OFFICIAL JOURNAL OF THE INTERNATIONAL ASSOCIATION FOR THE STUDY OF THE LIVER

Liver International ISSN 1478-3223

ORIGINAL ARTICLE

# The combination of liver stiffness measurement and NAFLD fibrosis score improves the noninvasive diagnostic accuracy for severe liver fibrosis in patients with nonalcoholic fatty liver disease

Salvatore Petta[1,†], Ester Vanni[2,†], Elisabetta Bugianesi[2], Vito Di Marco[2], Calogero Cammà[1], Daniela Cabibi[3], Lavinia Mezzabotta[2] and Antonio Craxì[1]

1 Sezione di Gastroenterologia, DiBiMIS, University of Palermo, Palermo, Italy
2 Division of Gastro-Hepatology, Department of Internal Medicine, San Giovanni Battista Hospital, University of Torino, Torino, Italy
3 Cattedra di Anatomia Patologica, University of Palermo, Palermo, Italy

**Keywords**
LSM – NAFLD – NFS

**Abbreviations**
LSM, liver stiffness measurement; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NFS, NAFLD fibrosis score.

**Correspondence**
Salvatore Petta, MD PhD
Sezione di Gastroenterologia Di.Bi.M.I.S
AOUP Policlinico Paolo Giaccone
Piazza delle Cliniche 2 90127 Palermo, Italy
Tel: +39 091 655 2145
Fax: +39 091 655 2156
e-mails: salvatore.petta@unipa.it; petsa@inwind.it

Received 4 December 2013
Accepted 26 April 2014

DOI:10.1111/liv.12584

## Abstract

*Background & Aims:* The accuracy of noninvasive tools for the diagnosis of severe fibrosis in patients with nonalcoholic fatty liver disease(NAFLD) in clinical practice is still limited. We aimed at assessing the diagnostic performance of combined noninvasive tools in two independent cohorts of Italian NAFLD patients. *Methods:* We analysed data from 321 Italian patients(179 Sicilian-training cohort, and 142 northern Italy-validation cohort) with an histological diagnosis of NAFLD. Severe fibrosis was defined as fibrosis $\geq$ F3 according to Kleiner classification. The APRI, AST/ALT, BARD, FIB-4, and NFS scores were calculated according to published algorithms. Liver stiffness measurement(LSM) was performed by FibroScan. Cut-off points of LSM, NFS and FIB-4 for rule-in or rule-out F3-F4 fibrosis were calculated by the reported formulas. *Results:* In the Sicilian cohort AUCs of LSM, NFS, FIB-4, LSM plus NFS, LSM plus FIB-4, and NFS plus FIB-4 were 0.857, 0.803, 0.790, 0.878, 0.888 and 0.807, respectively, while in the northern Italy cohort the corresponding AUCs were 0.848, 0.730, 0.703, 0.844, 0.850, and 0.733 respectively. In the training cohort, the combination of LSM plus NFS was the best performing strategy, providing false positive, false negative and uncertainty area rates of 0%,1.1% and 48% respectively. Similar results were obtained in the validation cohort with false positive, false negative and uncertainty area rates of 0%,7.3% and 40.8%. *Conclusions:* The combination of LSM with NFS, two complementary, easy-to-perform, and widely available tools, is able to accurately diagnose or exclude the presence of severe liver fibrosis, also reducing of about 50–60% the number of needed diagnostic liver biopsies.

Nonalcoholic fatty liver disease (NAFLD), the hepatic manifestation of insulin resistance (IR), is a growing cause of chronic liver disease worldwide with a prevalence of about 20–30% in the general population (1). The clinical relevance of NAFLD stems from the evidence that a relevant proportion of patients, especially those with nonalcoholic steatohepatitis (NASH), may develop cirrhosis and its complications (1, 2). The most significant risk factors for disease progression are the degree of liver damage and of metabolic impairment, and a recent revision (3) clearly underlined that baseline severity of liver fibrosis is the strongest determinant of liver and non-liver-related mortality in NAFLD patients. Liver biopsy, albeit invasive, painful and with potentially life-threatening complications (4), is currently the gold standard for diagnosing advanced fibrosis in patients with NAFLD but is impractical for widespread use as a prognostic tool. In the last years the need for a robust and economical population-based screening to identify individuals with advanced fibrosis has led to the development of several noninvasive, biochemical and instrumental tools, but results are contrasting and large areas of diagnostic uncertainty are left (5–7). Specifically, AST/ALT ratio, APRI, BARD, FIB-4 and NAFLD fibrosis score (NFS) are easy and

[†]Both these authors contributed equally to this study.

quite reliable scores to be used in clinical practice (8–12), (13) while liver stiffness measurement (LSM) using transient elastography (TE) is probably the most validated tool to assess the stage of fibrosis not only in CHC and NAFLD patients (14–19). The main target of all the afore-mentioned noninvasive tools should be the accurate prediction of severe liver fibrosis with low rates of false positive and false negative results, thus reducing the need for diagnostic liver biopsy, but each of them bears potential limitations. Our aim was to assess the performance of combined noninvasive tools for identifying advanced fibrosis in a cohort of consecutive patients with biopsy-proven NAFLD from Sicily and to validate the results in another independent cohort from northern Italy.

## Materials and methods

### Patients

We retrospectively analysed data from Italian patients with a clinical and histological diagnosis of NAFLD, arising from two referral centres for liver diseases. The study cohort included 205 patients prospectively recruited at the Gastrointestinal & Liver Unit of the Palermo University Hospital (training set), and 220 patients recruited at the Gastro-hepatology Division of the University Hospital Torino (validation set). 26 patients (12.7%) from the Sicilian, and 52 (23%) from the northern Italy cohort were failure to a valid LSM because of obesity or unavailable results. Accordingly, all analyses were performed on 321 patients (179 from Sicily, and 142 from northern Italy) with complete biochemical data and LSM availability.

Other causes of liver disease were ruled out, including alcohol intake (>20 g/day) evaluated by a questionnaire, viral and autoimmune hepatitis, hereditary haemochromatosis, and alpha1-antitrypsin deficiency. Patients with advanced cirrhosis, hepatocellular carcinoma, and current use of steatosis inducing drugs were excluded.

This study was carried out in accordance with the principles of the Helsinki Declaration and its appendices, and with local and national laws. Approval was obtained from the hospital Internal Review Boards and their Ethics Committees, and written informed consent for the study was got from all the patients.

### Clinical, laboratory assessment, and histology

Clinical and anthropometric data, including BMI, the presence of arterial hypertension, impaired fasting glucose (IFG), and type 2 diabetes, were collected at the time of enrolment. On the same day of liver biopsy, a 12-hour overnight fasting blood sample was drawn to determine serum levels of AST, ALT, PLT, albumin, total, HDL and LDL-cholesterol, triglycerides, plasma glucose concentration and insulin. Insulin resistance was calculated by the homoeostasis model assessment (HOMA) (20).

Histological slides were coded and read at each clinical centre by one expert pathologist, who was unaware of patients' identity and history. A minimum 15 mm-length of the biopsy specimen or the presence of at least ten complete portal tracts was required (21). Steatosis was assessed as the percentage of hepatocytes containing fat droplets (minimum 5%), and evaluated as a continuous variable. Kleiner classification (22) was used to compute steatosis and lobular inflammation, and to stage fibrosis from 0 to 4.

### Noninvasive fibrosis algorithms/tools

The APRI (AST, PLT), AST/ALT ratio (AST, ALT), BARD (BMI, AST, ALT, Diabetes), FIB-4 (age, AST, ALT, PLT) and NFS (age, IFG/Diabetes, BMI, PLT, albumin, AST/ALT) were calculated using the original reported formulas (8–12).

Transient elastography was performed by FibroScan (Echosens, Paris, France) medical device using the M probe. LSM was assessed on the same day of liver biopsy, before the procedure, by one expert operator who had previously done at least 100 determinations in patients with chronic liver disease. LSM was expressed in kilopascal (kPa) and calculated as the median value of ten successful acquisitions, defined by a success rate of at least 60%, and by an interquartile range lower than 30% as previously described (23) and as suggested by the manufacturing company.

### Statistics

Continuous variables were summarized as mean ± SD, and categorical variables as frequency and percentage.

The accuracy of each score and of their combinations for detection of severe fibrosis (F3–F4) was assessed using receiver operator characteristic curves described as AUC with 95% confidence intervals (95% CI). A patient was assessed as positive or negative according to whether the noninvasive marker value was greater than, less than, or equal to a given cut-off value. Connected with any cut-off value is the probability of a true positive (sensitivity) and the probability of a true negative (specificity). AUCs are compared using DeLongs test. Cut-off points of LSM, NFS and FIB-4 for the F3–F4 model were derived from literature. Specifically, for LSM, cut-offs of <7.9 KPa and of ≥9.6 KPa were used to rule-out and rule-in, respectively, severe fibrosis (15); for NFS, cut-offs of <−1.455 and of >0.676 were used to rule-out and rule-in, respectively, severe fibrosis (8); and for FIB-4, cut-offs of <1.30 KPa and of >2.67 were used to rule-out and rule-in, respectively, severe fibrosis (7). Accordingly, false negative and false positive rates of the single test, and of their combination, as well as sensitivity, specificity, positive likelihood ratio (LHR),

negative LHR, positive predictive value (PPV), and negative predictive value (NPV) are calculated.

Analysis was performed by SPSS 18.0.3, and by MEDCALC.

## Results

### Patient characteristics and histology

The baseline features of the 179 Sicilian (training set) and of the 142 northern Italian (validation set) NAFLD patients are shown in Table 1. Aminotransaminase levels tended to be mildly elevated with 60.9% in the training, and 64.1% in the validation cohorts; ALT was generally less than twice the upper limit of normal, and within the normal range in 12.8% and 20.4% of the cases respectively. The patient characteristics were slightly different between institutions (Table 1). Patients from Sicily were more likely to have obesity, hypertension and more severe lobular inflammation compared to northern Italian patients. The mean length of the liver fragments given for histology was 17 mm (range 15–31) for Sicilian patients, and 20 mm (range 15–40) for northern Italy patients. Histological staging was similar in the two cohorts and significant fibrosis (F3–4) was present in one out of five patients.

### Diagnostic performance of single noninvasive tools for the diagnosis of severe liver fibrosis

Figures 1A and B show the accuracy, in terms of AUC, of the different noninvasive tools to detect fibrosis ≥ F3 in both the training and validation cohorts. In the training cohort the three best performing noninvasive tools were LSM, NFS and FIB-4 with AUC values of 0.857, 0.803, and 0.790 respectively. Accordingly, The AUC value of LSM was significantly higher than those of BARD score ($P = 0.002$), APRI ($P = 0.003$), and AST/ALT ratio ($P < 0.001$), the AUC of NFS was significantly higher than those of BARD ($P = 0.02$), APRI ($P = 0.04$), AST/ALT ($P = 0.01$), and the AUC of FIB-4 significantly higher than those of APRI ($P = 0.007$) and

**Table 1.** Baseline demographic, laboratory, metabolic, and histological features of 321 Italian patients with nonalcoholic fatty liver disease

| Variable | Nonalcoholic fatty liver disease (Sicily *n* = 179) | Nonalcoholic fatty liver disease (northern Italy *n* = 142) | *P* value |
|---|---|---|---|
| Mean age – years | 45.4 ± 13.3 (18–72) | 43.9 ± 11.8 (18-–78) | 0.29 |
| Male gender | 121 (67.5) | 102 (71.8) | 0.41 |
| Mean body mass index – kg/m$^2$ | 29.3 ± 4.1 (18.8–41.6) | 27.4 ± 3.7 (15.7–37.4) | <0.001 |
| Body mass index – kg/m$^2$ | | | |
| <25 | 20 (11.1) | 34 (23.9) | |
| 25–29.9 | 84 (46.9) | 74 (52.2) | |
| ≥30 | 75 (41.8) | 34 (23.9) | 0.001 |
| Aspartate aminotransferase – IU/ml | 45.7 ± 32.5 (15–290) | 42.2 ± 24.4 (12–198) | 0.29 |
| Alanine Aminotransferase – IU/ml | 80.3 ± 53.6 (17–453) | 75.6 ± 45.8 (13–323) | 0.40 |
| Platelets – 10$^3$/mmc | 221.5 ± 61.5 (65–418) | 230.2 ± 70.0 (71–664) | 0.23 |
| Albumin – g/dl | 4.6 ± 0.3 (3.8–5.4) | 4.6 ± 0.4 (3.2–5.3) | 0.23 |
| Total bilirubin – mg/dl | 0.6 ± 0.3 (0.1–3.0) | 0.6 ± 0.3 (0.3–3.1) | 0.23 |
| INR | 0.9 ± 0.1 (0.8–2.0) | 0.9 ± 0.1 (0.8–2.0) | 0.23 |
| Arterial hypertension | 43 (24.0) | 17 (11.9) | 0.008 |
| Impaired fasting glucose | 49 (27.3) | 58 (40.8) | 0.01 |
| Type 2 diabetes | 35 (19.5) | 22 (15.4) | 0.34 |
| Cholesterol – mg/dl | 206.2 ± 47.6 (91–368) | 205.3 ± 46.9 (97–382) | 0.86 |
| HDL cholesterol – mg/dl | 49.5 ± 18.9 (18–175) | 49.0 ± 13.1 (24–113) | 0.79 |
| Triglycerides – mg/dl | 149.5 ± 83.1 (31–477) | 153.7 ± 104.4 (17–701) | 0.69 |
| Blood glucose – mg/dl | 98.5 ± 28.4 (64–307) | 99.2 ± 25.7 (70–245) | 0.83 |
| Insulin – μU/ml | 16.5 ± 10.3 (2.5–59.8) | 16.8 ± 17.6 (2–135) | 0.81 |
| HOMA Score | 4.20 ± 3.38 (0.40–21.5) | 4.38 ± 6.38 (0.36–53.5) | 0.74 |
| Liver stiffness measurement (KPa) | 9.0 ± 7.0 (3–63) | 8.2 ± 5.2 (3–38) | 0.24 |
| Histology | | | |
| Lobular inflammation | | | |
| 0–1 vs. 2–3 | 74 (41.3) | 7 (4.9) | <0.001 |
| Steatosis grade | | | |
| 1 (5–33%) | 63 (35.1) | 82 (58.7) | |
| 2 (>33–66%) | 64 (35.8) | 44 (31.7) | |
| 3 (>66%) | 52 (29.1) | 16 (9.6) | <0.001 |
| Stage of Fibrosis | | | |
| 3–4 | 41 (22.9) | 29 (20.4) | 0.59 |

Abbreviation: IU, international units; HOMA, homoeostasis model assessment; HDL, high density lipoprotein. Data are given as mean ± standard deviation (range), or as number of cases (%).

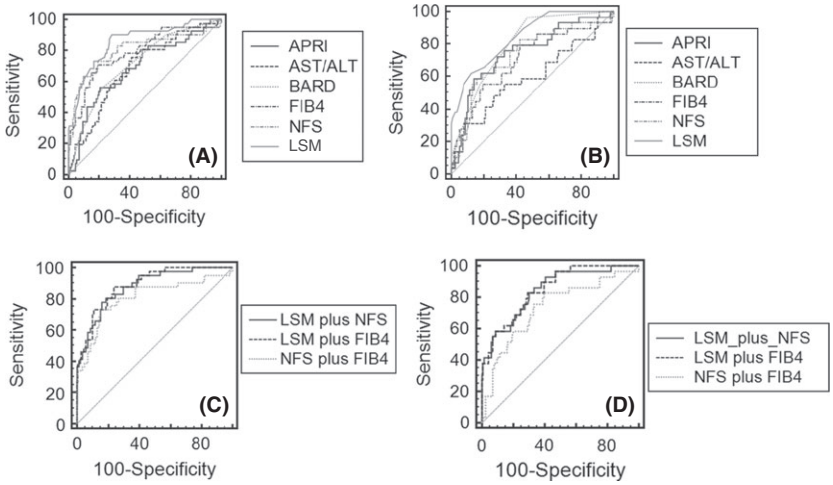


**Fig. 1.** Receiver operating characteristic (ROC) curve for singular and combined noninvasive tools on the basis of the presence of F3–F4 fibrosis. (A) ROC curve for APRI (AUC 0.688, 95% C.I. 0.593–0.783), AST/ALT (AUC 0.685, 95% C.I. 0.595–0.775), BARD (AUC 0.717, 95% C.I. 0.629–0.805), FIB4 (AUC 0.790, P5% C.I. 0.705–0.875), NFS (AUC 0.803, 95% C.I. 0.711–0.894), and liver stiffness measurement (LSM) (AUC 0.857, 95% C.I. 0.790–0.924) in the 179 Sicilian patients. The AUC value of LSM was significantly higher than those of BARD score ($P = 0.002$), APRI ($P = 0.003$), and AST/ALT ratio ($P < 0.001$); the AUC of NFS was significantly higher than those of BARD ($P = 0.02$), APRI ($P = 0.04$), AST/ALT ($P = 0.01$), and the AUC of FIB-4 significantly higher than those of APRI ($P = 0.007$) and AST/ALT ($P = 0.006$). (B) ROC curve for APRI (AUC 0.746, 95% C.I. 0.641–0.850), AST/ALT (AUC 0.597, 95% C.I. 0.472–0.723), BARD (AUC 0.803, 95% C.I. 0.723–0.882), FIB4 (AUC 0.703, 95% C.I. 0.589–0.816), NFS (AUC 0.730, 95% C.I. 0.621–0.838), and LSM (AUC 0.848, 95% C.I. 0.774–0.922) in the 142 northern Italy patients. The AUCs of LSM, NFS and FIB-4 were significantly higher than that of AST/ALT ratio ($P < 0.001$, $P = 0.03$, and $P = 0.03$ respectively), and the AUC of LSM was better than FIB-4 ($P = 0.03$). (C) ROC curve for LSM plus NFS (AUC 0.878, 95% C.I. 0.821–0.936), LSM plus FIB4 (AUC 0.888, 95% C.I. 0.836–0.941), and NFS plus FIB4 (AUC 0.807, 95% C.I. 0.719–0.895) in the 179 Sicilian patients. Both LSM plus NFS, and LSM plus FIB-4, were superior to NFS plus FIB-4 ($P = 0.03$ for both). LSM plus NFS had an AUC higher compared to FIB-4 ($P = 0.01$) and NFS ($P = 0.02$, as well as LSM plus FIB-4 compared to FIB-4 ($P = 0.006$) and to NFS ($P = 0.04$). (D) ROC curve for LSM plus NFS (AUC 0. 844, 95% C.I. 0.766–0.923), LSM plus FIB4 (AUC 0.850, 95% C.I. 0.777–0.922), and NFS plus FIB4 (AUC 0.733, 95% C.I.0.626–0.841) in the 142 northern Italy patients. Both LSM plus NFS, and LSM plus FIB-4, performed better than NFS ($P = 0.03$ for both), FIB-4 ($P = 0.03$ and $P = 0.02$ respectively), and NFS plus FIB-4 ($P = 0.04$ and $P = 0.03$ respectively).

AST/ALT ($P = 0.006$). In the replication cohort the AUC values of LSM, NFS and FIB-4 were 0.848, 0.730, and 0.703 respectively. In addition, their AUCs were significantly higher than that of AST/ALT ratio ($P < 0.001$, $P = 0.03$, and $P = 0.03$ respectively), and the AUC of LSM was better than FIB-4 ($P = 0.03$).

In view of the results in the training set, only LSM, FIB-4 and NFS scores were further analysed for diagnostic performance by using cut-offs previously reported in the literature to rule-in and rule-out F3-F4 fibrosis (7.9 KPa and 9.6 KPa for LSM, 1.30 and 2.67 for FIB-4, and −1.455 and 0.676 for NFS). Thus, in the training and validation cohorts, NFS provided the highest specificity (99.1% and 100% respectively), highest PPV (85.7% and 100% respectively) and highest positive LHR (37.3 and infinity respectively), while LSM had the highest sensitivity (85.3% and 68% respectively), highest NPV (95.2% and 91.2% respectively) and lowest negative LHR (0.2 and 0.3 respectively) (Table 2). In both the training and the validation cohorts, NFS had the lowest false positive rates (14.2% and 0% respectively), while LSM showed the lowest false negative rates (4.7% and 8.7% respectively) (Table 2). Interestingly, elevated ALT levels (ALT>100) (24) explained the 15.4% (2/13) and the 34.8% (8/23) of the false positive results of LSM in Sicilian and northern Italy cohorts respectively.

### Diagnostic performance of combined noninvasive tools for the diagnosis of severe liver fibrosis

We finally evaluated the performance of paired combinations of LSM, NFS and FIB-4 for the diagnosis of severe liver fibrosis (≥ F3) in both cohorts.

In the training cohort, the AUCs of the tested combinations were 0.878, 0.888 and 0.807 for LSM plus NFS, LSM plus FIB-4, and NFS plus FIB-4 respectively (Fig. 1C). Both LSM plus NFS, and LSM plus FIB-4, were superior to NFS plus FIB-4 ($P = 0.03$ for both). LSM plus NFS had an AUC higher compared to FIB-4 ($P = 0.01$) and NFS ($P = 0.02$), as well as LSM plus FIB-4 compared to FIB-4 ($P = 0.006$) and to NFS ($P = 0.04$). Similar results were obtained in the replication cohort, where LSM plus NFS, LSM plus FIB-4, and NFS plus FIB-4 had AUCs of 0.844, 0.850, and 0.745 respectively (Fig. 1D). Both LSM plus NFS, and LSM plus FIB-4, performed better than NFS ($P = 0.03$ for both), FIB-4 ($P = 0.03$ and $P = 0.02$ respectively), and NFS plus FIB-4 ($P = 0.04$ and $P = 0.03$ respectively).

**Table 2.** False positive rates, false negative rates, uncertainty area, sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive values and negative predictive values, of single or combined tools, according to rule-in and rule-out literature cut-offs, for the diagnosis of severe fibrosis in patients with nonalcoholic fatty liver disease

| Tools | ROC AUC (C.I.) | False positive (%) | False Negative (%) | Uncertainty area (%)* | Correctly classified (%) | Sensitivity (%) | Specificity (%) | Positive likelihood ratio | Negative likelihood ratio | Positive predictive value (%) | Negative predictive value (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sicilian cohort | | | | | | | | | | | |
| Fibroscan | 0.857 (0.790–0.924) | 23/52 (44.2) | 5/106 (4.7) | 21/179 (11.7) | 130/179 (72.6)† | 85.3 | 81.4 | 4.6 | 0.2 | 55.8 | 95.2 |
| FIB-4 | 0.790 (0.705–0.875) | 3/15 (20) | 12/123 (9.7) | 41/179 (22.9) | 123/179 (68.7) | 50 | 97.3 | 19 | 0.5 | 80 | 90.2 |
| NFS | 0.803 (0.711–0.894) | 1/7 (14.2) | 12/123 (9.7) | 49/179 (27.3) | 117/179 (65.3) | 33.3 | 99.1 | 37.3 | 0.7 | 85.7 | 90.2 |
| Fibroscan + FIB-4 | 0.888 (0.836–0.941) | 0/11 (0) | 2/86 (2.3) | 82/179 (45.8) | 95/179 (53.0) | 84.6 | 100 | Infinity | 0.2 | 100 | 97.6 |
| Fibroscan+NFS | 0.878 (0.821–0.936) | 0/5 (0) | 1/88 (1.1) | 86/179 (48) | 92/179 (51.3) | 83.3 | 100 | Infinity | 0.1 | 100 | 98.9 |
| FIB-4 + NFS | 0.807 (0.719–0.895) | 0/5 (0) | 9/109 (8.2) | 75/179 (41.8) | 105/179 (58.6) | 35.2 | 100 | Infinity | 0.7 | 100 | 91.3 |
| Northern Italy cohort | | | | | | | | | | | |
| Fibroscan | 0.848 (0.774–0.922) | 13/30 (43.3) | 8/91 (8.7) | 21/142 (14.7) | 100/142 (70.4)‡ | 68 | 86.4 | 5 | 0.3 | 56.7 | 91.2 |
| FIB-4 | 0.703 (0.589-0.816) | 4/9 (44.4) | 16/112 (14.2) | 21/142 (14.8) | 101/142 (71.1)§ | 23.8 | 96 | 5.9 | 0.8 | 55.5 | 85.7 |
| NFS | 0.730 (0.621–0.838) | 0/2 (0) | 13/109 (11.9) | 31/142 (21.8) | 98/142 (69.0)** | 13.3 | 100 | Infinity | 0.9 | 100 | 88 |
| Fibroscan + FIB-4 | 0.850 (0.777–0.922) | 2/7 (28.5) | 7/81 (8.6) | 54/142 (38) | 79/142 (55.6) | 41.6 | 97.3 | 15.8 | 0.6 | 71.4 | 91.3 |
| Fibroscan + NFS | 0.844 (0.766–0.923) | 0/2 (0) | 6/82 (7.3) | 58/142 (40.8) | 78/142 (54.9) | 25 | 100 | Infinity | 0.7 | 100 | 92.7 |
| FIB-4 + NFS | 0.733 (0.626–0.841) | 0/2 0 | 12/100 (12) | 40/142 (28.1) | 90/142 (63.3) | 14.3 | 100 | Infinity | 0.9 | 100 | 88 |

*Referred to patients with a value within rule-in and rule-out cut-offs.

†$P < 0.05$ compared to Fibroscan + FIB-4, Fibroscan + NFS and FIB-4 + NFS.

‡$P < 0.05$ compared to Fibroscan + FIB-4, Fibroscan + NFS.

§$P < 0.01$ compared to Fibroscan + FIB-4, Fibroscan + NFS.

**$P < 0.05$ compared to Fibroscan + FIB-4, Fibroscan + NFS.

In the training cohort, overall LSM plus NFS was the best performing combination strategy with specificity and PPV of 100%, and with sensitivity, NPV and negative LHR of 83.3%, 98.9% and 0.1 respectively (Table 2, upper panel). Similar results were observed in the validation cohort, where LSM plus NFS provided specificity and PPV of 100%, and sensitivity, NPV and negative LHR of 25%, 92.7% and 0.7% respectively (Table 2, bottom panel). Accordingly, both in the training and validation cohorts, LSM plus NFS had the lowest false positive (0% for both), and false negative rates (1.1% and 7.3% respectively) (Table 2 and Fig. 2).

## Discussion

Nonalcholic fatty liver disease, with an estimated prevalence around 25% in the general population (1), is becoming the leading cause of chronic liver diseases worldwide and its liver-related prognosis is finally decided by the amount of liver fibrosis accumulating over the years (3). The noninvasive identification of NAFLD patients at high risk for advanced liver disease who may benefit from more intensive management is still an unmet need. In this study, including NAFLD cohorts from two different centres as training and validation set respectively, we found that the combination of biochemical (NFS) with instrumental (LSM) noninvasive tools can significantly improve the diagnostic accuracy of severe liver fibrosis, by reducing the rates of false positive and particularly false negative results.

Similarly to other studies, we confirmed LSM and NFS, together with FIB-4 score, as the best performing single noninvasive tools for the diagnosis of severe liver fibrosis in NAFLD, with an acceptable/good diagnostic ability according to AUC values. However, when evaluating the performance of each score to rule-in or rule-out severe liver fibrosis, we observed high rates of false negative –for NFS and FIB-4-, and false positive –for LSM- results, in both cohorts. Specifically, lower false negative (up to 4.7%), but higher false positive (up to

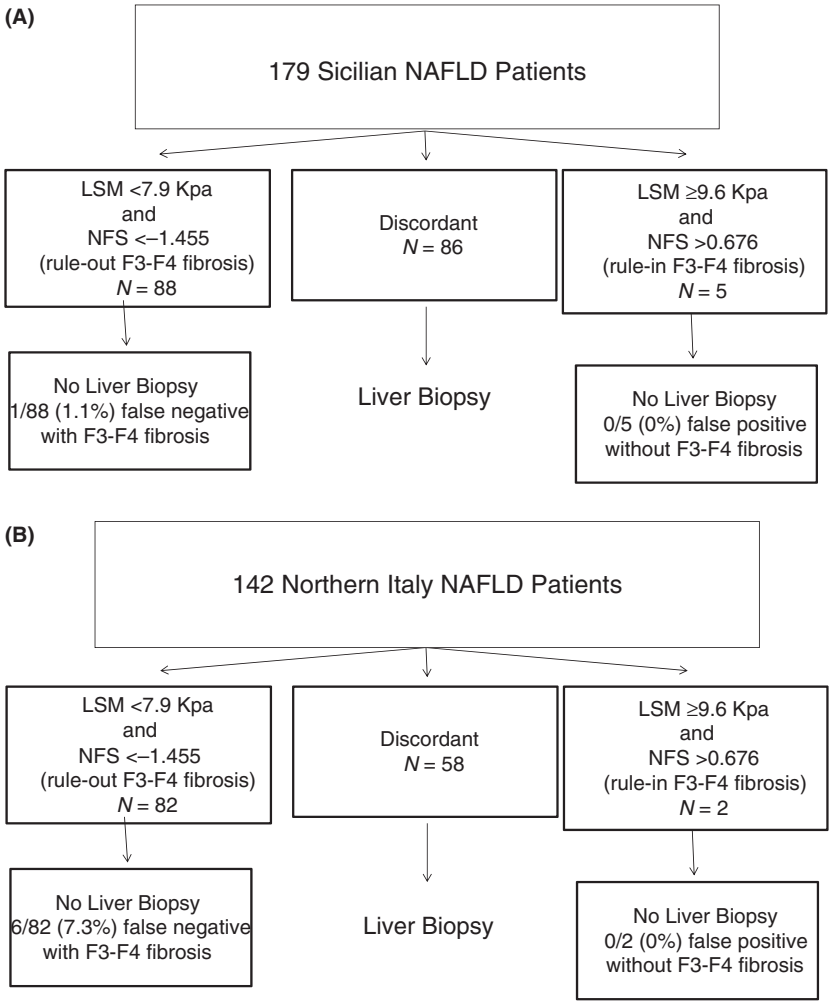


**Fig. 2.** False positive results, false negative results, and uncertainty area using combined noninvasive tools. (A) Liver stiffness measurement plus NFS for Sicilian cohort; (B) Liver stiffness measurement plus NFS for northern Italy cohort.

44.2%) rates were observed for LSM, while false positive rates for NFS were ranging from 0% in northern Italy to 14.2% in Sicilian cohort. Our data confirm the modest overall accuracy of the single noninvasive tools, and therefore the impossibility to rely on just one of them in clinical practice (5–17).

The applicability of single noninvasive tools is an important issue also in patients with chronic hepatitis C. In this clinical setting, recent studies showed that the combination of two noninvasive tools could help to generate algorithms with a higher diagnostic accuracy, thus limiting the number of liver biopsies (23, 25). The better performance of multiple vs. single tests is also confirmed in our study, particularly by the combination of LSM (an instrumental tool at low false positive results) with NFS (a clinical-biochemical tool at high false negative results), since this approach is able to reduce both false negative (range 1.1–7.3%) and false positive (0%) results, although increasing the number of patients in the 'uncertainty area'. The advantage of this approach is better evident in the Sicilian than in the northern Italy cohort, where false negative rates are only slightly improved. Whenever LSM would be nonreliable/available, then the combination of two clinical-biochemical tools (NFS and FIB-4) can reduce false positive results (at least in the Sicilian cohort), obtaining a similar false negative rate. The worse performance of combination strategies in northern Italy patients is probably because of the better performance of NFS in this cohort.

This study has limitations. First, it is aimed at optimizing the use of currently easy-to-perform and widely available noninvasive tools for NAFLD patients, even if, while clinical-biochemical scores are ever available, LSM could not be. In addition, we cannot rule out that the introduction of new techniques/tools such as ARFI (26), citokeratin-18 fragments serum levels (27), etc. could improve the performance of the proposed combination. Another methodological question is the potentially limited validity of the results in different populations and settings. Our study included two cohorts of Italian patients enrolled at tertiary care centres, and that are different in terms of prevalence of obesity, steatosis and liver inflammation. This issue could affect the interpretation of our results, even if the similar performance of the tested scores in the two population is a strength of our work. In any case our tested populations may be different from the majority of cases of NAFLD in the general population. Thus, it is plausible that the performance of the proposed algorithms could change according to the clinical, biochemical and metabolic characteristics of patients, and to the prevalence of severe fibrosis. In this line, to strength our study, in our analysis we used published standardized cut-offs and not cut-offs calculated from data of our populations, even if similar diagnostic performances were observed in both the two scenarios (data not shown). In addition, because of the complexity of NAFLD, the decision to perform the liver biopsy must take into account not only algorithms aimed to identify at risk patients, but also a complete evaluation of pros and contra of liver biopsy in the singular patient. Finally, the accuracy of liver biopsy is always a methodological issue in this kind of studies, related to sampling errors, technical processing of the specimens, and interobserver variability and both to the length of biopsy specimens and the number of portal spaces. However, we are confident that the minimum length of 15 mm and the presence of at least ten complete portal tracts minimize this bias.

In conclusion, we demonstrated that the combination of LSM with NFS, two complementary, easy-to-perform, and widely available tools, is able to accurately diagnose or exclude the presence of severe liver fibrosis, also reducing of about 50–60% the number of needed diagnostic liver biopsies. From a clinical standpoint, whenever LSM would be no reliable/available, we suggest to use the association between FIB-4 and NFS, characterized by an acceptable diagnostic performance. Therefore, to reduce the proportion of patients requiring liver biopsy to further stage the disease, it may be possible to develop an algorithm for investigation of patients with NAFLD that uses simple noninvasive tests as recently observed for NFS, APRI and FIB-4 in a large cohort of clinically diagnosed NAFLD patients (28). However, such an algorithm will need to be carefully evaluated in a prospective study.

## Acknowledgements

*Financial support*: This study was funded by grants from the FP7/2007–2013 (n°Health-F2-2009-241762), from PRIN 2009 ( Prot. N. 2009ARYX4T_004), and from PRIN 2010–2011 ( Prot. N. 2010C4JJWB).

*Conflict of interest*: The authors do not have any disclosures to report.

## References

1. Petta S, Muratore C, Craxì A. Non-alcoholic fatty liver disease pathogenesis: the present and the future. *Dig Liver Dis* 2009; **41**: 615–25.
2. Petta S, Craxì A. Hepatocellular carcinoma and non-alcoholic fatty liver disease: from a clinical to a molecular association. *Curr Pharm Des* 2010; **16**: 741–52.
3. Angulo P. Long-term mortality in nonalcoholic fatty liver disease: is liver histology of any prognostic significance? *Hepatology* 2010; **51**: 373–5.
4. Dienstag JL. The role of liver biopsy in chronic hepatitis C. *Hepatology* 2002; **36**: S152–60.
5. Petta S, Handberg A, Craxì A. Non invasive indexes for the assessment of patients with non-alcoholic fatty liver disease. *Curr Pharm Des* 2013; **19**: 5193–218.
6. Machado MV, Cortez-Pinto H. Non-invasive diagnosis of non-alcoholic fatty liver disease a critical appraisal. *J Hepatol* 2013; **58**: 1007–19.
7. Castera L, Vilgrain V, Angulo P. Noninvasive evaluation of NAFLD. *Nat Rev Gastroenterol Hepatol* 2013; **10**: 666–75.

8. McPherson S, Stewart SF, Henderson E, *et al.* Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut* 2010; **59**: 1265–9.
9. Shah AG, Lydecker A, Murray K, *et al.* Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009; **7**: 1104–12.
10. Angulo P, Hui JM, Marchesini G, *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007; **45**: 846–54.
11. Cales P, Laine F, Boursier J, *et al.* Comparison of blood tests for liver fibrosis specific or not to NAFLD. *J Hepatol* 2009; **50**: 165–73.
12. Harrison SA, Oliver D, Arnold HL, *et al.* Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut* 2008; **57**: 1441–7.
13. Musso G, Gambino R, Cassader M, *et al.* Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2011; **43**: 617–49.
14. Friedrich-Rust M, Ong MF, Martens S, *et al.* Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. *Gastroenterology* 2008; **134**: 960–74.
15. Wong VW, Vergniol J, Wong GL, *et al.* Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* 2010; **51**: 454–62.
16. Petta S, Di Marco V, Cammà C, *et al.* Reliability of liver stiffness measurement in non-alcoholic fatty liver disease: the effects of body mass index. *Aliment Pharmacol Ther* 2011; **33**: 1350–60.
17. Gaia S, Carenzi S, Barilli AL, *et al.* Reliability of transient elastography for the detection of fibrosis in non-alcoholic fatty liver disease and chronic viral hepatitis. *J Hepatol* 2011; **54**: 64–71.
18. Kohlhaas A, Durango E, Millonig G, *et al.* Transient elastography with the XL probe rapidly identifies patients with nonhepatic ascites. *Hepat Med* 2012; **1**: 11–8.
19. Myers RP, Pomier-Layrargues G, Kirsch R, *et al.* Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. *Hepatology* 2012; **55**: 199–208.
20. Matthews DR, Hosker JP, Rudenski AS, *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985; **28**: 412–9.
21. Colloredo G, Guido M, Sonzogni A, *et al.* Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. *J Hepatol* 2003; **39**: 239–44.
22. Kleiner DE, Brunt EM, Van Natta M, *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; **411**: 313–21.
23. Castera L, Vergniol J, Foucher J, *et al.* Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005; **128**: 343–50.
24. Mueller S, Millonig G, Sarovska L, *et al.* Increased liver stiffness in alcoholic liver disease: differentiating fibrosis from steatohepatitis. *World J Gastroenterol* 2010; **16**: 966–72.
25. Sebastiani G, Halfon P, Castera L, *et al.* SAFE biopsy: a validated method for large-scale staging of liver fibrosis in chronic hepatitis C. *Hepatology* 2009; **49**: 1821–7.
26. Palmeri ML, Wang MH, Rouze NC, *et al.* Noninvasive evaluation of hepatic fibrosis using acoustic radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. *J Hepatol* 2011; **55**: 666–72.
27. Joka D, Wahl K, Moeller S, *et al.* Prospective biopsy-controlled evaluation of cell death biomarkers for prediction of liver fibrosis and nonalcoholic steatohepatitis. *Hepatology* 2012; **55**: 455–64.
28. Kim D, Kim WR, Kim HJ, *et al.* Association between noninvasive fibrosis markers and mortality among adults with nonalcoholic fatty liver disease in the United States. *Hepatology* 2013; **57**: 1357–65.