Obesity Facts

Research Article

Obes Facts 2021;14:490–500
DOI: 10.1159/000517383

Development and Validation of a Novel Scoring System for Noninvasive Nonalcoholic Steatohepatitis Detection in Bariatric Patients

Adrian T. Billeter a  Sarah Wloka a  Rouven Behnisch b  Thomas Albrecht c
Stephanie Roessler c  Benjamin Goeppert c  Sebastian Mueller d  Felix Nickel a
Beat Müller a

aDepartment of General, Visceral and Transplant Surgery, University of Heidelberg, Heidelberg, Germany; bInstitute of Medical Biometry and Informatics, Heidelberg, Germany; cDepartment of Pathology, University of Heidelberg, Heidelberg, Germany; dDepartment of Internal Medicine, Salem Medical Center and Center for Alcohol Research, University of Heidelberg, Heidelberg, Germany

Keywords
Nonalcoholic fatty liver disease · Fatty liver · Noninvasive score · Nonalcoholic steatohepatitis · Bariatric surgery

Abstract

Introduction: Nonalcoholic fatty liver disease covers a broad spectrum. Simple steatosis has usually a benign course while nonalcoholic steatohepatitis (NASH) can progress into hepatocellular carcinoma, and cirrhosis. Therefore, differentiating patients with benign steatosis and NASH is crucial. Liver biopsy, the usual gold standard for NASH diagnosis, cannot be used as a screening method due to its associated risks. This is especially problematic for obese patients with a prevalence of nonalcoholic fatty liver disease (NAFLD) in >80% of patients. The aim of this study was therefore to develop and validate a noninvasive NASH screening test in a cohort of high-risk, morbidly obese patients. Methods: This prospective study examined diagnostic accuracy in accordance with STARD guidelines. 112 liver biopsies were consecutively assigned to either a training or validation cohort. Using the Bedossa histological scoring system, the cohorts were subdivided into NASH versus NAFLD/No NAFLD. Predictors of NASH were evaluated with receiver operating characteristic (ROC) curves. A model was then constructed using a backward stepwise logistic regression and evaluated in an independent validation cohort. Results: 53.5% of the patients had NASH and 4 patients had cirrhosis. Mean body mass index (BMI) was 49.8 ± 7.5 kg/m². Backward stepwise logistic regression identified 4 parameters associated with the presence of NASH: alanin-aminotransferase, albumin, BMI, and triglycerides. The noninvasive NASH detection score (NI-NASH-DS) had an ROC of 0.851 and 0.727 in the training and validation cohorts, respectively. Sensitivity and specificity were 77.1% and 88% in the training cohort and 88% and 48% in the validation cohort which was much better than the established noninvasive scores. Discussion/Conclusion: The NI-NASH-DS is easy-to-use, inexpensive, and noninvasive and can reliably detect NASH in patients with morbid obesity. Due to its simplicity, it can be used frequently and repeatedly.

Felix Nickel and Beat Müller share the senior authorship.
**Introduction**

The incidence of nonalcoholic fatty liver disease (NAFLD), the hepatic manifestation of the metabolic syndrome, is increasing worldwide, with a prevalence of 10–30% in the general population [1] and 50–90% among patients with obesity [2]. While steatosis is NAFLD’s only histologically unique feature, it can progress into nonalcoholic steatohepatitis (NASH), which is characterized by inflammation and ballooning in liver biopsy [3]. NASH can further advance into liver cirrhosis in 20–30% and is considered a relevant cause for liver cirrhosis of unknown etiology [4]. Additionally, NASH increases the risk for hepatocellular carcinoma (HCC) and these NASH-HCCs do not always follow the classical path, in which HCCs develop in the cirrhotic liver, as is common for alcoholic and virus-related liver disease. NASH-HCC often develops in the fibrotic but not yet cirrhotic liver making early detection difficult and resulting in reduced overall survival [5, 6]. Furthermore, due to the high prevalence of NAFLD in the population and the limitations of ultrasound in patients with obesity, regular screening for HCC with ultrasound is currently not recommended by either European or American guidelines [7, 8].

Liver biopsy is the only reliable method for diagnosing NAFLD or NASH but its invasive nature and expense precludes its widespread use as a screening method [9]. Many methods, using imaging methods, such as the FibroScan®, and noninvasive tests consisting of a panel of laboratory values, liver metabolism tests, or single biomarkers have been presented in the past [10–18]. So far, none of these tests have been established as routine tools for NASH diagnosis due to expense and/or lack of reliability. While steatosis has an overall risk of 1–2% to progress to cirrhosis, the risk for progression to cirrhosis in NASH is 20–30% [19], making a noninvasive tool to differentiate NAFLD from NASH an urgent need because these patients could benefit from interventions to treat NASH and surveillance programs [5]. The aim of this study was to develop and validate a noninvasive NASH detection score (NI-NASH-DS) in a cohort of high-risk, morbidly obese patients.

**Methods**

This study was conducted in accordance with the Standards for Reporting Diagnostic Accuracy (STARD) guidelines [20]. The study was approved by the local Ethics Committee (S-629/2013) and conducted according to the World Medical Association Declaration of Helsinki and its later amendments. All patients provided a written informed consent. Training and validation cohorts were defined before recruitment and analyzed separately. We prospectively recruited 112 consecutive patients who underwent laparoscopic sleeve gastrectomy or Roux-en-Y gastric bypass for treatment of morbid obesity at a University Hospital from November 2014 to April 2019 and who had liver biopsies taken intraoperatively. All patients undergoing bariatric surgery at our institution were considered eligible. Patients with regular alcohol consumption (20 g/day for women, 30 g/day for men) or chronic liver diseases were excluded.

Clinical data and blood samples were collected preoperatively. Liver stiffness was measured with a FibroScan® 1 day before surgery by an experienced gastroenterologist. A biopsy was taken from liver segment 3 during the bariatric operation. All patients gave their written consent to participate in this study. Furthermore, patients with increased risk for bleeding after liver biopsy, such as those with coagulation disorders or undergoing anticoagulant treatment, patients not agreeing to a liver biopsy, those who did not have a preoperative FibroScan measurement, those who did not have blood samples for IGFBP-1 (insulin-like growth factor-binding protein 1) measurements taken, and those with missing laboratory data were excluded. Clinical data included sex, age, body mass index (BMI), and comorbidities. Blood samples were used to measure laboratory parameters and IGFBP-1 and to calculate fibrosis scores.

**Fibrosis Scores**

Standard laboratory parameters included insulin (mU/L), fasting glucose (mg/dL), platelet count, high-density lipoprotein, mg/dL, low-density lipoprotein, mg/dL), triglycerides (TG, mg/dL), cholesterol (mg/dL), albumin (g/L), bilirubin (mg/dL), aspartate aminotransferase (U/L), alanine aminotransferase (ALT, U/L), alkaline phosphatase (U/L), and gamma glutamyl transferase (GGT, U/L). All these parameters were measured at the central clinical laboratory of our University Hospital and are part of our standard preoperative laboratory assessment. Laboratory-based fibrosis scores were calculated as shown in Table 1.

**Fibrosis Serum Markers**

IGFBP-1 was determined by enzyme-linked immunosorbent assays using a human IGFBP-1 enzyme-linked immunosorbent assay kit (ALPCO®, Salem, NH, USA) according to the manufacturer’s instructions in duplicates and analyzed using elisaanalysis.com software (Elisakit.com Pty Ltd.).

**Transient Elastography of the Liver**

Liver stiffness was measured by transient elastography using the FibroScan® device (FibroScan®; Echosens SA, Paris, France) 1 day prior to surgery by an experienced and trained gastroenterologist. The physicians performing the FibroScan had knowledge about the clinical data of each patient but were blinded to the results of the laboratory examination, calculated fibrosis scores, and results of the liver biopsy. All measurements were performed using the XL-probe for patients with obesity. The probe was placed in the intercostal space of the patient’s right side in projection to the right liver lobe. An examination was deemed valid when 10 measurements were included. Success rate, median value, and the interquartile range of liver stiffness were calculated. Cutoffs for liver fibrosis stage were based on those described by Wong et al. [14]: F0/F1 <7.0 kilopascal (kPa), F2 ≥7.0 kPa and <8.7 kPa, F3 ≥8.7 kPa and <10.3 kPa and F4 ≥10.3 kPa.
Liver Histology
The left liver lobe (segment 3) was biopsied at the end of surgery. To ensure adequate biopsy size and to minimize the risk of sampling error due to superficial sample extraction, a wedge biopsy technique was used. Liver biopsies were stored in formalin and then embedded in paraffin. Biopsies were processed by the Institute of Pathology at our institution. Staining was performed with hematoxylin and eosin, as well as Gomori. Two experienced and specialized liver pathologists, who were blinded to the patients’ clinical and biological data, evaluated the liver biopsies. Specimens were evaluated using the histopathological algorithm described by Bedossa et al. [21], as follows: Each specimen was first analyzed regarding steatosis, ballooning, lobular inflammation, and fibrosis. Next, the algorithm was applied, and each specimen was categorized as No NAFLD/NAFLD, or NASH. The Bedossa score is the clinically relevant score that being reliable allows differentiating patients with NASH from those without NASH which was developed in morbidly obese patients [21].

Statistical Analysis
The data from 112 patients were separated chronologically into 2 cohorts. Data from 60 patients were used to build the model (training cohort) and 52 patients were used to validate it (validation cohort). Categorial variables were described as numbers per group and binary variables as absolute and relative data. Continuous variables were described as means ± standard deviation (SD). According to histological analysis, the training and validation cohorts were divided into 2 groups (No NAFLD/NAFLD vs. NASH). A Man-Whitney U test was used to compare continuous data between the groups while the χ² test was used to compare the binary data. The predictive value of each variable was calculated using its receiver operating characteristics (ROC) and area under the curve (AUC), along with a 95% confidence interval. An AUROC of \( \gtrless 0.75 \) was considered good. In a second step, variables were included in a backward stepwise logistic regression analysis to identify which variables were associated with the presence or absence of NASH. Those variables with \( p < 0.05 \) were used to construct a scoring system. To evaluate the model, we performed a 5-fold cross-validation to reduce the risk of overfitting. Fitted model accuracy, which is the number of correct predictions divided by the total number of input samples, was then determined. To account for the chance of correct predictions of a random classifier, Cohen’s Kappa was calculated. After cross-validation, the fitted models’ AUROC as well as sensitivity and specificity were calculated. Youden index was used to calculate a cutoff value for the predicted probabilities. The model was then validated using the validation cohort. ROC curves were calculated. Statistical analysis was performed using SPSS (Version 25, German, SPSS Inc., Chicago, IL, USA) and R (Version 3.4, R Core Team 2017).

Results
Characteristics of the Patient Population
A total of 112 subjects were divided chronologically into training and validation cohorts. The clinical, laboratory, and liver biopsy data of the 112 patients who were included in the study are summarized in Table 2. The mean age was 41.8 ± 11.1 years and 28% of patients were male. All patients had obesity of at least class 2 with a mean BMI of 49.8 ± 7.5 kg/m². Out of the total cohort, 20.5% were histologically diagnosed with No NAFLD, 26% with NAFLD, and 53.5% with NASH. Previously unknown liver cirrhosis due to NASH was found in 4 patients. Except for 1 patient, all patients with NASH had liver fibrosis stages 3 and 4 in the training cohort. There were no significant differences between the training and validation cohorts except for IGFBP-1 being significantly lower in the training cohort \( (p < 0.001) \). To assess the diagnostic potential of each variable, the cohorts were divided into 2 groups (NASH vs. No NAFLD/NAFLD) since the aim of the NI-NASH-DS is to identify patients with NASH.

### Table 1. Fibrosis scores based on laboratory parameters and clinical data

| Score                      | Formula                                                                 | Ranges                                                                 |
|----------------------------|------------------------------------------------------------------------|-----------------------------------------------------------------------|
| AST/ALT-ratio (Gianinni et al., 2003) [39] | AST (U/L)/ALT (U/L) F3/4: >1; F0–2: <1 | F3/4: >1.5, F0–2: <0.5; indifferent: 0.5–1.5                        |
| APRI (Wai et al., 2003) [40]          | (AST/upper normal level)/platelet count (109/L) × 100                | F3/4: ≥2, F0–2: ≤1                                                   |
| Bard-score (Cichoż-Lach et al., 2012) [41] | AST/ALT-ratio ≥0.8 – 2 points; BMI ≥ 28 – 1 point; diabetes – 1 point | F3/4: >0.676, F0–2: ≤1                                              |
| NAFLD fibrosis score (Angulo et al., 2007) [42] | −1.675 + 0.037 × age (years) + 0.094 × BMI (kg/m²) + 1.13 × IFG/diabetes (yes = 1, no = 0) + 0.99 × AST/ALT-ratio - 0.013 × platelet count (109/L) – 0.66 × albumin (g/dL) | F3/4: >0.676, F0–2: ≤1 |

AST, aspartate aminotransferase; ALT, alanine aminotransferase; BMI, body mass index; IFG, impaired fasting glucose; NAFLD, nonalcoholic fatty liver disease.
In the training cohort, serum levels of ALT, GGT, albumin, HOMA-IR, glucose, TG, and total cholesterol were higher in patients with NASH than in patients without NASH. Aspartate aminotransferase/ALT ratio and liver stiffness were also significantly higher in patients with NASH (2).

Predictors of NASH
Diagnostic accuracy of each variable was determined by ROC analysis. An optimal cutoff value using Youden index as well as sensitivity and specificity were determined and are reported in Table 3.

***Table 2. Characteristics of the patient population***

| Parameter | All (n = 112) | Training cohort (n = 60) | Validation cohort (n = 52) |
|-----------|---------------|--------------------------|----------------------------|
|           | All no NAFLD and NAFLD (n = 25) | NASH (n = 35) | p value* | p value** |
| Age, years | 41.8±11.1 | 41.7±10.4 | 41.7±10.3 | 41.7±10.7 | 41.9±11.8 | 0.914 | 0.97 |
| Male, % | 32 (28) | 18 (30) | 5 (20) | 13 (37) | 14 (27) | 0.157 |
| BMI, kg/m² | 49.8±7.5 | 49±7.8 | 45±5.8 | 51.2±48.4 | 50.7±7.1 | 0.085 | 0.011 |
| Diabetes mellitus, % | 30 (26) | 20 (33.3) | 6 (24) | 14 (40) | 10 (19) | 0.094 | 0.199 |
| Hypertension, % | 47 (41.2) | 23 (38.3) | 8 (32) | 15 (43) | 24 (46) | 0.529 | 0.398 |
| Hypercholesterolemia, % | 5 (4.4) | 3 (5) | 1 (4) | 2 (6) | 2 (4) | 0.769 | 0.766 |
| ALT, U/L | 37.9±28.2 | 34.3±18.6 | 26±11.5 | 40.2±20.6 | 42.2±36 | 0.416 | 0.003 |
| AST, U/L | 28.4±15.8 | 26.7±13.9 | 22.8±7 | 29.5±16.8 | 30.4±17.7 | 0.355 | 0.095 |
| AP, U/L | 81.5±24 | 84.5±24.5 | 86±25.2 | 83.3±24.3 | 78.1±23 | 0.094 | 0.352 |
| GGT, U/L | 36.8±23.7 | 39.5±27.5 | 28±15.2 | 47.6±31.3 | 33.8±18.4 | 0.489 | 0.003 |
| Albumin, g/dL | 4.5±0.4 | 4.4±0.3 | 4.3±0.3 | 4.5±0.3 | 4.5±0.4 | 0.265 | 0.033 |
| Total bilirubin, mg/dL | 0.5±0.2 | 0.5±0.2 | 0.5±0.2 | 0.5±0.2 | 0.6±0.3 | 0.072 | 0.096 |
| Platelet count, n/L | 272.5±65.3 | 261.4±58.3 | 274.2±59.5 | 252.3±56.5 | 285.2±71 | 0.109 | 0.213 |
| HOMA-IR | 17.7±46 | 22.6±61.4 | 8.9±10 | 32.4±78.9 | 11.9±13.4 | 0.616 | 0.012 |
| Fasting glucose, mg/dL | 116.6±62 | 118.3±64.9 | 99.8±58.1 | 131.4±67.1 | 114.8±59.1 | 0.477 | 0.012 |
| TGs, mg/dL | 166.4±130.9 | 161.1±133.3 | 111.3±433 | 196.6±333 | 172.5±197 | 0.768 | 0.001 |
| Total cholesterol, mg/dL | 189±39.7 | 186.7±38.2 | 172.8±31.2 | 196.5±40 | 191.8±41.5 | 0.565 | 0.037 |
| HDL-cholesterol, mg/dL | 45.6±10.2 | 45.3±8.9 | 47.7±9 | 43.6±8.5 | 45.3±11.7 | 0.711 | 0.053 |
| LDL-cholesterol, mg/dL | 113.7±34.9 | 109±30.7 | 101.3±25.7 | 114.9±33.2 | 119±39 | 0.231 | 0.169 |
| IGFBP-1, µg/L | 1.6±1.6 | 1.3±1.8 | 1±2.1 | 0±1.3 | 1.8±1 | <0.001 | 0.051 |
| AST/ALT-ratio | 0.9±0.3 | 0.8±0.4 | 1±0.5 | 0.8±0.2 | 0.8±0.3 | 0.965 | 0.005 |
| APRI | 0.2±0.2 | 0.2±0.2 | 0.2±0.1 | 0.3±0.3 | 0.2±0.2 | 0.822 | 0.144 |
| Bard-score | 2.2±1.1 | 2.2±1.1 | 2.5±1.1 | 2±1 | 2.2±1.2 | 0.815 | 0.081 |
| NAFLD fibrosis score | −1±1.1 | −1±1.3 | −1.3±1.3 | −0.8±1.2 | −1±1.3 | 0.517 | 0.098 |
| Liver stiffness (FibroScan®, kPa) | 14±12.6 | 10.6±9.1 | 16.5±14.2 | n/a | 6/25/7 | 0.019 |

NAFLD, nonalcoholic fatty liver disease; ALT, alanine aminotransferase; BMI, body mass index; IGFBP-1, insulin-like growth factor-binding protein 1; TG, triglycerides; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; AP, alkaline phosphatase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SD, standard deviation; NASH, nonalcoholic steatohepatitis. The table shows mean ± SD for continuous variables, n (%) for binary variables and number per group for categorical variables. * Comparison of distribution of variables between training and validation cohorts using Mann-Whitney U test or χ² test. ** Comparison of distribution of selected variables between groups no NAFLD and NAFLD, versus NASH, using Mann-Whitney U test in training cohort or χ² test.

BMI with a cutoff value of 51 kg/m² showed an AUROC of 0.693 with a sensitivity and specificity of 48.6% and 88%, respectively. ALT, GGT, and TG provided good AUROC (AUROC = 0.75) with 0.725, 0.723, and 0.758, respectively. All 4 laboratory-based fibrosis scores as well as the serum marker IGFBP-1 (negative correlation) and FibroScan® revealed low AUROC to diagnose NASH and were therefore not included in the final model.

Construction of the NI-NASH-DS
To create the NI-NASH-DS identifying the presence or absence of NASH, significant variables differentiating...
NASH from No NAFLD/NAFLD were included in a backward stepwise logistic regression analysis. During this process, HOMA-IR tended to be excluded although it was significantly higher in patients with NASH. Furthermore, the addition of HOMA-IR did not influence the accuracy of the model resulting in the exclusion of the HOMA-IR from the model ultimately. The final NI-NASH-DS was composed of the variables BMI, ALT, TG, and albumin. The results of the logistic regression are displayed in Table 4. The parameters that were significant and ultimately included in the NI-NASH-DS are outlined in Figure 1 with their respective AUCs. Five-fold cross-validation yielded a model accuracy of 0.817 ± 0.07 with a Cohen’s Kappa of 0.62 ± 0.15. The final NI-NASH-DS’s mean AUROC was 0.851 ± 0.153 (95% confidence interval 0.785, 0.966; Fig. 1F). The regression formula based on backward stepwise logistic regression for predicting NASH based on these 4 variables is:

\[-32.771 + 0.227 \times \text{BMI} \text{ (kg/m}^2) + 0.062 \times \text{ALT (U/L)} + 0.024 \times \text{TG (mg/dL)} + 3.881 \times \text{albumin (g/dL)}.

Using the Youden index, a cutoff value of 0.649 to predict NASH was determined. Sensitivity and specificity were 77.1% and 88%, respectively. The NI-NASH-DS correctly identified 27 patients with NASH and falsely attributed NASH to 3 patients. Conversely, 22 patients with NAFLD/No NAFLD received the correct diagnosis but 8 patients with NASH were falsely categorized as NAFLD/No NAFLD.

### Table 3. ROC curve analysis of parameters associated with NASH in the training cohort (n = 60)

| Variable                  | AUROC (95% CI) | Cutoff       | Sensitivity, % | Specificity, % |
|---------------------------|----------------|--------------|----------------|----------------|
| BMI, kg/m²                | 0.693 (0.559–0.827) | 51 (kg/m²)   | 48.6           | 88             |
| ALT, U/L                  | 0.725 (0.598–0.852) | 22.5 (U/L)   | 82.8           | 52             |
| AST, U/L                  | 0.627 (0.485–0.769) | 27.5 (U/L)   | 42.8           | 84             |
| AP, U/L                   | 0.571 (0.481–0.723) | 95 (U/L)     | 80             | 40             |
| GGT, U/L                  | 0.723 (0.594–0.852) | 39.5 (U/L)   | 54.3           | 92             |
| Albumin, g/dL             | 0.662 (0.524–0.8) | 4.5 (g/dL)   | 60             | 68             |
| Total bilirubin, mg/dL    | 0.507 (0.357–0.656) | 0.8 (mg/dL)  | 94.2           | 12             |
| Platelet count, /nL       | 0.595 (0.448–0.741) | 262 (nL)     | 62.9           | 60             |
| HOMA-IR                   | 0.691 (0.555–0.827) | 7.2          | 68.6           | 68             |
| Glucose, mg/dL            | 0.691 (0.552–0.83) | 86.5 (mg/dL) | 71.4           | 68             |
| TG, mg/dL                 | 0.758 (0.636–0.88) | 141.5 (mg/dL)| 62.9           | 84             |
| Total cholesterol, mg/dL  | 0.659 (0.52–0.798) | 178.5 (mg/dL)| 62.9           | 68             |
| HDL-cholesterol, mg/dL    | 0.647 (0.506–0.787) | 37.5 (mg/dL) | 34.3           | 96             |
| LDL-cholesterol, mg/dL    | 0.606 (0.459–0.753) | 83 (mg/dL)   | 90.9           | 28             |
| IGFBP-1                   | 0.647 (0.505–0.788) | 0.35 (µL)    | 48.6           | 76             |
| AST/ALT-ratio             | 0.713 (0.583–0.842) | 0.755        | 60             | 80             |
| APRI                      | 0.611 (0.467–0.756) | 0.194        | 54.3           | 68             |
| Bard-score                | 0.627 (0.480–0.774) | 2            | 68.6           | 64             |
| NAFLD fibrosis score      | 0.626 (0.480–0.772) | −0.97        | 57.1           | 72             |
| FibroScan®                | 0.678 (0.539–0.817) | 8.9 (kPa)    | 65.7           | 68             |

NAFLD, nonalcoholic fatty liver disease; ALT, alanine aminotransferase; BMI, body mass index; IGFBP-1, insulin-like growth factor-binding protein 1; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; AP, alkaline phosphatase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SD, standard deviation; NASH, nonalcoholic steatohepatitis; ROC, receiver operating characteristic; CI, confidence interval; TG, triglycerides.

### Table 4. Results of logistic regression analysis including OR, p value, and coefficient of variables

| Parameter                  | OR (95% CI)    | p value | Regression coefficient |
|----------------------------|----------------|---------|------------------------|
| Constant                   | 0 (0.0–0.0)    | 0.001   | −32.771                |
| BMI                        | 1.255 (1.082–1.456) | 0.003   | 0.227                  |
| ALT                        | 1.064 (1.001–1.130) | 0.047   | 0.062                  |
| TG                         | 1.024 (1.006–1.043) | 0.011   | 0.024                  |
| Albumin                    | 48.468 (1.806–1301.1) | 0.021   | 3.881                  |

OR, odds ratio; CI, confidence interval; ALT, alanine aminotransferase; protein 1; TG, triglycerides.
Table 5. ROC curves, sensitivity, and specificity of our test and other noninvasive tests in the training and validation cohorts

| Variable                      | Training cohort | Validation cohort |
|-------------------------------|----------------|-------------------|
|                               | AUROC (95% CI) | cutoff | sensitivity, % | specificity, % | AUROC (95% CI) | cutoff | sensitivity, % | specificity, % |
| AST/ALT-ratio                 | 0.713 (0.583–0.842) | 0.755 | 60     | 80           | 0.613 (0.456–0.771) | 0.727 | 52     | 77.8          |
| APRI                          | 0.611 (0.467–0.756) | 0.194 | 54.3   | 68           | 0.613 (0.456–0.771) | 0.727 | 52     | 77.8          |
| Bard-score                    | 0.627 (0.480–0.774) | 2     | 68.6   | 64           | 0.745 (0.610–0.880) | 0.160 | 76     | 63            |
| NAFLD fibrosis score          | 0.626 (0.480–0.772) | −0.97 | 57.1   | 72           | 0.530 (0.381–0.680) | 2     | 56     | 55.6          |
| FibroScan®                   | 0.678 (0.539–0.817) | 8.9 (kPa) | 65.7   | 68           | 0.547 (0.382–0.712) | −1.973 | 88     | 40.7          |
| NI-NASH-DS                   | 0.851 (0.785, 0.966) | 0.649 | 77.1   | 88           | 0.727 (0.586–0.869) | 0.649 | 88     | 48            |

NI-NASH-DS, noninvasive NASH detection score; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; ROC, receiver operating characteristic; CI, confidence interval.

Fig. 1. ROC for albumin (A), ALT (B), BMI (C), IGFBP-1 (D), TG (E), and NI-NASH-DS (F) training cohort. ALT, alanine aminotransferase; AUC, area under the curve; BMI, body mass index; IGFBP-1, insulin-like growth factor-binding protein 1; ROCs, receiver operating characteristic; TG, triglycerides; NI-NASH-DS, noninvasive NASH detection score.
Validation of Results

Patients in the training and validation cohorts did not differ significantly in terms of BMI, ALT, or albumin and TG levels. The number of patients with NASH and no NAFLD were lower in the validation cohort (48.1% vs. 53.5%, \( p = 0.28 \) and 11.5% vs. 20.5%, \( p = 0.029 \)). Consequently, the validation cohort included a significantly higher number of patients with NAFLD than the training cohort (40.4% vs. 26%, \( p < 0.001 \)). ROC analysis of the NI-NASH-DS within the validation cohort demonstrated an AUROC of 0.727 (Fig. 2). After applying the calculated cutoff value of 0.649, sensitivity and specificity were 88% and 48%, respectively.

Comparison of Our Model with Noninvasive Tests

The comparison of the diagnostic accuracy for identifying NASH of our newly developed NI-NASH-DS with established noninvasive scores in the training cohort are provided in Table 5. Our scores perform better than all the other noninvasive tests in these morbidly obese patients.

**Discussion**

While many noninvasive diagnostic tests for NASH or liver fibrosis have been developed on patient cohorts with lower mean BMI or viral hepatitis, their accuracy among patients with morbid obesity and a risk of NASH is limited. To close this gap, the aim of this study was to develop a score for identifying patients with a high likelihood of NASH instead of liver fibrosis, since the pathogenesis of NASH-HCC does not always follow the classical path of cirrhosis followed by HCC [5]. Instead, many patients with NASH develop HCCs in the noncirrhotic liver. In this study, all patients with NASH save 1 already had stage 3 or 4 fibrosis, thus confirming the importance of identifying patients with NASH, due to their higher risk of developing liver fibrosis. Therefore, creating a scoring system that distinguishes NASH from NAFLD is clinically highly relevant, especially in patients with obesity, who are already at an increased risk for NASH. Patients identified by this test can then undergo further testing. In patients undergoing bariatric surgery, for example, an intraoperative liver biopsy could be performed easily.

We were able to build the NI-NASH-DS for diagnosing NASH using simple, readily available clinical information such as BMI and standard laboratory parameters. This new score differentiated NASH from NAFLD/No NAFLD much better than the other, established noninvasive scores. Using a backward stepwise logistic regression based on the variables BMI, ALT, TG, and albumin, the NI-NASH-DS’s mean AUROC in the training and validation cohorts were 0.851 and 0.727, respectively. Analysis of the training cohort showed that, with 1 exception, all patients with fibrosis stage 3 or higher were also classified as NASH. Therefore, in this cohort of morbidly obese patients with a high risk for NASH and progressive liver disease, the NI-NASH-DS can be used to identify patients with liver fibrosis stage of 3 or greater. The NI-NASH-DS is based on data from patients with a mean BMI of 49.8 ± 7.5 kg/m² who underwent bariatric surgery. Consequently, the NI-NASH-DS is especially aimed at this high-risk group of morbidly obese patients. Furthermore, due to the simplicity of the included parameters, it can be inexpensively repeated, thus compensating for the AUROC of around 0.8 for both training and validation cohorts.

A secondary aim of this study was to evaluate well-known fibrosis tests and novel, promising markers for detecting NASH, such as IGFBP-1. Their diagnostic accuracy was too low to identify patients with NASH by them-
Novel Noninvasive Score for NASH

more, our test is much easier to perform and less expensive than the NashTest, since we only included easily measurable parameters. A patient cohort with a comparable BMI was reported by Campos et al. [18], who introduced a score composed of simple clinical data (hypertension, type 2 diabetes, sleep apnea syndrome, and race) and routine laboratory parameters (AUROC 0.8). Furthermore, it was not validated with an independent cohort. The NAFLIC score by Sumida et al. [15, 31] also contains complex parameters that might not be attainable in everyday clinical practice. Additionally, patients treated with insulin or insulin sensitizers were excluded from their cohort, making their score ill-suited for a large portion of patients undergoing bariatric surgery [15]. Finally, their test was not evaluated in non-Japanese patients [31].

The German guidelines committee for fatty liver disease has reviewed the use of cytokeratin-18 fragment measurement as a diagnostic tool for NAFLD but not made any specific recommendations [32]. Two meta-analyses show sensitivities and specificities of 66%/82% [33] and 83%/71%, respectively [17]. Based on these results, Chen et al. [17] proposed using cytokeratin-18 fragments as a screening method for NASH. The results of the meta-analysis by Kwok et al. [33, 34] and another recent study found low sensitivities questioning the value of cytokeratin-18 fragments as a screening tool. The accuracy of this means of measurement is improved in combination with ALT, however, as shown by Anty et al. [13] in the Nice model (AUROC of 0.88 and 0.83 in the training and validation cohorts). The Nice model was based on a bariatric cohort with a mean BMI of 44.5 kg/m² [13]. Initial data on the LiMAX® test are promising, as it is the most accurate test to date (AUROC of 0.86). The LiMAX® cohort included bariatric patients with a high mean BMI of 53.8 kg/m². The authors found a correlation between the results of the LiMAX® test and histological hepatic inflammation and ballooning, which is also the significant difference between NAFLD and NASH [18]. The authors also found a correlation between LiMAX® test results and liver fibrosis in chronic liver disease [35]. A limitation of the LiMAX® test is that smoking interferes with its accuracy [36]. All of these tests (cytokeratin-18 fragment measurement, the Nice model, and LiMAX®) also pose significant challenges: they are relatively expensive, not readily available, and do not offer much higher diagnostic accuracy than our test. In contrast to the NI-NASH-DS, the simplicity of the NI-NASH-DS allows it to be repeated without incurring much expense.

Compared with other published tests, the NI-NASH-DS has some beneficial characteristics. However, in the next steps, the NI-NASH-DS should be validated in ex-
ternal and larger cohorts as well as in patients with lower BMI. Furthermore, since we provide the formula in this manuscript, it should be refined based on the results of other cohorts. Provided that these validation cohorts confirm our results, the NI-NASH-DS can be made accessible online as a calculator.

Our study has several limitations. First, the NI-NASH-DS was developed and validated on a relatively small patient cohort. While we did have a training and validation cohorts, the cohorts differed significantly in number of patients with NAFLD. However, separating the cohorts into NASH and NAFLD/No NAFLD groups, based on a clinically relevant differentiation, minimized this imbalance. Second, while this imbalance may affect the model, it also reflects the “real world” and this is a primary reason to validate such a test. In addition to the imbalance in NASH in the cohorts, the NI-NASH-DS was fitted for the training cohort explaining the gap in AUROC values (0.851 vs. 0.727). While we tried to minimize the risk of overfitting by using 5-fold cross-validation when creating the scoring system, the scoring system must be further evaluated in larger cohorts and subsequently adjusted. Next, our reference standard for diagnosing NASH is not devoid of sources of possible errors. By choosing an intraoperative approach rather than a percutaneous one, we tried to minimize the risk of sampling error, since the liver biopsy could be taken from the area appearing the least normal. However, the subcapsular liver biopsy may also lead to an overestimation of the degree of fibrosis while the timing of the biopsy after the operation may also alter the histology. To further reduce the potential for error and variation in the assessment, the reference and index tests were conducted in a standardized manner by 2 expert liver pathologists while the laboratory tests were performed in a standardized and certified laboratory. Furthermore, the liver stiffness measurements were performed by trained gastroenterologists who used the Fibroscan® in daily practice. Therefore, while some inter-observer and inter-measurement variations cannot be excluded, all measures were taken to reduce the risks. Additionally, the use of widely used and standardized, easy to determine clinical and laboratory parameters included in the score guarantees a reliable measurement in essentially all hospitals and practices.

Regarding the scoring system to differentiate NASH from NAFLD/No NAFLD, we deliberately chose to evaluate and divide the liver biopsies using the SAF score published by Bedossa et al. [21], rather than using the well-known NAFLD activity score (NAS) by Kleiner et al. [37] because it has been emphasized that the NAS was not developed as a diagnostic tool for separating NASH from NAFLD [38]. In contrast, the SAF score provides a clinically relevant differentiation while being at the same time less bound to errors. Critical views that the SAF score might oversimplify NASH because it only contains 3 parameters were disproven by a blinded validation [21].

Conclusion

The NI-NASH-DS proposed in this study provides a simple, cost-efficient, and noninvasive screening method for identifying patients with morbid obesity with NASH in a high-risk cohort. Due to the simplicity of the included parameters, the NI-NASH-DS can be used easily and repeatedly, during regular outpatient clinic visits in patients with obesity and a high risk for NASH. Early identification and therapy for patients with NASH is critical because of its associated risks for progressing into cirrhosis and NASH-HCC.

Statement of Ethics

All procedures in studies involving human participants were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval was obtained from the Institutional Ethics Committee (S-629/2013). Written informed consent was obtained from all individual participants included in the present study.

Conflict of Interest Statement

The authors declare that they have no conflicts of interest.

Funding Sources

There was no funding for this study.

Author Contributions

Adrian T. Billeter contributed to study idea, acquisition, analysis, and interpretation of the data and drafting of the manuscript. Sarah Wloka contributed to study idea, acquisition, analysis, and interpretation of the data and drafting of the manuscript. Rouven Behnisch contributed to analysis and interpretation of the data. Thomas Albrecht, Stephanie Roessler, and Benjamin Goeppert
References

1. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther.* 2011 Aug;34(3):274–85.

2. Stal P. Liver fibrosis in non-alcoholic fatty liver disease: diagnostic challenge with prognostic significance. *World J Gastroenterol.* 2015 Oct 21;21(39):11077–87.

3. Yeh MM, Brunt EM. Pathological features of fatty liver disease. *Gastroenterology.* 2014 Oct;147(4):754–61.

4. Roeb E, Steffen HM, Bantel H, Baumann U, Canbay A, Demir M, et al. [Skl Guideline non-alcoholic fatty liver disease]. *Z Gastroenterol.* 2015 Jul;53(7):668–723.

5. Piscaglia F, Svegliati-Baroni G, Barchetti A, Pecorelli A, Marinelli S, Tiribelli C, et al. Clinical patterns of hepatocellular carcinoma in nonalcoholic fatty liver disease: a multicenter prospective study. *Hepatology.* 2016 Mar;63(3):827–38.

6. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology.* 2018 Jan;67(1):328–57.

7. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol.* 2018 Jul;69(1):182–236.

8. Marrero JA, Kullik LM, Sirlin CB, Zhu AX, Finn RS, Abecasis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology.* 2018 Aug;68(2):723–50.

9. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association for the Study of Liver Diseases. *American Association for the Study of Liver Diseases.* 2012 Jun;55(6):2005–23.

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. *BMJ Gastroenterol.* 2006 Nov 10;6:34.

11. Campos GM, Bambha K, Vittinghoff E, Rabl C, Posselt AM, Ciovica R, et al. A clinical scoring system for predicting nonalcoholic steatohepatitis in morbidly obese patients. *Hepatology.* 2008 Aug;47(6):1916–23.

12. Grafigna MN, Belli SH, de Larrafaaga G, Finboim H, Estepo C, Peres S, et al. [Insulin-like growth factor-binding protein-1: a new biochemical marker of nonalcoholic fatty liver disease?]. *Acta Gastroenterol Latinoam.* 2009 Mar;39(1):30–7.

13. Anty R, Ianneli A, Patouraux S, Bonnafous S, Lavallard VJ, Senni-Buratti M, et al. A new composite model including metabolic syndrome, alanine aminotransferase and cyto-keratin-18 for the diagnosis of non-alcoholic steatohepatitis in morbidly obese patients. *Aliment Pharmacol Ther.* 2010 Dec;32(11–12):1315–22.

14. Wong VW, Vernaio J, Wong GL,oucher J, Chan HL, Le Bail B, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology.* 2010 Feb;51(2):454–62.

15. Sumida Y, Yoneda M, Hydro H, Yamauchi K, Ono M, Fujii H, et al. A simple clinical scoring system using ferritin, fasting insulin, and type IV collagen 7S for predicting steatohepatitis in nonalcoholic fatty liver disease. *J Gastroenterol.* 2011 Feb;46(2):257–68.

16. Wong VW, Vergnoli J, Wong GL, Foucher J, Chan AH, Chermak F, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. *Am J Gastroenterol.* 2012 Dec;107(12):1862–71.

17. Chen J, Zhu Y, Zheng Q, Jiang J. Serum cyto-keratin-18 in the diagnosis of non-alcoholic steatohepatitis: a meta-analysis. *Hepat Res.* 2014 Aug;44(8):854–62.

18. Alizai PH, Lurie I, Kroh A, Schmitz S, Luedde T, Andruszkow J, et al. Noninvasive evaluation of liver function in morbidly obese patients. *Gastroenterol Res Pract.* 2019;2019:4307462.

19. Perumpail BJ, Khan MA, Yoo ER, Cholankeril G, Kim D, Ahmed A. Clinical epidemiology and disease burden of nonalcoholic fatty liver disease. *World J Gastroenterol.* 2017 Dec 21;23(47):8283–76.

20. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015 Oct 28;351:h5257.

21. Bedossa P, Poitou C, Veyrie N, Bourliot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology.* 2012 Nov;56(5):1751–9.

22. Angulo P, Bugianesi E, Bjornsson ES, Charatcharoenwitthaya P, Mills PR, Bannister F, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology.* 2013 Oct;145(4):782–9.e4.

23. Park KS, Lee YS, Park HW, Seo SH, Jang BG, Hwang JY, et al. Factors associated or related to with pathological severity of nonalcoholic fatty liver disease. *Korean J Intern Med.* 2004 Mar;19(1):19–26.

24. Dixon JB, Bhalal PS, O’Brien PE. Nonalcoholic fatty liver disease: predictors of nonalcoholic steatohepatitis and liver fibrosis in the severely obese. *Gastroenterology.* 2001 Jul;121(1):91–100.

25. Mofrad P, Contos MJ, Haque M, Sargeant C, Fisher RA, Luketic VA, et al. Clinical and histological spectrum of nonalcoholic fatty liver disease associated with normal ALT values. *Hepatology.* 2003 Jun;37(6):1286–92.

26. Wong RJ, Tran T, Kaufman H, Niles J, Gish R. Increasing metabolic co-morbidities are associated with higher risk of advanced fibrosis in nonalcoholic steatohepatitis. *PLOS One.* 2019;14(8):e0220612.

27. Samuel VT, Liu ZX, Qu X, Elder BD, Bilz S, Befroy D, et al. Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J Biol Chem.* 2004 Jul 30;279(31):32345–53.

28. Jornayvaz FR, Shulman GI. Dicacylglycerol activation of protein kinase Ca and hepatic insulin resistance. *Cell Metab.* 2012 May 2;15(5):574–84.

29. Fierbinteanu-Braticevici C, Baicus C, Tribus M, Jornayvaz FR, Shulman GI. Diacylglycerol activation of protein kinase Ca and hepatic insulin resistance. *Cell Metab.* 2012 May 2;15(5):574–84.

30. Fierbinteanu-Braticevici C, Baicus C, Tribus M, Jornayvaz FR, Shulman GI. Diacylglycerol activation of protein kinase Ca and hepatic insulin resistance. *Cell Metab.* 2012 May 2;15(5):574–84.

31. Fierbinteanu-Braticevici C, Baicus C, Tribus M, Jornayvaz FR, Shulman GI. Diacylglycerol activation of protein kinase Ca and hepatic insulin resistance. *Cell Metab.* 2012 May 2;15(5):574–84.

32. Roeb E, Steffen HM, Bojunga J. [Non-alcoholic fatty liver diseases]. *Dtsch Med Wochenschr.* 2015 Dec;140(24):1846–67.
33 Kwok R, Tse YK, Wong GL, Ha Y, Lee AU, Ngu MC, et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease – the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment Pharmacol Ther.* 2014 Feb;39(3):254–69.

34 Cusi K, Chang Z, Harrison S, Lomonaco R, Bril F, Orsak B, et al. Limited value of plasma cytokeratin-18 as a biomarker for NASH and fibrosis in patients with non-alcoholic fatty liver disease. *J Hepatol.* 2014 Jan;60(1):167–74.

35 Buechter M, Thimm J, Baba HA, Bertram S, Willuweit K, Gerken G, et al. Liver maximum capacity: a novel test to accurately diagnose different stages of liver fibrosis. *Digestion.* 2019;100(1):45–54.

36 Kasicka-Jonderko A, Loska D, Jonderko K, Kaminska M, Blonska-Fajrowska B. Interference of acute cigarette smoking with [(1)(3)C] methacetin breath test. *Isotopes Environ Health Stud.* 2011 Mar;47(1):34–41.

37 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 Jun;41(6):1313–21.

38 Brunt EM, Kleiner DE, Behling C, Contos MJ, Cummings OW, Ferrell LD, et al. Misuse of scoring systems. *Hepatology.* 2011 Jul;54(1):369–71. author reply 70–1.

39 Giannini E, Risso D, Botta F, Chiaroniello B, Fasoli A, Malfatti F, et al. Validity and clinical utility of the aspartate aminotransferase-alanine aminotransferase ratio in assessing disease severity and prognosis in patients with hepatitis C virus-related chronic liver disease. *Arch Intern Med.* 2003;163(2):218–24.

40 Wai CT, Greenson JK, Fontana RJ, Kalbflieisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology.* 2003;38(2):518–26.

41 Cichoż-Lach H, Geliński K, Prozorow-Król B, Swatek J, Słomka M, Lach T. The BARD score and the NAFLD fibrosis score in the assessment of advanced liver fibrosis in nonalcoholic fatty liver disease. *Med Sci Monit.* 2012;18(12):CR735–40.

42 Angulo P, Hui JM, Marchesini G, Bugiansi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology.* 2007;45(4):846–54.