 64:1388–1402.
- Kleiner DE, Brunt EM, van Natta M, Behling C, Contos MJ, Cummings OW, *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; 41:1313–1321.
- Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, *et al.* Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012; 56:1751–1759.
- Bedossa P. FLIP Pathology Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of non-alcoholic fatty liver disease. *Hepatology* 2014; 60:565–575.
- Ratzliff V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005; 128:1898–1906.
- Younossi ZM, Gramlich T, Liu YC, Matteoni C, Petrelli M, Goldblum J, *et al.* Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. *Mod Pathol* 1998; 11:560–565.
- Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA. NASH Clinical Research Network (CRN). Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* 2011; 53:810–820.
- Brunt EM. Nonalcoholic fatty liver disease: pros and cons of histologic systems of evaluation. *Int J Mol Sci* 2016; 13:17.
- Kessoku T, Ogawa Y, Yoneda M, Imajo K, Sumida Y, Eguchi Y, *et al.* Simple scoring system for predicting cirrhosis in nonalcoholic fatty liver disease. *World J Gastroenterol* 2014; 20:10108–10114.
- Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, *et al.* Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with non-alcoholic fatty liver disease. *Gastroenterology* 2015; 149:389–397.
- Poynard T, Munteanu M, Charlotte F, Perazzo H, Ngo Y, Deckmyn O, *et al.* Diagnostic performance of a new noninvasive test for NASH using simplified histological reference. (AASLD late breaker 2016) Submitted. [In press].
- Munteanu M, Tiniakos D, Anstee Q, Charlotte F, Marchesini G, Bugianesi E, *et al.* Diagnostic performance of FibroTest, SteatoTest,

- and ActiTest in patients with NAFLD using the SAF-score as histological reference. *Aliment Pharmacol Ther* 2016; 44:877–889.
- 15 Boursier J, de Ledinghen V, Poynard T, Guécho 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.
  - 16 Poynard T, Ratziu V, Naveau S, Thabut D, Charlotte F, Messous D, *et al.* The diagnostic value of biomarkers (SteatoTest) for the prediction of liver steatosis. *Comp Hepatol* 2005; 4:10.
  - 17 Poynard T, Munteanu M, Imbert-Bismut F, Charlotte F, Thabut D, Le Calvez S, *et al.* Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. *Clin Chem* 2004; 50:1344–1355.
  - 18 Poynard T, Munteanu M, Luckina E, Perazzo H, Ngo Y, Royer L, *et al.* Liver fibrosis evaluation using real-time shear wave elastography: applicability and diagnostic performance using methods without a gold standard. *J Hepatol* 2013; 58:928–935.
  - 19 Poynard T, Pham T, Perazzo H, Munteanu M, Luckina E, Elaribi D, *et al.* Real-time shear wave versus transient elastography for predicting fibrosis: applicability, and impact of inflammation and steatosis. A non-invasive comparison. *PLoS One* 2016; 11:e0163276.
  - 20 Poynard T, Lebray P, Ingiliz P, Varaut A, Varsat B, Ngo Y, *et al.* Prevalence of liver fibrosis and risk factors in a general population using non-invasive biomarkers. *BMC Gastroenterol* 2010; 10:40.
  - 21 Houot M, Ngo Y, Munteanu M, Marque S, Poynard T. Systematic review with meta-analysis: direct comparisons of biomarkers for the diagnosis of fibrosis in chronic hepatitis C and B. *Aliment Pharmacol Ther* 2016; 43:16–29.
  - 22 Chou R, Wasson N. Blood tests to diagnose fibrosis or cirrhosis in patients with chronic hepatitis C virus infection: a systematic review. *Ann Intern Med* 2013; 158:807–820.
  - 23 Poynard T, Munteanu M, Ngo Y, Castera L, Halfon P, Ratziu V, *et al.* ActiTest accuracy for the assessment of histological activity grades in patients with chronic hepatitis C, an overview using Obuchowski measure. *Gastroenterol Clin Biol* 2010; 34:388–396.
  - 24 Poynard T, Munteanu M, Deckmyn O, Ngo Y, Drane F, Messous D, *et al.* Applicability and precautions of use of liver injury biomarker FT. A reappraisal at 7 years of age. *BMC Gastroenterol* 2011; 11:39.
  - 25 Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43:551–558.
  - 26 Hintze J. NCSS 2013, LCC. Kaysville, Utah, USA. <http://www.ncss.com>.
  - 27 Pais R, Charlotte F, Fedchuk L, Bedossa P, Lebray P, Poynard T, *et al.* A systematic review of follow-up biopsies reveals disease progression in patients with non-alcoholic fatty liver. *J Hepatol* 2013; 59:550–556.
  - 28 McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol* 2015; 62:1148–1155.