



Article

External Validation of Surrogate Indices of Fatty Liver in the General Population: The Bagnacavallo Study

Francesco Giuseppe Foschi ^{1,†}, Fabio Conti ¹, Marco Domenicali ², Pierluigi Giacomoni ¹, Alberto Borghi ², Vittoria Bevilacqua ¹, Lucia Napoli ¹, Dante Berardinelli ¹, Mattia Altini ¹, Alessandro Cucchetti ², Giorgio Ercolani ², Andrea Casadei-Gardini ³ , Stefano Bellentani ⁴, Amalia Gastaldelli ⁵ , Claudio Tiribelli ⁴ , Giorgio Bedogni ^{4,*} and Bagnacavallo Study Group ‡

- ¹ Azienda Unità Sanitaria della Romagna, Ospedale degli Infermi, 48018 Faenza, Italy; francesco.foschi@auslromagna.it (F.G.F.); fabio.conti@auslromagna.it (F.C.); pierluigi.giacomoni@alice.it (P.G.); vittoria.bevilacqua@auslromagna.it (V.B.); lucia.napoli@auslromagna.it (L.N.); dante.berardinelli@auslromagna.it (D.B.); mattia.altini@auslromagna.it (M.A.)
- ² Faculty of Medicine and Surgery, University of Bologna, 40138 Bologna, Italy; marco.domenicali@gmail.com (M.D.); alberto.borghi@auslromagna.it (A.B.); alessandro.cucchetti2@unibo.it (A.C.); giorgio.ercolani@auslromagna.it (G.E.)
- ³ Faculty of Medicine and Surgery, University of Modena and Reggio Emilia, 41121 Modena, Italy; casadeigardini@gmail.com
- ⁴ Italian Liver Foundation, 34012 Basovizza, Italy; bellentanistefano@gmail.com (S.B.); ctliver@fegato.it (C.T.)
- ⁵ Italian National Research Council, 56124 Pisa, Italy; amalia@ifc.cnr.it
- * Correspondence: giorgiobedogni@gmail.com
- † These authors contributed equally to the present work.
- ‡ Membership of the Bagnacavallo Study Group is provided in the Acknowledgements.



Citation: Foschi, F.G.; Conti, F.; Domenicali, M.; Giacomoni, P.; Borghi, A.; Bevilacqua, V.; Napoli, L.; Berardinelli, D.; Altini, M.; Cucchetti, A.; et al. External Validation of Surrogate Indices of Fatty Liver in the General Population: The Bagnacavallo Study. *J. Clin. Med.* **2021**, *10*, 520. <https://doi.org/10.3390/jcm10030520>

Received: 28 December 2020
Accepted: 24 January 2021
Published: 1 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: We externally validated the fatty liver index (FLI), the lipid accumulation product (LAP), the hepatic steatosis index (HSI), and the Zhejiang University index (ZJU) for the diagnosis of fatty liver (FL) and non-alcoholic fatty liver disease (NAFLD) in the general population. The validation was performed on 2159 citizens of the town of Bagnacavallo (Ravenna, Italy). Calibration was evaluated by calculating the calibration slope and intercept and by inspecting calibration plots; discrimination was evaluated using the c-statistic. The average calibration slope was 1 and the average intercept was 0 for all combinations of outcomes and indices. For the diagnosis of FL, the c-statistic was 0.85 for FLI, 0.83 for ZJU, 0.82 for HSI, and 0.80 for LAP; for the diagnosis of NAFLD, the c-statistic was 0.77 for FLI, 0.76 for ZJU, 0.75 for HSI, and 0.74 for LAP. All indices were strongly correlated with each other. In conclusion, FLI, LAP, HSI, and ZJU perform similarly well to diagnose FL and NAFLD in the Bagnacavallo population, even if FLI has a small advantage as discrimination is concerned.

Keywords: cross-sectional study; diagnostic techniques and procedures; validation study; fatty liver; non-alcoholic fatty liver disease

1. Introduction

Fatty liver (FL), the most common liver disease worldwide, has been classified into non-alcoholic fatty liver disease (NAFLD) and alcoholic fatty liver disease (AFLD) for almost 40 years [1]. Such dichotomization has been increasingly criticized, so an international panel of experts has recently proposed to abandon the NAFLD definition, adopting instead the more comprehensive definition of metabolic dysfunction-associated fatty liver disease (MAFLD), which has the advantage of being independent of alcohol intake [2–4].

Independently of its etiology, FL is operationally defined as visible steatosis in more than 5% of hepatocytes at liver biopsy or as an intrahepatic triglyceride content of at least 5.6% at magnetic resonance spectroscopy (MRS) or magnetic resonance imaging [5]. Liver biopsy can be performed only in selected patients followed at secondary or tertiary care

centers and the use of magnetic resonance techniques is restricted to a few research centers because of its cost [5].

The most common method to diagnose FL in clinical practice and epidemiological research is liver ultrasonography (LUS) [5]. Another method is the measurement of the controlled attenuation parameter using FibroScan, which is however expensive and not readily available [5]. When LUS is not available, the presently suggested method to diagnose FL is the calculation of surrogate indices of FL [5]. Such indices are referred to as “biomarkers” by the European Association for the Study of the Liver (EASL) guidelines on NAFLD [5].

As for any diagnostic test, the performance of FL biomarkers should be externally validated in terms of calibration and discrimination [6–8]. However, as it happens for most diagnostic research [6–8], calibration is often neglected by the available validation studies of FL biomarkers, with some notable exceptions [9–11]. Calibration is nonetheless the primary requirement to perform decision-making and inform patients, and a test with high discrimination but low (or unknown) calibration is not clinically useful [6–8].

Using LUS as the reference method, we evaluated the calibration and discrimination of FL biomarkers at diagnosing FL and NAFLD in the general population of the Bagnacavallo study [12–14]. The study was reported following the TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidelines [6] (Appendix A).

2. Subjects and Methods

2.1. Source of Data

The validation of FL biomarkers was performed using data collected during the Bagnacavallo Study [12–14]. The Bagnacavallo study was aimed at evaluating the prevalence of and the risk factors for FL in a cross-section of the general population of a Northern Italy town, and at developing a cohort of subjects from the general population where the association between FL and incident health outcomes could be studied. The study was approved by the Ethics Committee of Area Vasta Romagna-IRST (reference number 112), and all subjects gave their written informed consent.

2.2. Participants

As described in detail elsewhere [13], 3933 citizens of the town of Bagnacavallo (Ravenna, Italy) aged 30 to 60 years, were studied between October 2005 and March 2009. Altered liver enzymes were defined as alanine transaminase (ALT) > 40 U/L and/or aspartate transaminase (AST) > 37 U/L, i.e., the upper limit of normal of the laboratory. After exclusion of subjects with hepatitis B virus (HBV) infection, hepatitis C virus (HCV) infection, and unavailability of LUS, the Bagnacavallo cross-sectional analysis was performed on 349 citizens with and 1810 without altered liver enzymes [13]. The same sample of 2159 citizens was analyzed here. All participants underwent a detailed clinical history and physical examination [15]. Alcohol intake was assessed by interview [13]. Weight and height were measured following international guidelines [16] and waist circumference was measured at the midpoint between the last rib and the iliac crest [17]. Body mass index (BMI) was calculated as weight (m)/height (m)² [18]. Performed blood tests included: (1) glucose; (2) triglycerides; (3) total cholesterol; (4) high-density lipoprotein (HDL) cholesterol; (5) low-density lipoprotein (LDL) cholesterol; (6) ALT; (7) AST; (8) GGT. Systolic and diastolic blood pressure was measured using a sphygmomanometer following international guidelines. (The recommended method of measurement of blood pressure remained the same during the study period.) The metabolic syndrome was diagnosed using the harmonized international definition [19].

2.3. Outcomes

The primary outcome was LUS-diagnosed FL and the secondary outcome was LUS-diagnosed NAFLD. We focused on LUS not because it is the gold-standard method but

because it is the method most commonly employed [5]. We did not consider other diagnostic methods besides LUS, e.g., MRS, because we wanted to control the error attributable to the choice of the reference method.

LUS was performed by five experienced physicians using the same methodology employed by the Dionysos Nutrition & Liver Study [15]. In detail, normal liver was defined as the absence of liver steatosis or other liver abnormalities; light FL was defined as the presence of slight bright liver or hepatorenal echo contrast without intrahepatic vessels blurring and no deep attenuation; moderate FL as the presence of mild bright liver or hepatorenal echo contrast without intrahepatic vessel blurring and with deep attenuation; and severe FL as diffusely severe bright liver or hepatorenal echo contrast, with intrahepatic vessels blurring (no visible borders) and deep attenuation without visibility of the diaphragm. For the present analysis, FL was coded as 0 = normal liver and 1 = any degree of FL.

NAFLD was defined as FL associated with ethanol intake <2 alcohol units (20 g/day) in women and <3 alcohol units (30 g/day) in men testing negative for HBV surface antigen and anti-HCV antibodies and not under treatment with steatogenic drugs [5]. Alcoholic fatty liver disease (AFLD) was defined as FL associated with ethanol intake ≥ 2 alcohol units in women and ≥ 3 alcohol units in men testing negative for hepatitis B surface antigen and anti-HCV antibodies and not under treatment with steatogenic drugs [5]. For the present analysis, NAFLD was coded as 0 = normal liver or AFLD and 1 = NAFLD.

2.4. Predictors

We identified five non-patented FL biomarkers, developed using LUS as the reference method, for potential inclusion into the study: fatty liver index (FLI) [20], lipid accumulation product (LAP) [17], hepatic steatosis index (HSI) [21], Zhejiang University index (ZJU) [22], and the index of non-alcoholic steatohepatitis (ION) [23].

FLI is suggested by the European Association for the Study of the Liver (EASL) as a “biomarker” of liver steatosis [5]. Other “biomarkers” suggested by the EASL are SteatoTest [24], which is based on a proprietary formula and could not be validated here, and the NAFLD-liver fat score (NAFLD-LFS) [25], which was developed using MRS as the reference method and was therefore not considered here. We were also unable to calculate NAFLD-LFS because insulin, which is a required predictor of NAFLD-LFS, was available only in 1415 (66%) of our 2159 subjects. For the same reason and because of the unavailability of hip circumference, we could not to calculate the ION index, which requires both insulin and the waist-to-hip ratio. (We could have imputed the missing values of insulin [12], but we chose not to do so because insulin is a key predictor of FL [20], and key predictors should not be missing when developing or validating prediction models [7].)

FLI and LAP were developed to predict FL while HSI and ZJU were developed to predict NAFLD. All biomarkers were developed in cross-sections of individuals from the general population (FLI, LAP) or health-care facilities (HSI, ZJU), by matching individuals with FL or NAFLD to individuals without it. The formulae for calculating FLI, LAP, HSI and ZJU are given in Table A3 of Appendix B.

2.5. Sample Size

We did not perform any formal sample size calculation but were quite confident that, with 896/2159 (42%) cases of FL and 567/2159 (26%) cases of NAFLD, we could attain a precise assessment of the performance of the biomarkers [13]. At least 200 events and 200 non-events are in fact required for reasonable external validation of model performance [6,7].

2.6. Missing Data

There were no missing data.

2.7. Statistical Analysis

Most continuous variables were not Gaussian-distributed, and all are reported as median (50th percentile) and interquartile range (25th and 75th percentiles). Discrete variables are reported as the number and proportion of subjects with the characteristic of interest.

Calibration was evaluated by applying Van Calster's three-level hierarchy [8,26]. Level 1 of this hierarchy is "mean calibration" or "calibration-in-the-large", which compares the observed event rate with the average predicted risk; level 2 is "weak calibration", which consists of a logistic calibration analysis testing whether the calibration slope is 1 and the calibration intercept is 0, and is aimed at detecting overestimation or underestimation of risk; lastly, level 3 is "moderate calibration", which uses a "calibration plot" to test whether the predicted risks correspond to the observed event rates. Such a graph plots the predicted (expected) outcome probabilities (x-axis) against the observed outcome frequencies (y-axis). As suggested by the TRIPOD guidelines, we performed the calibration using tenths of the predicted risk and superimposed a line obtained by locally weighted scatterplot smoothing [6]. A well-calibrated model shows predictions lying or around the 45° line of the calibration plot [6]. Discrimination was evaluated using Harrell's c-statistic [27].

Statistical analysis was performed using Stata 16.1 (Stata Corporation, College Station, TX, USA) with the *pmcalplot* module [28], and R 4.0.3 (R Core Team 2020, Vienna, Austria) with the *val.prob.ci.2* function [8]. R code was run from within Stata using the *rcall* package [29].

3. Results

3.1. Study Population

The measurements of the 2159 citizens who took part in the study are given in Table 1 and are discussed in greater detail elsewhere [12]. FL was diagnosed in 896/2159 (42%, 95% CI 39 to 44%) and NAFLD in 567/2159 (26%, 24 to 28%) citizens.

Table 1. Measurements of the study subjects.

	<i>n</i> = 2159
Altered liver enzymes	349 (16.2%)
Male sex	1079 (50.0%)
Age (years)	49 (41; 56)
Body mass index (kg/m ²)	25.5 (23.0; 29.0)
Normal liver	1263 (58.5%)
Fatty liver	896 (41.5%)
Non-alcoholic fatty liver disease	567 (26.3%)
Alcoholic fatty liver disease	329 (15.2%)
Waist circumference (cm)	101.0 (94.0; 108.0)
Glucose (mg/dL)	89 (83; 97)
Triglycerides (mg/dL)	102 (71; 153)
Total cholesterol (mg/dL)	209 (185; 235)
HDL-cholesterol (mg/dL)	59 (49; 71)
LDL-cholesterol (mg/dL)	128 (105; 152)
Systolic blood pressure (mm Hg)	130 (120; 140)
Diastolic blood pressure (mm Hg)	80 (80; 90)
Metabolic syndrome	615 (28.5%)
Alanine transaminase (U/L)	22 (16; 32)
Aspartate transaminase (U/L)	22 (18; 26)
Gamma-glutamyltransferase (U/L)	19 (13; 32)
Alcohol intake (units/day)	2 (0; 4)
Fatty liver index (FLI)	46 (21; 76)
Lipid accumulation product (LAP)	44 (28; 75)
Hepatic steatosis index (HSI)	39 (35; 44)
Zhejiang University index (ZJU)	36 (33; 41)

Continuous variables are reported as median (50th percentile) and interquartile range (25th and 75th percentiles). Discrete variables are reported as the number and proportion of subjects with the characteristic of interest. HDL = high-density lipoprotein; LDL = low-density lipoprotein.

3.2. Diagnosis of FL

The average expected rate of FL (42%) equaled the average observed rate (42%) for all biomarkers, showing a satisfactory mean calibration (Table 2).

Table 2. Calibration and discrimination of the fatty liver index, lipid accumulation product, hepatic steatosis index and Zhejiang University index at diagnosing fatty liver and non-alcoholic fatty liver disease.

Fatty Liver				
	FLI	LAP	HSI	ZJU
Expected event rate [†]	0.42 (0.40 to 0.43)			
Calibration intercept	0.00 (−0.11 to 0.11)	0.00 (−0.10 to 0.10)	0.00 (−0.10 to 0.10)	0.00 (−0.10 to 0.10)
Calibration slope	1.00 (0.92 to 1.08)	1.00 (0.89 to 1.11)	1.00 (0.91 to 1.09)	1.00 (0.91 to 1.09)
C-statistic	0.85 (0.84 to 0.87)	0.80 (0.78 to 0.82)	0.82 (0.80 to 0.83)	0.83 (0.81 to 0.85)
Non-Alcoholic Fatty Liver Disease				
	FLI	LAP	HSI	ZJU
Expected event rate ^{††}	0.26 (0.25 to 0.28)	0.26 (0.24 to 0.28)	0.26 (0.25 to 0.28)	0.26 (0.25 to 0.28)
Calibration intercept	0.00 (−0.11 to 0.11)	0.00 (−0.10 to 0.10)	0.00 (−0.10 to 0.10)	0.00 (−0.10 to 0.10)
Calibration slope	1.00 (0.89 to 1.11)	1.00 (0.84 to 1.16)	1.00 (0.88 to 1.12)	1.00 (0.88 to 1.12)
C-statistic	0.77 (0.75 to 0.79)	0.74 (0.71 to 0.76)	0.75 (0.72 to 0.77)	0.76 (0.74 to 0.78)

Values are averages and 95% confidence intervals; [†] vs. observed event rate of 0.42 (0.39 to 0.44); ^{††} vs. observed event rate of 0.26 (0.24 to 0.28). Abbreviations: FLI = fatty liver index; LAP = lipid accumulation product; HSI = hepatic steatosis index; ZJU = Zhejiang University index.

At logistic calibration, the average calibration slope was 1 and the average intercept was 0 for all biomarkers, showing a satisfactory weak calibration (Table 2). Lastly, the examination of calibration plots (Figure 1) showed an acceptable profile of moderate calibration for all predictors. FLI had the highest (0.85) c-statistic, followed by ZJU (0.83), HSI (0.82), and LAP (0.80).

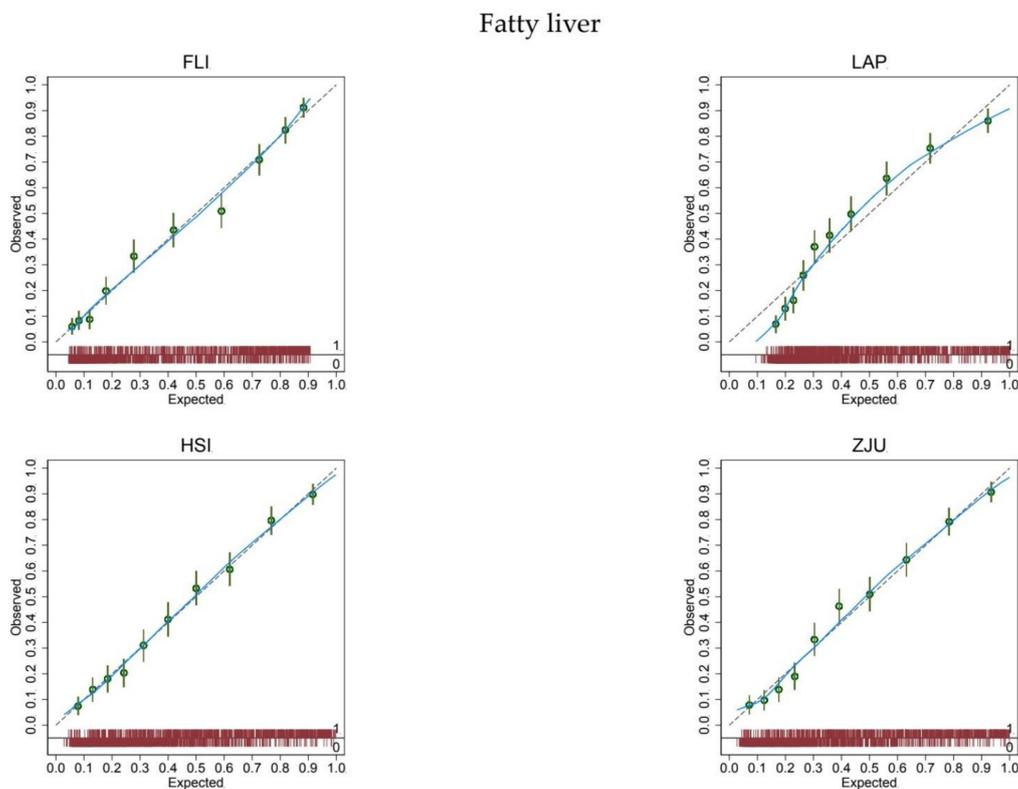


Figure 1. Calibration plots for the diagnosis of fatty liver.

The expected (predicted) risk is divided into 10 equally sized groups (tenths). The green dots and spikes on the diagonal line are average risks and 95% confidence intervals. The dotted line is the reference line of calibration. The blue line connecting the green dots is obtained by locally weighted scatterplot smoothing. The red spike plot at the bottom gives the distribution of fatty liver (0 = no; 1 = yes). Abbreviations: FLI = fatty liver index; LAP = lipid accumulation product; HSI = hepatic steatosis index; ZJU = Zhejiang University index.

3.3. Diagnosis of NAFLD

The expected rate of NAFLD (26%) equaled the observed rate (26%) for all biomarkers, showing a satisfactory mean calibration (Table 2). At logistic calibration, the average calibration slope was 1 and the average intercept was 0 for all biomarkers, showing a satisfactory weak calibration (Table 2). Lastly, the examination of calibration plots showed an acceptable profile of moderate calibration for all predictors (Figure 2). FLI had the highest (0.77) c-statistic, followed by ZJU (0.76), HSI (0.75), and LAP (0.74).

NAFLD

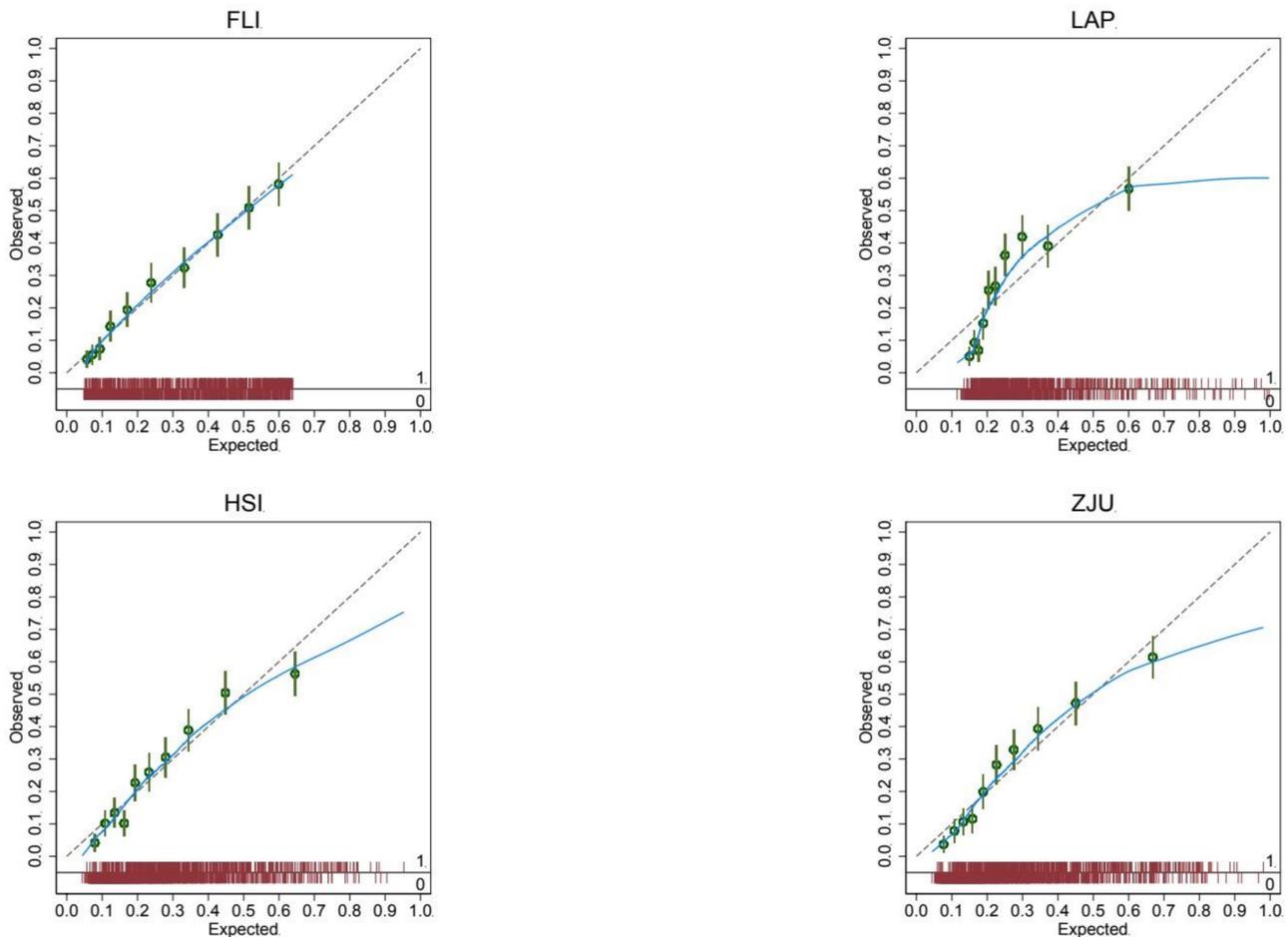


Figure 2. Calibration plots for the diagnosis of non-alcoholic fatty liver disease.

The expected (predicted) risk is divided into 10 equally sized groups (tenths). The green dots and spikes on the diagonal line are average risks and 95% confidence intervals. The dotted line is the reference line of calibration. The blue line connecting the green dots is obtained by locally weighted scatterplot smoothing. The red spike plot at the bottom gives the distribution of NAFLD (0 = no; 1 = yes). Abbreviations: NAFLD = non-alcoholic

fatty liver disease; FLI = fatty liver index; LAP = lipid accumulation product; HSI = hepatic steatosis index; ZJU = Zhejiang University index.

3.4. Association between Biomarkers

Further analysis revealed a strong association between all biomarkers (Figure 3), partially explained by the use of the same or highly correlated predictors (Table A2 of Appendix B).

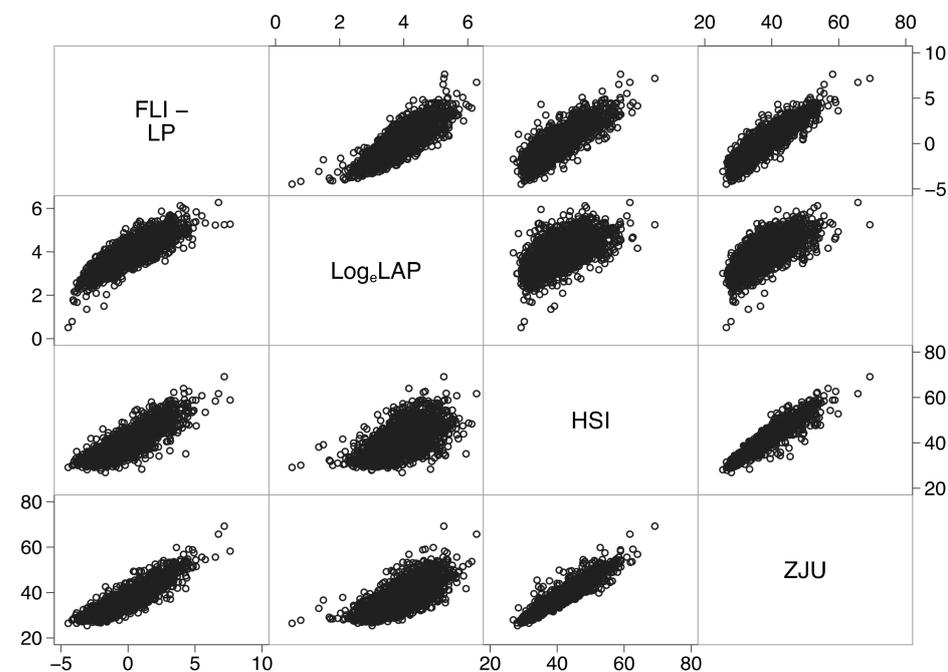


Figure 3. Correlation matrix showing strong associations between all biomarkers. Abbreviations: FLI-LP = fatty liver index-linear predictor (see Table A3 of Appendix B); Log_eLAP = natural logarithm of the lipid accumulation product; HSI = hepatic steatosis index; ZJU = Zhejiang index.

For instance, the linear predictor of FLI explained 72% of the variance of HSI, 81% of the variance of ZJU, and 51% of the variance of log_e-transformed LAP. Moreover, ZJU explained 89% of the variance of HSI. The similar performance of these biomarkers at diagnosing FL and NAFLD (Table 1) is thus likely to be partially explained by their underlying mutual association.

4. Discussion

In the present study, we took advantage of the Bagnacavallo cross-sectional study of liver disease [13] to externally validate FLI [20], LAP [17], HSI [21], and ZJU [22] for the diagnosis of FL and NAFLD in the general population. All biomarkers showed an acceptable mean, weak, and moderate calibration for the diagnosis of FL (Figure 1 and Table 2) and NAFLD (Figure 2 and Table 2) [8].

We hypothesized that FLI would perform better than LAP, HSI, and ZJU at diagnosing FL and possibly NAFLD in the present population. (We had some reservations about NAFLD because FLI was purposely developed to predict FL.) Our hypothesis was based on the fact that FLI was developed in the general population of a town (Campogalliano, Modena, Italy) similar to the one studied here (Bagnacavallo, Ravenna, Italy) [20]. We expected, however, in line with the available evidence [9–11], that FLI had to be recalibrated for proper use in the Bagnacavallo population. We were thus surprised to find that FLI had a satisfactory mean, weak and moderate calibration, and that it could be applied without modification to the Bagnacavallo population for the diagnosis of both FL and NAFLD. We were even more surprised to find that biomarkers (LAP, HSI and ZJU) developed in

different populations (US, Korea and China) showed a satisfactory profile of mean, weak, and moderate calibration in the Bagnacavallo population.

The strengths of the present study are that it was performed in a representative sample of the general population, that it enrolled a high number of subjects, and that it had a high observed event rate for both FL and NAFLD. A sample size of at least 200 subjects with and 200 without the outcome of interest is presently suggested for proper validation of a diagnostic test [7]. With its 896 citizens with and 1263 without FL, and 567 citizens with and 1592 without NAFLD, the Bagnacavallo Study is thus in an excellent position to serve as a platform to externally validate biomarkers of FL.

A limitation of the present study is the unavailability of some predictors needed to calculate the ION index [23], which was one of the biomarkers that we identified as theoretically suitable for validation in this population. The ION index employs insulin, which was available only in a subsample of subjects, and hip circumference, which was not measured in the Bagnacavallo study. Moreover, the partial availability of insulin and the unavailability of C-reactive protein impeded us to diagnose MAFLD and to evaluate the performance of FL biomarkers at diagnosing this newly proposed entity, which is expected to attract much attention in coming years [2,3]. Furthermore, even if we chose to include studies that used only LUS to diagnose FL to reduce the error attributable to the choice of the reference method, LUS is known to offer an accurate assessment of FL only starting from an intrahepatic triglyceride content of 10% [30].

External calibration is more important than discrimination at establishing the value of a test for a given diagnosis, but has not been taken into account by most diagnostic studies of FL biomarkers, with some notable exceptions [9–11]. This is not to say that discrimination is irrelevant because, in the presence of an acceptable calibration, the greater discrimination is preferable. Another problem of most diagnostic studies is that they compare an externally derived predictor with an internally derived one and go on to declare the latter superior to the former [6]. This is, however, largely expected on both theoretical and empirical grounds, and is one of the primary reasons why external validation of diagnostic models is so important [6–8].

The similar performance of the biomarkers at diagnosing FL and NAFLD in the present study is likely to be partially explained by their underlying mutual association (Figure 3). This finding, which awaits replication in other populations, suggests that the same set of predictors may be employed to develop a common algorithm for the prediction of FL. The re-estimation of some or all regression coefficients or the updating of the algorithm with new predictors should be done only if it increases its performance [7].

5. Conclusions

In conclusion, FLI, LAP, ZJU, and HSI show similar performance at diagnosing FL and NAFLD in the Bagnacavallo population, even if FLI has greater discrimination. FLI, LAP, ZJU, and HSI are strongly associated, which is likely to explain their similar performance. Further studies are needed to evaluate the use of these surrogate indices for the diagnosis of MAFLD [31], the diagnostic entity which is expected to attract much attention in coming years [2–4]. We hope that the findings of the present study will stimulate the routine reporting of calibration in future studies of surrogate indices of fatty liver because it is the use of this metric that made it possible to detect their similar performance in the Bagnacavallo population.

Author Contributions: Conceptualization, F.G.F., M.D., P.G., A.C.-G. and G.B.; Data curation, F.C., M.D., A.C.-G. and G.B.; Formal analysis, G.B.; Funding acquisition, F.G.F. and C.T.; Investigation, F.G.F., F.C., M.D., P.G., A.B., V.B., L.N., D.B., M.A., A.C., G.E. and A.C.-G.; Methodology, G.B.; Project administration, F.G.F., P.G. and C.T.; Software, G.B.; Writing—original draft, G.B.; Writing—review and editing, F.G.F., S.B., A.G., C.T. and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: The present analysis was sponsored by a research grant from Gilead (Milan, Italy) to the Italian Liver Foundation (Basovizza, Trieste, Italy). The Sponsor had no role in the design and

conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation and review of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Area Vasta Romagna-IRST (reference number 112).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Acknowledgments: The Bagnacavallo Study Group: Pietro Andreone, Anna Chiara Dall’Aglio, Mauro Bernardi, Lauro Bucchi, Francesca Dazzani, Fabio Falcini, Arianna Lanzi, Alessandra Ravaioli, Margherita Rimini, Giulia Rovesti, Gaia Saini, Giuseppe Francesco Stefanini.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) checklist [6].

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	2
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	2
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	2
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	2
	5b	Describe eligibility criteria for participants.	2
	5c	Give details of treatments received, if relevant. Comment: no treatment was administered outside good clinical practice.	NAP

Table A1. Cont.

Section/Topic	Item	Checklist Item	Page
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	3
	6b	Report any actions to blind assessment of the outcome to be predicted. Comment: biomarkers were calculated after the end of the study.	NAP
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	13
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors. Comment: biomarkers were calculated after the end of the study.	NAP
Sample size	8	Explain how the study size was arrived at.	3
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	3
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	13
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done. Comment: there was no need to recalibrate any biomarker.	NAP
Risk groups	11	Provide details on how risk groups were created, if done.	4
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	3, 12
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	2
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	2
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	2
Model performance	16	Report performance measures (with CIs) for the prediction model.	5
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance). No attempt was made to update any biomarker.	NAP

Table A1. *Cont.*

Section/Topic	Item	Checklist Item	Page
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	9
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	9
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	9
Implications	20	Discuss the potential clinical use of the model and implications for future research.	9
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets. Comment: The data presented in this study are available on reasonable request from the corresponding author.	NAP
Funding	22	Give the source of funding and the role of the funders for the present study.	9

NAP = not applicable.

Appendix B

Table A2. Predictors employed by the biomarkers.

	Biomarkers			
	FLI	LAP	HSI	ZJU
Triglycerides	✓	✓		✓
BMI	✓		✓	✓
GGT	✓			
Waist	✓	✓		
ALT:AST			✓	✓
T2DM			✓	
Sex			✓	
Glucose				✓

Abbreviations: BMI = body mass index; GGT = gamma-glutamyltransferase; ALT = alanine transaminase; AST = aspartate transaminase; T2DM = type 2 diabetes mellitus.

Table A3. Multivariable Prediction Models Employed by the Biomarkers.

Abbreviations and Units of Measurements	
altsalt	alanine transaminase (U/L)/aspartate transaminase (U/L)
bmi	body mass index (kg/m ²)
exp	exponential operator
female	female gender (1 = female; 0 = male)
ggt	gamma-glutamyltransferase (U/L)
gmmol	glucose (mmol/L)
log _e	natural logarithm
fli_lp	fatty liver index-linear predictor

Table A3. Cont.

Abbreviations and Units of Measurements	
t2dm	type 2 diabetes mellitus (1 = yes; 0 = no)
tg	triglycerides (mg/dL)
tgmmol	triglycerides (mmol/L)
wc	waist circumference (cm)
Fatty liver index (FLI)	
fli_lp	$0.953 \cdot \log_e(\text{tg}) + 0.139 \cdot \text{bmi} + 0.718 \cdot \log_e(\text{ggT}) + 0.053 \cdot \text{wc} - 15.745$
FLI	$[(\exp(\text{fli_lp}) / (1 + \exp(\text{fli_lp})))] \cdot 100$
Lipid accumulation product (LAP)	
LAP	$(\text{wc} - k) \cdot \text{tgmmol}$ with $k=65$ if sex=male or $k=58$ if sex=female
Hepatic steatosis index (HSI)	
HSI	$8 \cdot \text{altast} + \text{bmi} + 2 \cdot \text{t2dm} + 2 \cdot \text{female}$
Zhejiang University index (ZJU)	
ZJU	$\text{bmi} + \text{gmmol} + \text{tgmmol} + 3 \cdot \text{altast} + 2 \cdot \text{female}$

References

- Fouad, Y.; Waked, I.; Bollipo, S.; Gomaa, A.; Ajlouni, Y.; Attia, D. What's in a name? Renaming 'NAFLD' to 'MAFLD'. *Liver Int.* **2020**, *40*, 1254–1261. [[CrossRef](#)]
- Eslam, M.; Newsome, P.N.; Anstee, Q.M.; Targher, G.; Gomez, M.R.; Zelber-Sagi, S.; Wong, V.W.-S.; Dufour, J.-F.; Schattenberg, J.; Arrese, M. A new definition for metabolic associated fatty liver disease: An international expert consensus statement. *J. Hepatol.* **2020**, *73*, 202–209. [[CrossRef](#)] [[PubMed](#)]
- Eslam, M.; Sanyal, A.J.; George, J.; Sanyal, A.; Neuschwander-Tetri, B.; Tiribelli, C.; Kleiner, D.E.; Brunt, E.; Bugianesi, E.; Yki-Järvinen, H.; et al. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. *Gastroenterology* **2020**, *158*, 1999–2014. [[CrossRef](#)] [[PubMed](#)]
- Ratziu, V.; Rinella, M.; Beuers, U.; Loomba, R.; Anstee, Q.M.; Harrison, S.; Francque, S.; Sanyal, A.; Newsome, P.N.; Younossi, Z. The times they are a-changin' (for NAFLD as well). *J. Hepatol.* **2020**, *73*, 1307–1309. [[CrossRef](#)] [[PubMed](#)]
- European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL–EASD–EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J. Hepatol.* **2016**, *64*, 1388–1402. [[CrossRef](#)] [[PubMed](#)]
- Moons, K.G.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [[CrossRef](#)] [[PubMed](#)]
- Steyerberg, E.W. *Clinical Prediction Models*; Springer International: Cham, Switzerland, 2019.
- Van Calster, B.; Nieboer, D.; Vergouwe, Y.; De Cock, B.; Pencina, M.J.; Steyerberg, E.W. A calibration hierarchy for risk models was defined: From utopia to empirical data. *J. Clin. Epidemiol.* **2016**, *74*, 167–176. [[CrossRef](#)] [[PubMed](#)]
- Shen, Y.N.; Yu, M.X.; Gao, Q.; Li, Y.Y.; Huang, J.J.; Sun, C.M.; Qiao, N.; Zhang, H.X.; Wang, H.; Lu, Q.; et al. External validation of non-invasive prediction models for identifying ultrasonography-diagnosed fatty liver disease in a Chinese population. *Medicine* **2017**, *96*, e7610. [[CrossRef](#)] [[PubMed](#)]
- Meffert, P.J.; Baumeister, S.E.; Lerch, M.M.; Mayerle, J.; Kratzer, W.; Volzke, H. Development, external validation, and comparative assessment of a new diagnostic score for hepatic steatosis. *Am. J. Gastroenterol.* **2014**, *109*, 1404–1414. [[CrossRef](#)]
- Wan, F.; Pan, F.; Ayonrinde, O.T.; Adams, L.A.; Mori, T.A.; Beilin, L.J.; O'Sullivan, T.A.; Olynyk, J.K.; Oddy, W.H. Validation of fatty liver disease scoring systems for ultrasound diagnosed non-alcoholic fatty liver disease in adolescents. *Dig. Liver Dis.* **2020**. [[CrossRef](#)]
- Foschi, F.G.; Domenicali, M.; Giacomoni, P.; Dall'Aglio, A.C.; Conti, F.; Borghi, A.; Bevilacqua, V.; Napoli, L.; Mirici, F.; Cucchetti, A.; et al. Is there an association between commonly employed biomarkers of liver fibrosis and liver stiffness in the general population. *Ann. Hepatol.* **2020**, *19*, 380–387. [[CrossRef](#)] [[PubMed](#)]
- Foschi, F.G.; Bedogni, G.; Domenicali, M.; Giacomoni, P.; Dall'Aglio, A.C.; Dazzani, F.; Lanzi, A.; Conti, F.; Savini, S.; Saini, G.; et al. Prevalence of and risk factors for fatty liver in the general population of Northern Italy: The Bagnacavallo Study. *BMC Gastroenterol.* **2018**, *18*, 177. [[CrossRef](#)] [[PubMed](#)]
- Rimini, M.; Casadei-Gardini, A.; Ravaioli, A.; Rovesti, G.; Conti, F.; Borghi, A.; Dall'Aglio, A.C.; Bedogni, G.; Domenicali, M.; Giacomoni, P.; et al. Could Inflammatory Indices and Metabolic Syndrome Predict the Risk of Cancer Development? Analysis from the Bagnacavallo Population Study. *J. Clin. Med.* **2020**, *9*, 1177. [[CrossRef](#)] [[PubMed](#)]
- Bedogni, G.; Miglioli, L.; Masutti, F.; Tiribelli, C.; Marchesini, G.; Bellentani, S. Prevalence of and risk factors for nonalcoholic fatty liver disease: The Dionysos nutrition and liver study. *Hepatology* **2005**, *42*, 44–52. [[CrossRef](#)] [[PubMed](#)]

16. Lohman, T.G.; Roche, A.F.; Martorell, R. *Anthropometric Standardization Reference Manual*; Human Kinetics Books: Champaign, IL, USA, 1991.
17. Bedogni, G.; Kahn, H.S.; Bellentani, S.; Tiribelli, C. A simple index of lipid overaccumulation is a good marker of liver steatosis. *BMC Gastroenterol.* **2010**, *10*, 98. [[CrossRef](#)] [[PubMed](#)]
18. National Institutes of Health. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. The Evidence Report. National Institutes of Health. *Obes. Res.* **1998**, *6* (Suppl. 2), 51S–209S.
19. Alberti, K.G.; Eckel, R.H.; Grundy, S.M.; Zimmet, P.Z.; Cleeman, J.I.; Donato, K.A.; Fruchart, J.C.; James, W.P.; Loria, C.M.; Smith, S.C.; et al. Harmonizing the Metabolic Syndrome A Joint Interim Statement of the International Diabetes Federation Task Force on Epidemiology and Prevention National Heart, Lung, and Blood Institute American. *Circulation* **2009**, *120*, 1640–1645. [[CrossRef](#)]
20. Bedogni, G.; Bellentani, S.; Miglioli, L.; Masutti, F.; Passalacqua, M.; Castiglione, A.; Tiribelli, C. The Fatty Liver Index: A simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol.* **2006**, *6*, 33. [[CrossRef](#)]
21. Lee, J.H.; Kim, D.; Kim, H.J.; Lee, C.H.; Yang, J.I.; Kim, W.; Kim, Y.J.; Yoon, J.H.; Cho, S.H.; Sung, M.W.; et al. Hepatic steatosis index: A simple screening tool reflecting nonalcoholic fatty liver disease. *Dig. Liver Dis.* **2010**, *42*, 503–508. [[CrossRef](#)]
22. Wang, J.; Xu, C.; Xun, Y.; Lu, Z.; Shi, J.; Yu, C.; Li, Y. ZJU index: A novel model for predicting nonalcoholic fatty liver disease in a Chinese population. *Sci. Rep.* **2015**, *5*, 16494. [[CrossRef](#)]
23. Otgonsuren, M.; Estep, M.J.; Hossain, N.; Younossi, E.; Frost, S.; Henry, L.; Hunt, S.; Fang, Y.; Goodman, Z.; Younossi, Z.M. A single non-invasive model to diagnose non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH). *J. Gastroenterol. Hepatol.* **2014**, *29*, 2006–2013. [[CrossRef](#)] [[PubMed](#)]
24. Poynard, T.; Ratziau, V.; Naveau, S.; Thabut, D.; Charlotte, F.; Messous, D.; Capron, D.; Abella, A.; Massard, J.; Ngo, Y.; et al. The diagnostic value of biomarkers (SteatoTest) for the prediction of liver steatosis. *Comp Hepatol.* **2005**, *4*, 10. [[CrossRef](#)] [[PubMed](#)]
25. Kotronen, A.; Peltonen, M.; Hakkarainen, A.; Sevastianova, K.; Bergholm, R.; Johansson, L.M.; Lundbom, N.; Rissanen, A.; Ridderstråle, M.; Groop, L.; et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* **2009**, *137*, 865–872. [[CrossRef](#)] [[PubMed](#)]
26. Stevens, R.J.; Poppe, K.K. Validation of clinical prediction models: What does the “calibration slope” really measure. *J. Clin. Epidemiol.* **2019**. [[CrossRef](#)]
27. Harrell, F. *Regression Modeling Strategies*; Springer: Cham, Switzerland, 2016.
28. Ensor, J.; Snell, K.I.; Martin, E.C. PMCALPLOT: Stata Module to Produce Calibration Plot of Prediction Model Performance. Statistical Software Components. 2018. Available online: <https://ideas.repec.org/c/boc/bocode/s458486.html> (accessed on 28 January 2021).
29. Haghish, E.F. Seamless interactive language interfacing between R and Stata. *Stata J.* **2019**, *19*, 61–82. [[CrossRef](#)]
30. Hernaez, R.; Lazo, M.; Bonekamp, S.; Kamel, I.; Brancati, F.L.; Guallar, E.; Clark, J.M. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: A meta-analysis. *Hepatology* **2011**, *54*, 1082–1090. [[CrossRef](#)]
31. Xu, Z.; Li, H.; Tian, S.; Wu, J.; Li, X.; Liu, Z.L.; Li, S.; Chen, Y.L.; Xiao, J.; Wei, J.Y.; et al. Blood biomarkers for the diagnosis of hepatic steatosis in metabolic dysfunction-associated fatty liver disease. *J. Hepatol.* **2020**, *73*, 1264–1265. [[CrossRef](#)]