

Validation of noninvasive biomarkers (FibroTest, SteatoTest, and NashTest) for prediction of liver injury in patients with morbid obesity

Guillaume Lassailly^{a,b}, Robert Caiazzo^{c,d}, Antoine Hollebecque^{a,b}, David Buob^e, Emmanuelle Leteurtre^e, Laurent Arnalsteen^{c,d}, Alexandre Louvet^{a,b}, Marie Pigeyre^f, Violeta Raverdy^d, H el ene Verkindt^d, Marie France Six^d, Carole Eberle^d, Alexandre Patrice^d, S ebastien Dharancy^{a,b}, Monique Romon^f, Fran ois Pattou^{c,d} and Philippe Mathurin^{a,b}

Background Liver biopsy is considered as the gold standard for assessing nonalcoholic fatty liver disease (NAFLD) histologic lesions in patients with morbid obesity. The aim of this study was to determine the diagnostic utility of noninvasive markers of fibrosis (FibroTest), steatosis (SteatoTest), and steatohepatitis (NashTest, ActiTest) in these patients.

Materials and methods Two hundred and eighty-eight patients presenting with interpretable baseline operative biopsy and biomarkers, in an ongoing prospective cohort of patients treated with bariatric surgery, were included. Histology (NAFLD activity score, or NAFLD scoring system) and biochemical measurements were centralized and blinded to other characteristics. The area under the receiver operating characteristic curves (AUROC), sensitivity, specificity, positive and negative predictive values were assessed. Weighted AUROC (Obuchowski method) was used to prevent multiple testings and a spectrum effect.

Results The prevalence of advanced fibrosis (bridging) was 6.9%, advanced steatosis (>33%) was 48%, and steatohepatitis was 6.9% (NAFLD scoring system >4). Weighted AUROCs of the tests were as follows (mean, 95% confidence interval, significance): FibroTest for

advanced fibrosis: 0.85, 0.83–0.87, $P < 0.0001$; SteatoTest for advanced steatosis: 0.81, 0.79–0.83, $P < 0.0001$; and ActiTest for steatohepatitis: 0.77, 0.73–0.81, $P < 0.0001$.

Conclusion In patients with morbid obesity, the diagnostic performances of the FibroTest, SteatoTest, and ActiTest were statistically significant, thereby possibly reducing the need for biopsy in this population. *Eur J Gastroenterol Hepatol* 23:499–506 © 2011 Wolters Kluwer Health | Lippincott Williams & Wilkins.

European Journal of Gastroenterology & Hepatology 2011, 23:499–506

Keywords: liver fibrosis, morbid obesity, noninvasive biomarkers, steatohepatitis

^aService des Maladies de l'Appareil Digestif, Universit e Lille2, ^bUnit e Inserm U 995, ^cUnit e Inserm U 859, ^dServices de Chirurgie Endocrinienne, ^ed'Anatomie Pathologique and ^fNutrition, CHRU de Lille, Universit e Lille2, France

Correspondence to Pr Philippe Mathurin, Service des Maladies de l'Appareil Digestif, Universit e Lille2, H opital Claude Huriez 1er  tage Est, Avenue Michel Polonovski, Lille 59037, France
Tel: +33 3 20 44 55 97; fax: +33 3 20 44 55 64;
e-mail: philippe.mathurin@chru-lille.fr

Guillaume Lassailly, Robert Caiazzo, and Antoine Hollebecque contributed equally to this study

Received 15 November 2010 Accepted 28 February 2011

Introduction

Severe obesity is associated with decreased life expectancy [1]. In terms of liver injury, severe obesity is implicated in development of nonalcoholic fatty liver disease (NAFLD) [2]. The conceptual model of NAFLD has been referred to as the '2-hit hypothesis', with accumulation of fat as the first hit and development of an inflammatory response in fatty liver as the second [3,4]. In the first hit, accumulation of fat in hepatocytes renders the liver more vulnerable to subsequent insults.

In obese patients, bariatric surgery induces weight loss [5], decreases cardiovascular risk factors [6], and improves long-term survival [7,8]. With regard to liver injury, the benefit/risk ratio of bariatric surgery has been re-evaluated by long-

term prospective studies. Liver failure was observed in obese patients treated with jejunio-ileal bypass, a historical procedure no longer in use [9]. Early improvement of liver injury is observed after surgery in patients treated with recent bariatric surgical procedures (biliointestinal bypass, gastric bypass, gastric band) [10–12]. Using liver biopsy, we performed a 5-year prospective study to evaluate fibrosis and steatohepatitis in 381 severely obese patients after bariatric surgery. Five years after surgery, levels of fibrosis had increased significantly, but 95.7% of patients maintained a fibrosis score of at least F1. The percentage of patients with steatosis decreased from 37.4% before surgery to 16%, while the NAFLD score fell from 1.97 to 1, and ballooning from 0.2 to 0.1. Inflammation remained unchanged. The percentage of patients with probable or

definite nonalcoholic steatohepatitis (NASH) decreased significantly over 5 years from 27.4 to 14.2% [13].

Noninvasive biomarkers of liver injury have been extensively validated in chronic viral hepatitis and, more recently, in patients with alcoholic fatty liver disease and NAFLD [14]. These biomarkers have not been specifically validated in patients with morbid obesity. The aim of this study was to assess five biomarkers of fibrosis [FibroTest (FT)], steatosis [SteatoTest (ST)], necrosis, and inflammation [alanine aminotransferase (ALT), ActiTest (AT), and NashTest (NT)] using baseline serum and biopsy of our ongoing prospective cohort of patients treated with bariatric surgery. This should lead to reduction in the need for repeated biopsies in the assessment of liver injury before surgery and in postoperative follow-up of these patients.

Materials and methods

Patients

Severely obese patients referred to our unit for evaluation in view of bariatric surgery were considered for inclusion [13]. To be eligible for the study, all patients had to have fulfilled the following criteria: (i) morbid obesity (BMI > 40 kg/m²) or severe obesity (BMI > 35 kg/m²), at least one comorbidity factor (arterial hypertension, diabetes mellitus) for at least 5 years, and resistance to medical treatment; (ii) absence of medical or psychological contraindications for bariatric surgery; (iii) absence of current excessive drinking, as defined by average daily consumption of alcohol of 20 g/day for women and 30 g/day for men, and no history of past excessive drinking for a period longer than 2 years at any time in the past 20 years; (iv) absence of long-term consumption of hepatotoxic drugs; and (v) negative screening for chronic liver diseases including negative testing for hepatitis B surface antigen and hepatitis C virus antibodies, and no evidence of genetic hemochromatosis.

Surgical methods

Details of surgical methods were given elsewhere [13]. In summary, from 1994 to 2001, only biliointestinal bypass and the gastric band (Lap-Band System, INAMED Health, Santa Barbara, California, USA) were proposed. After 2001, gastric bypass (partitioning of the upper stomach to create a small gastric pouch and gastrojejunostomy to re-establish gastrointestinal continuity) was performed. Starting in 2004, biliointestinal bypass was no longer performed and was completely replaced by gastric bypass. When surgery was planned, patients were free to choose the surgical procedure. Informed written consent was obtained from all patients, and the study was conducted in conformity with the Helsinki Declaration.

Clinical and biological data

The following clinical and biological features were assessed prospectively before and at 1 and 5 years after surgery:

weight, BMI, blood pressure, ALT, γ -glutamyl transferase (GGT), prothrombin time, platelets, serum triglyceride, cholesterolemia, fasting blood glucose, and fasting insulin. Diabetes, hypercholesterolemia, and hypertriglyceridemia were defined as follows: fasting blood glucose greater than 1.26 g/l, cholesterolemia greater than 2.4 g/l, and serum triglyceride greater than 1.5 g/l.

Histological study

Liver biopsies were classified by two pathologists (E.L. and D.B.) blinded to the order of the biopsies and clinical and biological data. Liver biopsies were performed during the operative procedure. Biopsies were routinely stained with hematoxylin and eosin and Masson's trichrome.

During the prospective part of the study, histological features were scored according to the same criteria as those used in the FT/AT [14,15], ST [16], and NT [17] validations of NAFLD, and those used in the NAFLD scoring system (NAS) [18,19].

Fibrosis was prospectively scored using a predetermined scoring system equivalent to the METAVIR scoring system [20] and used in the first FT validation in NAFLD [15]. Fibrosis was staged on a scale of 0–4: F0 = no fibrosis, F1 = portal fibrosis or perivenular fibrosis without septa, F2 = few septa, F3 = numerous septa without cirrhosis, and F4 = cirrhosis.

Steatosis was quantified by low-power to medium-power evaluation of parenchymal involvement by steatosis (percentage of steatosis). Steatosis was scored using the NAS from 0 to 3 with a four-grade scoring system from S0 to S3: S0 = no steatosis or less than 5%, S1 = 5–33%, S2 = greater than 33–66%, and S3 = greater than 66% [19].

NASH was classified using the NAS [19]. Biopsies were assigned at random to pathologists. NAS is defined as the unweighted sum of scores for steatosis (0–3), lobular inflammation (0–3), and ballooning (0–2), thus ranging from 0 to 8. Cases with NAS of 0–2 were not considered as NASH; in contrast, most cases with scores of 5 or greater were diagnosed as NASH. Cases with activity scores of 3 and 4 were considered as borderline (probable or possible) NASH [19].

Biomarker measurements

FT, ST, AT, and NT (Biopredictive, Paris, France; FibroSURE LabCorp, Burlington, North Carolina, USA) were determined as published earlier [14]. Published recommended pre-analytical and analytical procedures were used [14]. The FT included α 2-macroglobulin, apolipoprotein A1, haptoglobin, total bilirubin, and GGT, adjusted for age and sex. The AT included the same five components in addition to ALT. The ST and NT included the same six components as AT plus serum glucose, triglycerides, and cholesterol, adjusted for age, sex, and BMI.

FT, AT, and ST scores ranged from 0 to 1.00, with higher scores indicating greater probability of significant lesions.

Predetermined FT conversion for the METAVIR fibrosis stage scoring system was 0.00–0.27 for F0, greater than 0.27–0.48 for F1, greater than 0.48–0.58 for F2, greater than 0.58–0.74 for F3, greater than 0.74 for F4 [14].

Predetermined AT conversion for the METAVIR activity grade scoring system was 0.00–0.17 for A0, greater than 0.17–0.52 for A1, greater than 0.52–0.62 for A2, greater than 0.62 for A3 [21].

Predetermined ST conversion for steatosis grade was 0.00–0.57 for S0, greater than 0.57–0.69 for S1, and greater than 0.69–1.000 for S2–S3 [16].

The NT is a 3-category score for predicting three NAS categories: 0.25 indicates ‘No-NASH’, 0.50 indicates ‘Possible NASH’, and 0.75 indicates ‘NASH’ [17].

Patent references were as follows: FT–AT: United States Patent of Trademark Office: #6631330, ST: #2009011132, and NT: #20080145864. Components were assessed using preanalytical and analytical recommended methods: GGT, ALT, serum glucose, triglycerides, cholesterol, and total bilirubin were measured using a Hitachi 917 (Mannheim, Germany). Standard manufacturer algorithms were used to exclude a high-risk profile of false-negatives/false-positives [14].

Statistical analysis

To account for a spectrum effect [22,23] and prevent multiple testing risks [22], the primary endpoint for each biomarker performance was the Obuchowski measure. This measure is a multinomial version of the area under the receiver operating characteristic curve (AUROC). With N categories of gold standard outcome (histological fibrosis stage or activity grade) and AUROC_{st}, the estimate of the AUROC of diagnostic tests for differentiating between categories s and t , the Obuchowski measure is a weighted average of $N(N-1)/2$ with different AUROC_{st} corresponding to all pairwise comparisons between two of the N categories. Each pairwise comparison was weighted to take into account the distance between grades or stages (i.e. the number of units on the ordinal scale). A penalty function proportional to the difference in METAVIR units between grades was defined, the penalty function was 0.33 when the difference between stages was 1, 0.67 when the difference was 2, and 1 when the difference was 3. The Obuchowski measure can be interpreted as the probability that the noninvasive index will correctly rank two randomly chosen patient samples from different activity grades according to the weighting scheme, with a penalty for misclassifying patients [22]. Note that the overall Obuchowski measure is not equivalent to the usual AUROC, as measurements are weighted according to distance between stages.

Secondary outcomes consisted of the AUROC according to the standard definition of liver injury and predictive values using predetermined cutoffs as defined in validation of biomarkers in NAFLD. Advanced fibrosis, defined as being higher than F1 (patients F2, F3, or F4) was presumed when the FT result was greater than 0.48. Fibrosis (patients with fibrosis stage of at least F1) was presumed when the FT result was greater than 0.27. Advanced steatosis, defined by a NAS above S1 (or more than 33% steatosis, i.e. S2/S3) was presumed for a ST greater than 0.69. Stage S1 (NAS definition: steatosis between 5 and 33%) was defined as a ST greater than 0.38 [16]. For NASH the definition used consisted of NAS categories (with no NASH: NAS < 3, possible NAS as 3 or 4, and NASH as > 4) [19]. The categories were defined, respectively, by a NT equal to 0.25, 0.5, or 0.75. The precision of the AT for diagnosis of NASH (NAS > 4) was also studied. The diagnostic threshold of the AT used to predict NASH was that initially elaborated for evaluating the histological activity of viral hepatitis. A NASH (NAS > 4) was presumed when the AT was 0.29, that is, METAVIR activity stage A1. Possible NASH (NAS > 2) was defined as an AT greater than 0.17 (equivalent to stage A0). Sensitivity analysis of biomarker diagnoses was carried out in patients with diabetes versus patients without diabetes, as the risk of liver injury may be different in patients with diabetes.

Results

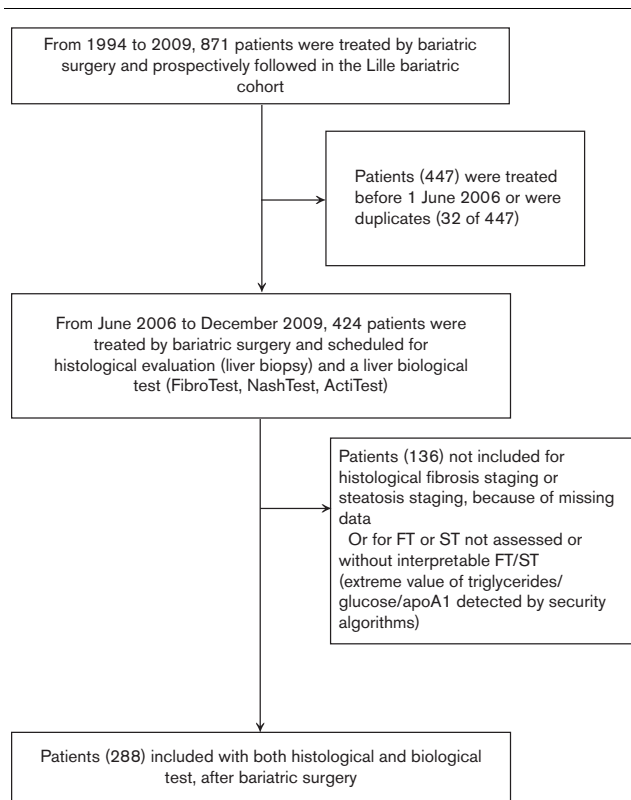
Patients included

Between 1994 and 2009, 871 patients were consecutively hospitalized for morbid obesity and were followed prospectively. From June 2006 to December 2009, biologic measures for FT and ST were routinely made. Four hundred and twenty-four patients were treated by a bariatric surgical procedure during that period. Two hundred and eighty-eight (67.9%) patients were included in the diagnostic study. One hundred and thirty-six patients were excluded for the following reasons: the biological collection sample lacked at least one biomarker ($n = 76$), patient renouncement ($n = 14$), extreme values detected by the security algorithm [$n = 3$, one of whom had triglyceride levels at 54 g/l, another with fasting glucose of 13.3 μ mol/l, and a third with apoA1 at 0.1 g/l]. The biological profile was considered as being at high risk of false-positives/false-negatives if the switch of one parameter from the usual median value induced variability in the biomarker (FT or ST) of greater than 0.30, if there was absence of liver tissue because of failure of the biopsy procedure ($n = 11$), or if the liver sample size was considered uninterpretable by our pathologists for histological scoring ($n = 32$) (Fig. 1).

Performance of FibroTest for diagnosis of fibrosis

Prevalence of advanced fibrosis was 6.9%. For diagnosis of advanced fibrosis defined using the F2, F3, F4, METAVIR-like definition, the AUROC mean (95% confidence

Fig. 1



Flow chart of the study. FT, FibroTest; ST, SteatoTest.

interval; significance vs. random) of FT was 0.82 (0.67–0.90; $P < 0.0001$) (Fig. 2a).

The Obuchowski measure was assessed after grouping F2, F3, and F4 because of the small sample size. FT overall weighted accuracy was 0.85 (0.83–0.87) (Table 1). Weighted accuracy was similar in patients with diabetes 0.82 (0.79–0.85) and in the overall group. FT values according to each stage are given in Fig. 3a.

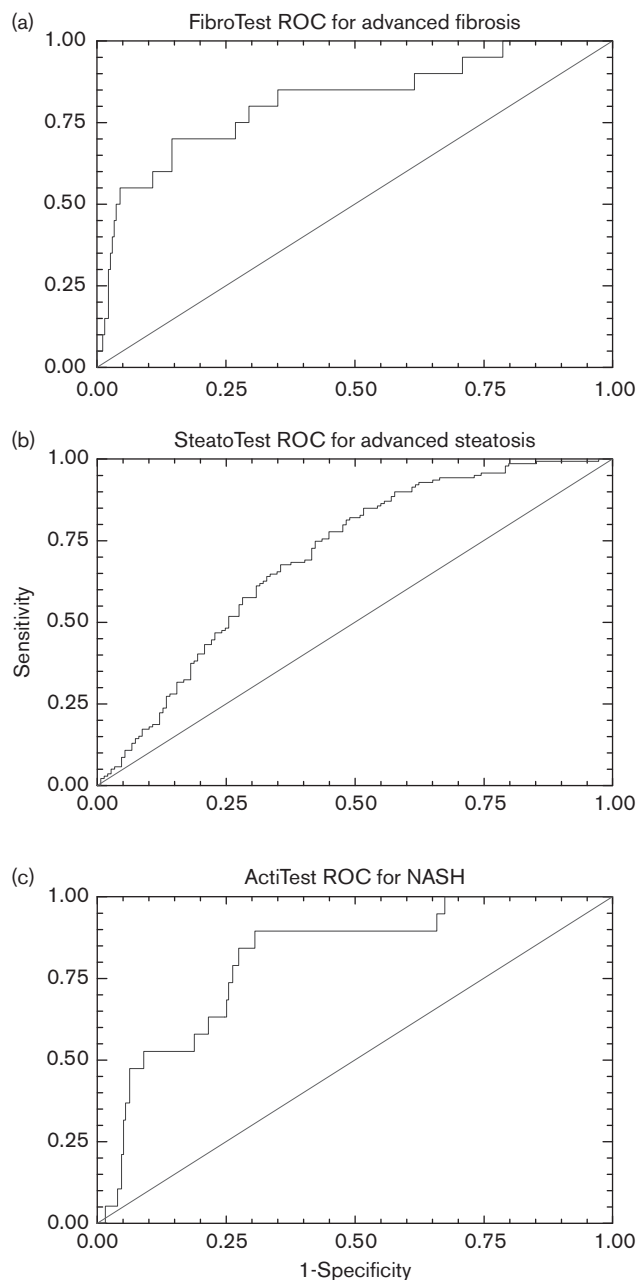
Performance of SteatoTest for diagnosis of steatosis

Advanced steatosis (> 33%) was present in 48.3% patients. For diagnosis of advanced steatosis, the AUROC of ST was 0.70 (0.63–0.75; $P < 0.0001$) (Fig. 2b). ST, overall, weighted accuracy for all steatosis grade pairwise comparisons (S0/S1/S2/S3) (Obuchowski measure) of ST was 0.81 (0.79–0.83); for ALT it was 0.74 (0.72–0.76; lower than ST: $Z = 2.3$, $P < 0.0001$) (Table 2). Weighted accuracy was similar in patients with diabetes 0.76 (0.70–0.82) and in the overall group. ST values according to each grade are given in Fig. 3b.

Performance of NashTest and ActiTest in diagnosis of NASH

NT was missing for 14 patients. The prevalence of NASH was 6.9%; for NASH or possible NASH it was 27.0%

Fig. 2



Receiver operating characteristic (ROC) curves of biomarkers for diagnosis of liver injury in patients with morbid obesity. (a) Advanced fibrosis defined as METAVIR F2F3F4 (significant fibrosis). (b) Advanced steatosis defined as steatosis NAS=S2S3 (percentage steatosis $\geq 33\%$). (c) Nonalcoholic steatohepatitis (NASH) defined as NAS greater than 4.

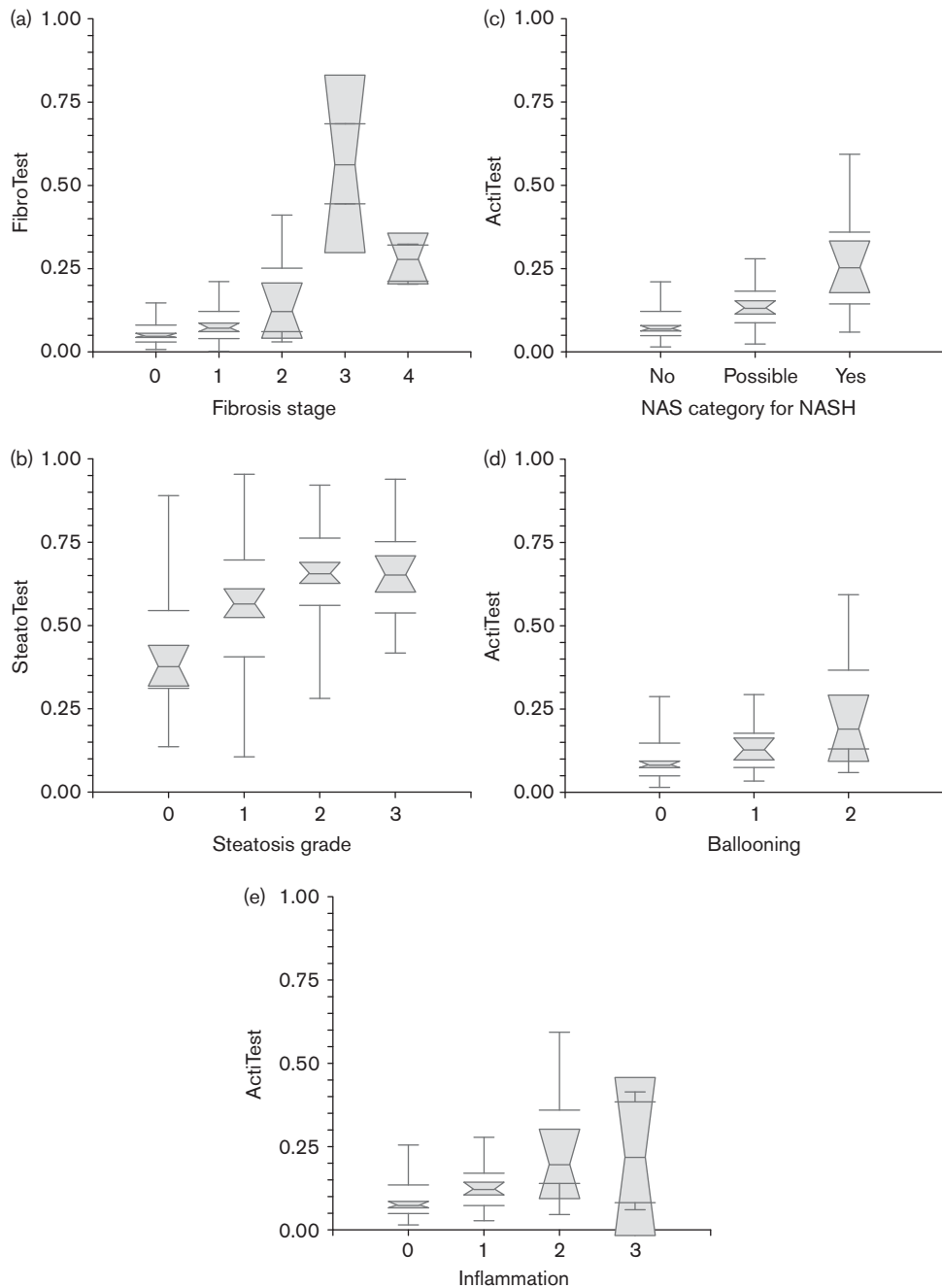
(Table 3). The concordance rate between histological NAS and that predicted by the NT was 43.1% ($P < 0.0001$), but with a weak κ -reliability test (0.14). Among 76 patients presumed to be no-NASH by NT, 68 were no-NASH, six were possible NASH, and two were NASH at biopsy; among 183 presumed possible-NASH by NT, 124 were no-NASH, 46 were possible NASH, and 13 were NASH at

Table 1 Accuracy (weighted area under the receiver operating characteristic curves) of the FibroTest for pairwise fibrosis stage diagnosis in 288 patients with morbid obesity

	F1 (n=98)	F2/F3/F4 (n=20)
F0 (n=170)		
FibroTest	0.525 (0.013)	0.693 (0.056)
F1 (n=98)		
FibroTest		0.621 (0.064)

Overall mean (standard error) accuracy of the FibroTest (Obuchowski measure) was 0.847 (0.006); not significant. Note that the overall Obuchowski measure is not equivalent to a usual area under the receiver operating characteristic curves as weighted according to distance between stages.

Fig. 3



Box plots of biomarkers according to liver injury. (a) FibroTest and METAVIR fibrosis stage. (b) SteatoTest and NAS steatosis grade. (c) ActiTest and NAS score for NASH. (d) ActiTest and NAS score for ballooning. (e) ActiTest and NAS score for inflammation.

Table 2 Accuracy (weighted area under the receiver operating characteristic curves) of the SteatoTest and alanine aminotransferase for pairwise steatosis grade diagnosis in 288 patients with morbid obesity

	S1 (n=113)	S2–S3 (n=139)
S0 (n=36)		
SteatoTest	0.657 (0.042)	0.795 (0.035)
Alanine aminotransferase	0.506 (0.028)	0.564 (0.028)
S1 (n=113)		
SteatoTest		0.535 (0.034)
Alanine aminotransferase		0.538 (0.023)

The overall mean (standard error) accuracy of the SteatoTest [Obuchowski measure, 0.807 (0.013)] was significantly greater than that of ALT [0.743 (0.008), $Z=4.7$, $P<0.0001$].

Note that the overall Obuchowski measure was not equivalent to the usual area under the receiver operating characteristic curves as weighted according to distance between stages.

Table 3 Comparison of characteristics of patients included or not included in the diagnostic study

Characteristics	Patients not included (n=136) ^a	Patients included in biomarker validation (n=288) ^a
Female sex: no (%)	104 (76.5%)	220 (76.4%)
Age (years): mean (SD)	41.2 (11.6)	41.6 (12.8)
BMI (kg/m ²): mean (SD)	49.4 (8)	48.6 (8.9)
Diabetes mellitus: no (%)	82/122 (32.8%)	92 (31.9%)
Arterial hypertension: no (%)	87/123 (70.7%)	174 (60.4%)
Dyslipidemia: no (%)	65/124 (47.6%)	167 (58.0%)
Cholesterolemia (mmol/l): mean (SD)	4.9 (0.9)	4.94 (0.88)
Serum triglycerides (mmol/l): mean (SD)	1.48 (0.75)	1.58 (0.88)
ALT (IU/l): mean (SD)	30.8 (17.8)	34 (23)
GGT (IU/l): mean (SD)	38.6 (31.2)	44 (48)
Fasting blood glucose (mmol/l): mean (SD)	6.2 (2.7)	6.2 (2.4)
Fibrosis F0/F1/F2/F3/ F4: no (%) (METAVIR scoring system)	39 (60%)/23 (35.4%)/1 (1.5%)/2 (3%)/0 (N=65) ^b	170 (59.0%)/98 (34.0%)/13 (4.5%)/2 (0.7%)/5 (1.7%)
Inflammation I0/I1/I2/I3 (Kleiner score)	56 (75.6)/13 (17.6%)/5 (6.7%)/(0%) (N=74) ^b	196 (71.3%)/64 (23.3%)/11 (4.0%)/4 (1.5%)
Ballooning B0/B1/B2: no (%) (Kleiner score)	65 (87.8%)/6 (8.1%)/3 (4%) (N=74) ^b	237 (86.2%)/24 (8.7%)/14 (5.1%)
Steatosis S0/S1/ S2–S3 (Kleiner score)	10 (13.1%)/34 (44.7%)/32 (42.1%) (N=76) ^b	36 (12.5%)/113 (39.2%)/139 (48.3%)
Extent of steatosis (%): mean (SD)	33.3 (26.1)	31.8 (25.0)
NAS (Kleiner score)		
0–2 No NASH	54 (73.9%)	199 (72.6%)
3–4 Possible	15 (20.5%)	56 (20.4%)
5–8 NASH	4 (5.4%)	19 (6.9%)

ALT, alanine aminotransferase; GGT, γ -glutamyl transferase; NAS, nonalcoholic fatty liver disease scoring system; NASH, nonalcoholic steatohepatitis.

^aSee for example Fig. 1.

^bValues for *n/N* patients (all histological values were not available for the 136 patients, not included).

biopsy; among 15 patients presumed to be NASH by NT, seven were no-NASH, four were possible NASH, and four were NASH at biopsy.

Table 4 Accuracy (weighted area under the receiver operating characteristic curves) of the ActiTest and alanine aminotransferase for pairwise nonalcoholic fatty liver disease scoring system category diagnosis in 288 patients with morbid obesity

	Possible NASH (n=98)	NASH (n=20)
No NASH (n=170)		
ActiTest	0.598 (0.033)	0.720 (0.057)
ALT	0.518 (0.025)	0.693 (0.058)
Possible NASH (n=98)		
ActiTest	0.598 (0.033)	0.634 (0.064)
ALT	0.518 (0.025)	0.667 (0.066)

The overall mean (standard error) accuracy of the ActiTest [Obuchowski measure, 0.774 (0.018)] was significantly greater than that of ALT [0.734 (0.016), $Z=2.7$, $P=0.006$].

Note that the overall Obuchowski measure was not equivalent to the usual area under the receiver operating characteristic curves as weighted according to distance between stages.

ALT, alanine aminotransferase; NASH, nonalcoholic steatohepatitis.

Using the quantitative biomarker AT, the AUROC were highly significant for diagnosis of NASH: AT = 0.81 (0.70–0.88; $P=0.002$) (Fig. 2c), and for diagnosis of possible-NASH or NASH = 0.74 (0.67–0.80; $P<0.0001$). Overall accuracy for all NAS category pairwise comparisons (No-NASH/possible-NASH/NASH) (Obuchowski measure) of AT was: 0.77 (0.73–0.81; $P<0.0001$); for ALT it was 0.74 (0.70–0.78; $P<0.0001$) lower than AT: $Z=2.7$, $P=0.006$. Details are given in Table 4. Weighted accuracy was similar in patients with diabetes 0.83 (0.79–0.87) and the overall group. AT values according to each NAS category are given in Fig. 3c.

Sensitivity analysis restricted to patients with high-quality liver samples

We performed sensitivity analysis according to the quality of liver samples that was considered satisfactory using earlier published criteria [24,25]. In this sensitivity analysis, restricted to patients ($n=251$) with a high quality liver sample (≥ 10 mm or at least six portal tracts), the AUROC mean of FT [0.80 (0.64–0.90; $P<0.0001$] for diagnosis of advanced fibrosis, the AUROC mean of ST [0.7 (0.64–0.76), $P<0.0001$] for diagnosis of advanced steatosis, and the AUROC of AT [0.799 (0.68–0.88), $P=0.004$] for diagnosis of NASH remained highly significant.

Sensitivity, specificity and predictive values

Diagnostic values according to predetermined cutoffs are detailed in Table 5. For fibrosis, the positive predicted value (PPV) was 87.5% for diagnosis of fibrosis greater than F0 using the 0.27 cutoff; the negative predictive value (NPV) for fibrosis greater than F1 was 93.8% using the 0.48 cutoff.

For steatosis, the PPV of the ST was 92.4% for diagnosis of steatosis greater than S0 using the 0.38 cutoff; the NPV for steatosis greater than S1 was 59.3% using the 0.69 cutoff.

For steatohepatitis the NPV of the AT was 96.0% for diagnosis of NASH (NAS > 4) using the 0.29 cutoff; the

Table 5 Sensitivity, specificity, and predictive values of biomarkers according to predetermined cutoffs

Biomarker (cutoff)	Disease (prevalence; %)	Se (%)	NPV (%)	Sp (%)	PPV (%)
FibroTest (0.27)	>F0 (41.0)	14/118 (11.9) ^a	168/272 (61.8)	168/170 (98.8)	14/16 (87.5)
FibroTest (0.48)	>F1 (6.9)	1/20 (5.0)	267/286 (93.4)	267/268 (99.6)	1/2 (50.0)
SteatoTest (0.38)	>S0 (87.5)	219/252 (86.9)	18/51 (35.3)	18/36 (50.0)	219/237 (92.4)
SteatoTest (0.69)	>S1 33 (48.3)	58/139 (41.7)	118/199 (59.3)	118/149 (79.2)	58/89 (65.2)
ActiTest (0.29)	NAS>4 (6.9)	9/19 (47.4)	239/249 (96.0)	239/255 (93.7)	9/25 (36.0)
ActiTest (0.17)	NAS>2 (27.4)	29/75 (38.7)	167/213 (78.4)	167/199 (83.9)	29/61 (47.5)

NAS, nonalcoholic fatty liver disease scoring system; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity.

^aNumber of patients *n/N* and percentage.

PPV for possible/NASH or NASH (NAS > 2) was 47.5% using 0.17 cutoff.

Discussion

This independent study of patients with morbid obesity confirms the earlier observed accuracy of biomarkers in diagnosis of liver injury in patients with NAFLD.

Weighted accuracy was crucial for assessing these validations, as the spectrum of liver injury varies significantly between different populations of NAFLD [22].

FibroTest area under the receiver operating characteristic curves for fibrosis

Like other studies [4,12,13], this study also used the METAVIR Score for grading fibrosis, as there is high inter-observer agreement (κ approximately = 0.8) [26]. Currently, the Brunt Score, specifically developed in patients with NAFLD, is considered the standard. However, both scores enable classifying those patients according to the presence or absence of significant fibrosis.

Owing to a low prevalence of advanced fibrosis, performances that fell between stages F2, F3, and F4 could not be assessed in the present population. The performance of biomarkers must be validated in larger populations of obese patients so as to be able to compare the performance between stages F2, F3, and F4 and between the different NAS categories. As in other studies [11,27], we observed, in our morbid obesity population, a low prevalence of advanced fibrosis 6.9% versus 24 and 32% in the NT training and validation groups, respectively, which included many patients with NASH [17]; ST validation subgroups showed 23–50% advanced fibrosis, but included patients with viral hepatitis C and alcoholic liver disease [16]. Likewise, FT validation studies in NAFLD patients included 15–24% of fibrosis, but were carried out in patients with NASH [15]. Owing to the limited number of patients with advanced fibrosis, it was not possible in this study to compare diagnostic sensitivity between the advanced fibrosis stages.

For advanced fibrosis, the observed FT AUROC was 0.82 (95% CI 0.67–0.90) and the weighted AUROC was 0.85 (95% CI 0.83–0.87), which were no different than the FT AUROC observed in the initial validation study of the reference center and the multicenter validations: 0.86 (95% confidence interval = 0.77–0.91) and 0.75 (95% confidence

interval = 0.61–0.83), respectively [15]. These FT performances were also similar to those observed in FT meta-analyses of different liver diseases [14].

Other noninvasive biomarkers such as the Fibrometer warrant evaluation in obese patients with NAFLD. However, we were unable to evaluate this score as hyaluronic acid, one of its components, was not prospectively assessed in our patients.

SteatoTest area under the receiver operating characteristic curve for steatosis

For steatosis, the spectrum was in sharp contrast to that of fibrosis, with a low prevalence of patients without disease. Only 12.5% had no steatosis, or less than 5%, versus 24 and 32% in training and validation groups, respectively, in NT validation [17]; 23–50% among subgroups in ST validation [16]; and 15–24% among populations in FT validation in NAFLD patients [15]. It was therefore not possible to assess the accuracy of the ST between patients without steatosis and those with less than 5% steatosis. The ST was not designed for comparison between steatosis greater than 66 versus 33–66%. This study confirms that there was no significant difference between these two grades of steatosis (Fig. 3b).

For advanced steatosis, the observed AUROC of 0.70 (95% confidence interval = 0.63–0.75) and the weighted AUROC of 0.81 (95% confidence interval = 0.79–0.83) were also similar to the AUROC observed in the initial validation study, that is, 0.79 (95% confidence interval = 0.75–0.83) [16].

NashTest, ActiTest, and steatohepatitis

For NASH, the prevalence of overt NASH at biopsy using the NAS was low in this study, only 6.9 versus 17 and 36% in training and validation groups in NT validation, and 17 and 34% among populations in the FT validation in NAFLD patients. Therefore, the power of this study in validating agreement between NT and biopsy was low, with only 15 cases presumed by the NT and 19 at biopsy. We observed significant accuracy of the AT for diagnosis of overt NASH as well as for pairwise comparison between NAS categories. The AT had initially been designed for necroinflammatory histological activity diagnosis in chronic hepatitis C and B. Indeed, it will be interesting to determine the AT performance in

patients with morbid obesity and in other NAFLD populations. AT included ALT in its formula, but had higher accuracy than ALT alone 'for diagnosis of the NAS category (Table 4)' as shown earlier in patients with chronic hepatitis C [21].

One limitation of liver biomarker validation is the absence of a perfect gold standard. Liver biopsy, even when the length is greater than 25 mm, also has a high percentage of false-positives/false-negatives [28].

Conclusion

Despite these limitations, independent validation of FT, ST, and NT in patients with morbid obesity should reduce the need for liver biopsy in these patients. The accuracy of AT as a biomarker of NASH and ballooning must be further investigated. Studies should now determine whether use of these biomarkers facilitates the evaluation of the impact of bariatric surgery and other treatments for morbid obesity.

Acknowledgements

This work was supported by the French Ministry of Health (Programme Hospitalier de Recherche Clinique), and financial support was provided by OSEO in the project IT-DIAB.

Conflicts of interest: none declared.

References

- Adams KF, Schatzkin A, Harris TB, Kipnis V, Mouw T, Ballard-Barbash R, et al. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med* 2006; **355**:763–778.
- Teli MR, Day CP, Burt AD, Bennett MK, James OF. Determinants of progression to cirrhosis or fibrosis in pure alcoholic fatty liver. *Lancet* 1995; **346**:987–990.
- Day CP, James OF. Steatohepatitis: a tale of two 'hits'? *Gastroenterology* 1998; **114**:842–845.
- Ratziu V, Giral P, Charlotte F, Bruckert E, Thibault V, Theodorou I, et al. Liver fibrosis in overweight patients. *Gastroenterology* 2000; **118**:1117–1123.
- Maggard MA, Shugarman LR, Suttrop M, Maglione M, Sugerman HJ, Livingston EH, et al. Meta-analysis: surgical treatment of obesity. *Ann Intern Med* 2005; **142**:547–559.
- Dixon JB, O'Brien PE, Playfair J, Chapman L, Schachter LM, Skinner S, et al. Adjustable gastric banding and conventional therapy for type 2 diabetes: a randomized controlled trial. *JAMA* 2008; **299**:316–323.
- Adams TD, Gress RE, Smith SC, Halverson RC, Simper SC, Rosamond WD, et al. Long-term mortality after gastric bypass surgery. *N Engl J Med* 2007; **357**:753–761.
- Sjostrom L, Narbro K, Sjostrom CD, Karason K, Larsson B, Wedel H, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007; **357**:741–752.
- Castillo J, Fabrega E, Escalante CF, Sanjuan JC, Herrera L, Hernanz F, et al. Liver transplantation in a case of steatohepatitis and subacute hepatic failure after biliopancreatic diversion for morbid obesity. *Obes Surg* 2001; **11**:640–642.
- Angulo P. NAFLD, obesity, and bariatric surgery. *Gastroenterology* 2006; **130**:1848–1852.
- Dixon JB, Bhathal PS, O'Brien PE. Nonalcoholic fatty liver disease: predictors of nonalcoholic steatohepatitis and liver fibrosis in the severely obese. *Gastroenterology* 2001; **121**:91–100.
- Mathurin P, Gonzalez F, Kerdraon O, Leteurtre E, Arnalsteen L, Hollebecque A, et al. The evolution of severe steatosis after bariatric surgery is related to insulin resistance. *Gastroenterology* 2006; **130**:1617–1624.
- Mathurin P, Hollebecque A, Arnalsteen L, Buob D, Leteurtre E, Caiazzo R, et al. Prospective study of the long-term effects of bariatric surgery on liver injury in patients without advanced disease. *Gastroenterology* 2009; **137**:532–540.
- Halfon P, Munteanu M, Poynard T. FibroTest-ActiTest as a non-invasive marker of liver fibrosis. *Gastroenterol Clin Biol* 2008; **32**:22–39.
- Ratziu V, Massard J, Charlotte F, Messous D, Imbert-Bismut F, Bonyhay L, et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2006; **6**:6.
- 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.
- Poynard T, Ratziu V, Charlotte F, Messous D, Munteanu M, Imbert-Bismut F, et al. Diagnostic value of biochemical markers (NashTest) for the prediction of non alcoholic steato hepatitis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2006; **6**:34.
- Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol* 1999; **94**:2467–2474.
- 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.
- Poynard T, Mathurin P, Lai CL, Guyader D, Poupon R, Tainturier MH, et al. A comparison of fibrosis progression in chronic liver diseases. *J Hepatol* 2003; **38**:257–265.
- 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.
- Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008; **54**:1372–1378.
- Poynard T, Halfon P, Castera L, Munteanu M, Imbert-Bismut F, Ratziu V, et al. Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; **53**:1615–1622.
- Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med* 2001; **344**:495–500.
- Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poynard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001; **357**:1069–1075.
- The METAVIR cooperative group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology* 1994; **20**:15–20.
- Merriman RB, Ferrell LD, Patti MG, Weston SR, Pabst MS, Aouizerat BE, Bass NM. Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology* 2006; **44**:874–880.
- Ratziu 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.