Diabetes, Metabolic Syndrome and Obesity

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Development and Validation of a Risk Prediction Model for NAFLD: A Study Based on a Physical Examination Population

Chunmei Yang[1,2], Tingwan Du[1], Yueying Zhao[1], Youhui Qian[1], Jiashi Tang[1], Xiaohong Li[2,*], Ling Ma[1,3,*]

[1]Department of Nutrition and Food Hygiene, School of Public Health, Southwest Medical University, Luzhou, 646000, People's Republic of China; [2]Health Management Center, The Affiliated Hospital, Southwest Medical University, Luzhou, 646000, People's Republic of China; [3]Environmental Health Effects and Risk Assessment Key Laboratory of Luzhou, School of Public Health, Southwest Medical University, Luzhou, 646000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiaohong Li; Ling Ma, Email 1848922774@qq.com; xjml@swmu.edu.cn

Purpose: To construct and validate a precise and personalized predictive model for non-alcoholic fatty liver disease (NAFLD) to enhance NAFLD screening and healthcare administration.

Patients and Methods: A total of 730 participants' clinical information and outcome measurements were gathered and randomly divided into training and validation sets in a ratio of 3:7. Using the least absolute shrinkage and selection operator (LASSO) regression and multiple logistic regression, a nomogram was established to select risk predictor variables. The NAFLD prediction model was validated through the receiver operating characteristic (ROC) curve, calibration plot, and decision curve analysis (DCA).

Results: After random grouping, the cohort comprised 517 in the training set and 213 in the validation set. The prediction model employed nine of the 20 selected variables, namely gender, hypertension, waist circumference, body mass index, blood platelet, triglycerides, high-density lipoprotein cholesterol, plasma glucose, and alanine aminotransferase. ROC curve analysis yielded an area under the curve values of 0.877 (95% Confidence Interval [CI]: 0.848–0.907) for the training set and 0.871 (95% CI: 0.825–0.917) for the validation set. Optimal critical values were determined as 0.472 (0.786, 0.825) in the training set and 0.457 (0.743, 0.839) in the validation set. Calibration curves for both sets showed proximity to the ideal diagonal, with P-values of 0.972 and 0.370 for the training and validation sets, respectively ($P > 0.05$). DCA indicated favorable clinical applicability of the model.

Conclusion: We constructed a nomogram model that could complement traditional NAFLD detection methods, aiding in individualized risk assessment for NAFLD.

Keywords: non-alcoholic fatty liver disease, nomogram, prediction model, LASSO

Introduction

Non-alcoholic fatty liver disease (NAFLD) encompasses a range of liver conditions, including non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH), potentially leading to cirrhosis and hepatocellular carcinoma.[1] Globally, approximately 30% of the population is affected by NAFLD, with recent data indicating a surge in prevalence from 23.8% in 2001 to 32.9% in 2018 in china.[2,3] The increasing incidence of NAFLD and its intricate association with metabolic dysfunction, underscores the risk of cirrhosis in patients lacking timely diagnosis.[1] Hence, it is imperative to prioritize prevention and screening measures for NAFLD.

Current monitoring techniques for NAFLD include liver biopsy, ultrasound, vibration-controlled transient elastography, and magnetic resonance imaging (MRI), among others.[3–6] However, these methods have their limitations, such as the invasive nature of liver biopsy despite being the gold standard for diagnosis and the high cost of MRI.[3–6] Given these limitations, exploring supplementary predictive models alongside conventional testing methods is justified.

Received: 4 September 2023
Accepted: 19 December 2023
Published: 8 January 2024

Diagnostic predictive models aim to diagnose prevalent diseases by analyzing the clinical symptoms and characteristics of the study population.[7] The nomogram is a visual prediction model that can present complex statistical models and associated risk factors as risk scores, aiding communication between physicians and patients.[8,9] Moreover, the nomogram is widely used in oncology, reproductive medicine, cardiovascular and cerebrovascular disorders, orthopedic chronic conditions, and other medical ailments.[8] Vitamin D (VD), a fat-soluble vitamin, deficiency is associated with various diseases.[10] In a previous study, VD deficiency was speculated to be a risk factor for patients with NAFLD.[11] Meanwhile, through reviewing the literature, controversies about the correlation between platelet count and NAFLD were further uncovered.[12,13] Despite the use of nomograms in NAFLD studies, their potential inclusion of VD and platelets as predictors remains unexplored.[14,15] Therefore, further studies are warranted to investigate if serum 25(OH)D_3 and platelets can be used as predictors in NAFLD prediction models.

In this study, we incorporated two additional indicators, serum 25(OH)D_3 and platelets, to the NAFLD prediction model, in addition to physical examination indicators mentioned previously. The aim is to investigate whether a superior predictive model can be devised for NAFLD and to suggest new approaches for identifying those with a higher risk of NAFLD.

Materials and Methods

Research Subjects

Participants were recruited from the Health Management Center of The Affiliated Hospital, Southwest Medical University in China between April 2019 and September 2022. Based on the Guidelines for the Prevention and Treatment of Non-alcoholic Fatty Liver Disease (2018 update),[16] NAFLD was diagnosed based on the following criteria: participants were between the ages of 18 and 70; diagnosis of fatty liver disease was via ultrasound with alcohol intake ≤140 g/week in women and ≤210 g/week in men in the previous 12 months. Total alcohol intake was calculated based on how much beer, grape wine, and liquor were consumed (Alcohol intake (beer) = bottle * 600 mL * 0.04, Alcohol intake (grape wine) = Liang * 50 mL * 0.1, Alcohol intake (liqueur) = Liang * 50 mL * 0.38).[17] Exclusion criteria encompassed pregnant or breastfeeding women, renal disease, and medication affecting the liver or VD metabolism.[11] The sample size for this study was determined using the logistic regression sample size formula proposed by Peduzzi et al (N = 10 k/p),[18,19] where k represents the number of independent variables, which was 20 in this study, and p represents the proportion of NAFLD outcomes, calculated to be 52% in this study. Based on these values, a sample size of 384 cases was deemed necessary. However, the sample size exceeded this value, amounting to 730 in the current study. This study was approved by the Ethical Research Committee of The Affiliated Hospital, Southwest Medical University (NO. KY2023357). All procedures were performed in accordance with the Declaration of Helsinki.

Methods

Participants underwent a morning physical examination following 12h of dietary restriction, 24h of refraining from strenuous physical activity, and 48h of abstaining from alcohol consumption. Data including demographic characteristics, physical examination details, serum parameters, and ultrasound findings were collected from the medical examination database. Variables such as gender, age, waist circumference (WC), body mass index (BMI), systolic blood pressure-(SBP), diastolic blood pressure (DBP), past and present medical history (hypertension, diabetes, NAFLD), serum 25(OH)D_3, blood platelet (PLT), albumin (ALB), aspartate aminotransferase (AST), alanine aminotransferase(ALT), γ-glutamine transferase (GGT), alkaline phosphatase (ALP), plasma glucose (GLU), total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and low density lipoprotein cholesterol (LDL-C) were gathered. The corresponding continuous variables were transformed into categorical variables based on clinical relevance and practicality. According to the Working Group on Obesity in China, individuals were classified as non-overweight (BMI <24kg/m^2), overweight (24kg/m^2≤ BMI <28kg/m^2), and obese (BMI ≥28kg/m^2).[20] Individual serum VD levels were categorized as VD deficiency (VD <12ng/mL), VD insufficiency (12ng/mL ≤ VD <20ng/mL), and VD sufficiency (VD ≥20ng/mL).[11] Furthermore, high TC, high TG, high LDL-C, and low HDL-C were defined as TC ≥ 5.2 mmol/L, TG ≥ 1.7 mmol/L, LDL-C ≥ 3.4 mmol/L, and HDL-C <1 mmol/L, respectively.[21] High SPB and high DBP were also defined as SPB ≥ 140

mmHg and DBP ≥ 90 mmHg.[22] High GLU was defined as GLU ≥ 6.1 mmol/L.[23] The remaining numerical data were analyzed in quartiles (25%, 50%, 75%) for analysis purposes.

Statistical Methods

Data analysis utilized SPSS 26.0 and R 4.2.3 software with significance set at P-values <0.05. The training set of 517 individuals was randomly selected from the 730 individuals, with the remaining individuals forming the validation set. Descriptive statistics involved percentages for categorical variables. LASSO regression utilizing the R package "glmnet" mitigated the risk of overfitting and selected independent variables in the training set. Logistic regression analysis was employed to construct the predictive model due to the dichotomous nature of the dependent variable in this study. Multivariate logistic regression was analyzed using backward LR. Moreover, assumptions for regression analysis were met. The risk indicators identified by logistic regression contributed to a nomogram model creation using the R package "rms". The ROC and calibration curves were plotted using the R package "pROC" and "rms", respectively, to assess the discrimination and calibration of the model. Additionally, DCA was performed using the R package "rmda" to confirm the model's clinical usefulness.

Results

Participates' Characteristics

A total of 730 participants were randomized in a 7:3 ratio between the training (n = 517) and validation (n = 213) sets. No significant overall differences in clinical data were observed between these sets ($P > 0.05$) (Table 1).

Table 1 Clinical Data Comparison Between the Two Groups [n (%)]

Characteristics	Training Set (n=517)	Validation Set (n=213)	P-value
NAFLD	269 (52.00)	112 (52.60)	0.892
Gender, Male	260 (50.30)	95 (44.60)	0.162
Hypertension	98 (19.00)	43 (20.20)	0.701
Diabetes	27 (5.20)	18 (8.50)	0.099
SBP <140 mmHg	460 (89.00)	187 (87.80)	0.648
DBP < 90 mmHg	478 (92.50)	193 (90.60)	0.405
TC <5.2 mmol/L	305 (59.00)	119 (55.90)	0.437
TG <1.7 mmol/L	315 (60.90)	113 (53.10)	0.049
HDL-C <1 mmol/L	89 (17.20)	42 (19.70)	0.423
LDL-C <3.4 mmol/L	341 (66.00)	142 (66.70)	0.854
GLU <6.1 mmol/L	452 (87.40)	188 (88.30)	0.755
Age (years)			0.477
≤38	132 (25.53)	56 (26.30)	
39–48	118 (22.82)	39 (18.30)	
49–55	122 (23.60)	59 (27.70)	
≥56	145 (28.05)	59 (27.70)	
BMI (kg/m^2)			0.979
<24	227 (43.90)	93 (43.70)	

(*Continued*)

Table 1 (Continued).

Characteristics	Training Set (n=517)	Validation Set (n=213)	P-value
24–28	201 (38.90)	82 (38.50)	
≥28	89 (17.20)	38 (17.80)	
WC (cm)			0.911
≤76	134 (25.92)	51 (23.94)	
77–82	107 (20.70)	45 (21.13)	
83–89	132 (25.53)	59 (27.70)	
≥90	144 (27.85)	58 (27.23)	
VD (ng/mL)			0.517
<12	71 (13.70)	32 (15.02)	
12–20	291 (56.30)	110 (51.64)	
≥20	155 (30.00)	71 (33.33)	
PLT (10^9/L)			0.323
≤195.75	121 (23.40)	61 (28.60)	
195.76–231.99	136 (26.30)	46 (21.60)	
232.00–269.99	127 (24.60)	56 (26.30)	
≥270.00	133 (25.70)	50 (23.50)	
ALT (U/L)			0.572
≤17.70	137 (26.50)	46 (21.60)	
17.71–23.99	123 (23.80)	56 (26.30)	
24.00–34.02	130 (25.10)	56 (26.30)	
≥34.03	127 (24.60)	55 (25.80)	
AST (U/L)			0.593
≤19.60	138 (26.70)	47 (22.10)	
19.61–22.74	123 (23.80)	57 (26.80)	
22.75–27.82	128 (24.80)	55 (25.80)	
≥27.83	128 (24.80)	54 (25.40)	
GGT (U/L)			0.300
≤14.88	119 (23.02)	63 (29.58)	
14.89–22.09	133 (25.73)	49 (23.00)	
22.10–35.72	135 (26.11)	49 (23.00)	
≥35.73	130 (25.15)	52 (24.41)	

(*Continued*)

Table 1 (Continued).

Characteristics	Training Set (n=517)	Validation Set (n=213)	P-value
ALB (g/L)			0.514
≤44.80	138 (26.70)	49 (23.00)	
44.81–46.49	124 (24.00)	47 (22.10)	
46.50–48.19	130 (25.10)	56 (26.30)	
≥48.20	125 (24.20)	61 (28.60)	
ALP (U/L)			0.080
≤58.98	128 (24.80)	54 (25.40)	
58.99–71.39	129 (25.00)	53 (24.90)	
71.40–87.04	142 (27.50)	42 (19.70)	
≥87.05	118 (22.82)	64 (30.00)	

Abbreviations: NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; WC, waist circumference; VitD, vitamin D; PLT, blood platelet; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; GLU, plasma glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ-glutamine transferase; ALB, albumin; ALP, alkaline phosphatase.

Independent Predictors in the Training Set

Utilizing the LASSO regression model, 20 variables were included from the training set to construct the LASSO regression path and cross-check diagrams (Figure 1). Nine variables with nonzero coefficients were identified for multivariate logistic regression analysis, revealing gender, hypertension, WC, PLT, BMI, TG, HDL-C, GLU, and ALT as independent predictors of NAFLD (Table 2).

Developing and Validation of the Predictive Nomogram

Based on the outcomes of the above two models, a predictive nomogram was developed incorporating gender, hypertension, WC, PLT, BMI, TG, HDL-C, GLU, and ALT to visually represent NAFLD risk prediction (Figure 2). The model assigns

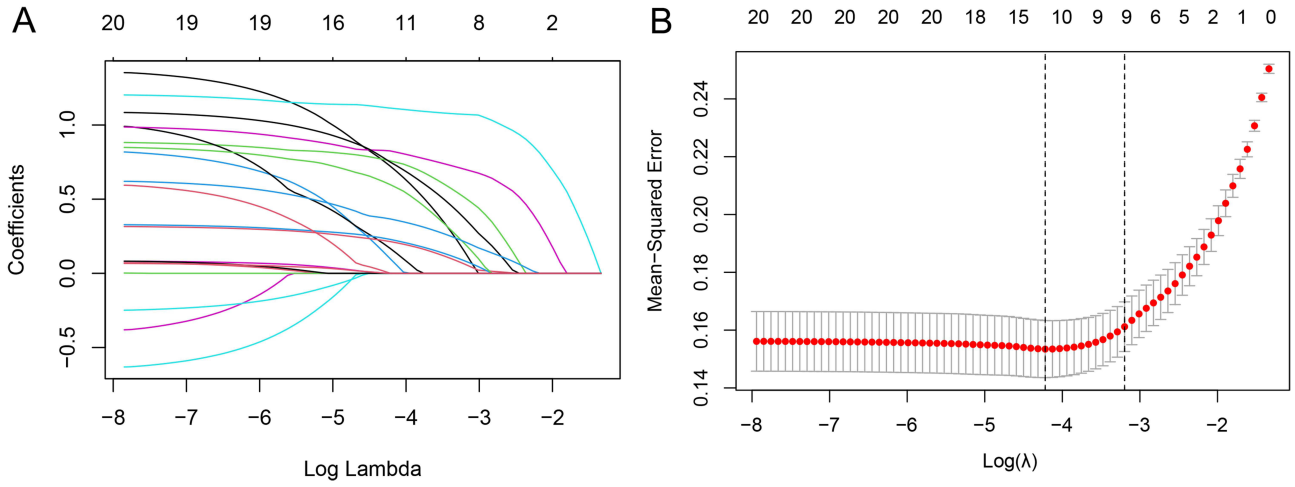

Figure 1 LASSO regression with 10-fold cross-validation for predictor selection. The LASSO coefficients for the 20 variables (**A**). Selecting the tuning parameter (lambda) of the deviation in the LASSO regression is determined by the minimum criteria (left dotted line) and the 1-SE criteria (right dotted line) (**B**). Based on 1-SE criteria (right dotted line), 9 nonzero coefficients were selected as predictors in the present study.
Abbreviations: LASSO, least absolute shrinkage and selection operator; SE, standard error.

Table 2 Risk Factors for NAFLD Based on Multivariate Logistic Regression

Characteristics	β	SE	P-value	OR (95% CI)
Gender				
Male			-	1.0 (ref)
Female	1.293	0.254	<0.001	3.644 (2.216–5.993)
Hypertension				
No			-	1.0 (ref)
Yes	0.796	0.283	0.005	2.217 (1.273–3.859)
TG				
<1.7			-	1.0 (ref)
≥1.7	0.888	0.216	<0.001	2.43 (1.593–3.708)
HDL-C				
<1			-	1.0 (ref)
≥1	−0.957	0.306	0.002	0.384 (0.211–0.699)
GLU				
<6.1			-	1.0 (ref)
≥6.1	1.016	0.394	0.010	2.761 (1.275–5.98)
BMI				
<24				1.0 (ref)
24–28	1.188	0.265	< 0.001	3.28 (1.952–5.513)
≥28	2.026	0.419	< 0.001	7.581 (3.332–17.247)
WC				
≤76			-	1.0 (ref)
77–82	0.394	0.31	0.204	1.482 (0.808–2.721)
83–89	0.885	0.348	0.011	2.423 (1.226–4.790)
≥90	1.339	0.433	0.002	3.814 (1.632–8.916)
PLT				
≤195.75			-	1.0 (ref)
195.76–231.99	0.899	0.288	0.002	2.457 (1.397–4.32)
232.00–269.99	0.939	0.291	0.001	2.558 (1.446–4.524)
≥270.00	1.014	0.288	< 0.001	2.758 (1.570–4.845)
ALT				
≤19.60			-	1.0 (ref)
19.61–22.74	0.429	0.289	0.317	1.536 (0.873–2.705)
22.75–27.82	1.052	0.29	< 0.001	2.862 (1.622–5.053)
≥27.83	1.516	0.315	< 0.001	4.553 (2.457–8.438)
Constant	−3.609	0.809	0.000	0.027

Abbreviations: SE, standard error; OR, odds ratio; ref, reference; CI, Confidence Interval; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; GLU, plasma glucose; BMI, body mass index; WC, waist circumference; PLT, blood platelet; ALT, alanine aminotransferase.

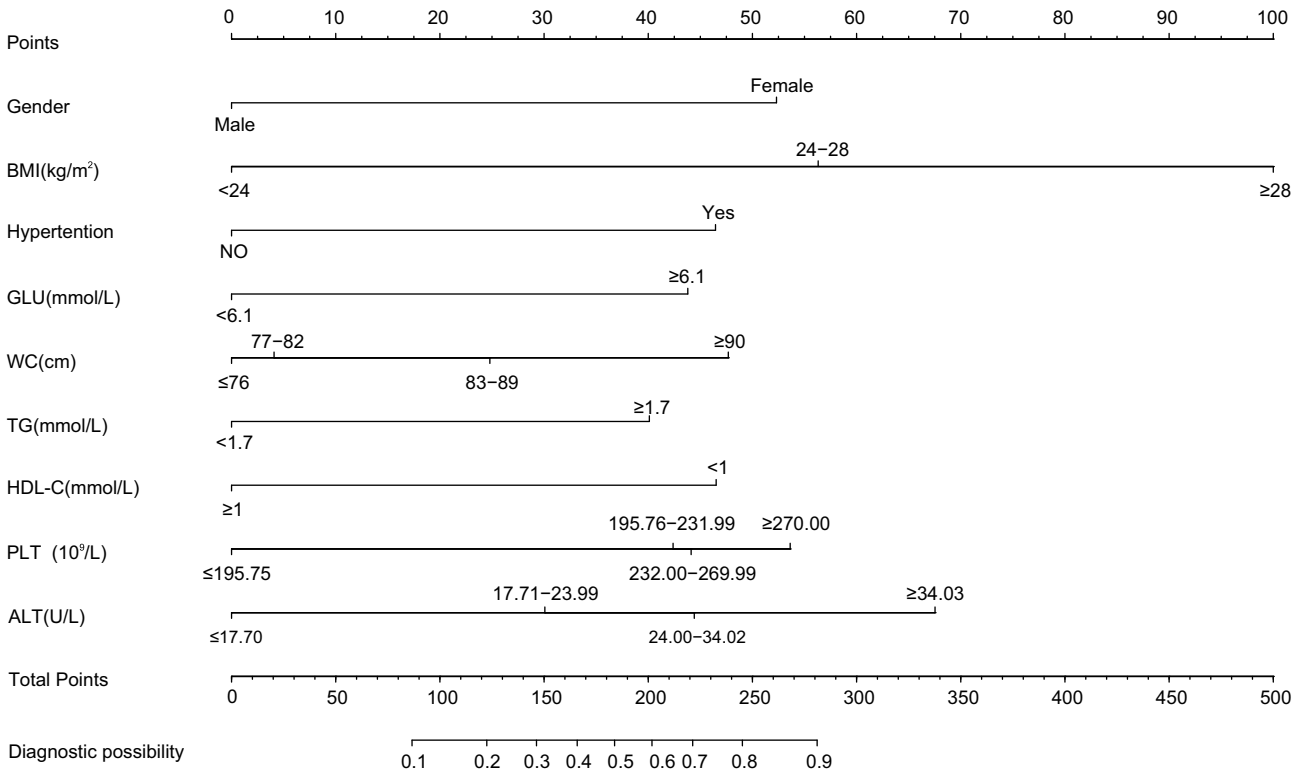

Figure 2 Nomogram for the prediction of NAFLD.
Abbreviations: NAFLD, non-alcoholic fatty liver disease; BMI, body mass index; WC, waist circumference; PLT, blood platelet; ALT, alanine aminotransferase; TG, triglycerides; GLU, plasma glucose.

small-scale scores for specific indicator values, ranging from 0 to 10. The cumulative score was derived by aggregating the scores acquired for each of the nine indicators in the physical examination data. The vertical alignment of an individual's total score corresponded to the diagnosis, indicating the probability of NAFLD occurrence in that individual.

Assessment of the prediction model via the ROC curves revealed (Figure 3) areas under the ROC curve for the training set and the validation set as 0.877 (95% CI: 0.848–0.907) and 0.871 (95% CI: 0.825–0.917), respectively. The

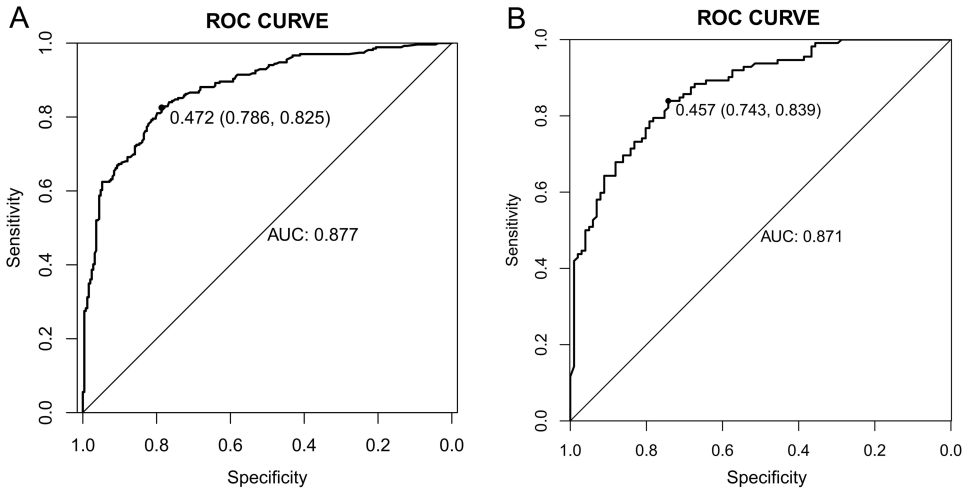

Figure 3 ROC curves. Training set (**A**). Validation set (**B**).
Abbreviations: ROC, receiver operating characteristic; AUC, the area under the ROC curve.

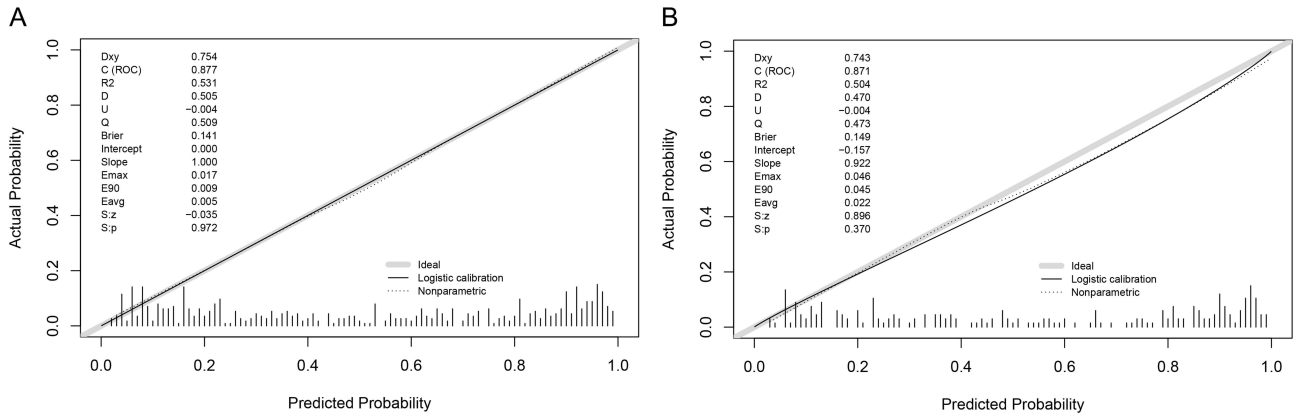

Figure 4 Calibration curve for predicting probability of NAFLD. Training set (**A**). Validation set (**B**).

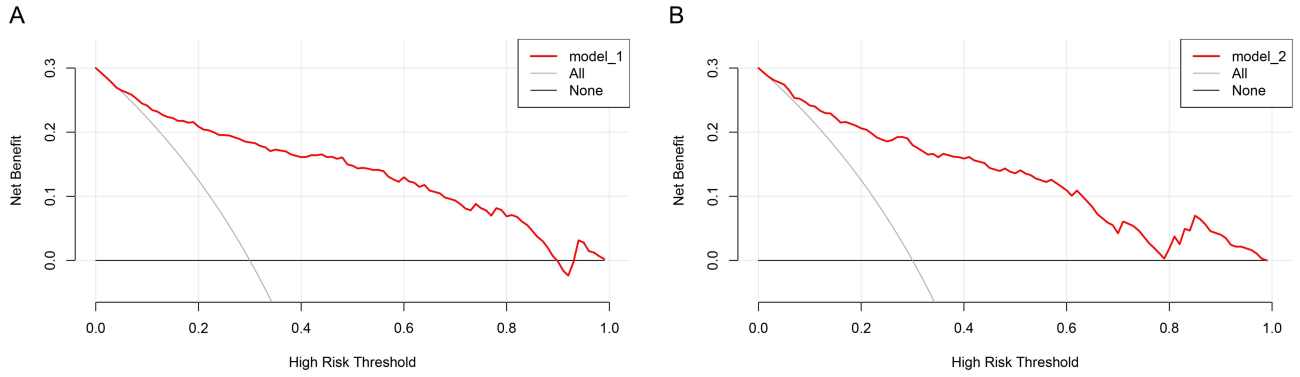

Figure 5 The decision curve analysis of NAFLD prediction. Training set (**A**). Validation set (**B**).

optimal critical values were determined as 0.472 (0.786, 0.825) in the training set and 0.457 (0.743, 0.839) in the validation set. These results suggest that this model may have good discriminative ability.

A calibration plot and Hosmer–Lemeshow test were used for the calibration of the nomogram model. The results demonstrated close alignment of calibration curves for both the training and validation sets with the ideal diagonal, as evidenced by P-values of 0.972 and 0.370, respectively ($P > 0.05$) (Figure 4). This suggests that the model exhibits consistency with the observed data and may possesses a commendable calibration capability.

Evaluating the clinical utility through DCA (Figure 5) revealed favorable net gains in both the training and validation sets over a larger threshold range, suggesting the potential practical utility of the model in clinical settings.

Discussion

This study conducted among a southwest Chinese medical examination population utilized LASSO regression to identify nine factors, namely gender, hypertension, WC, PLT, BMI, HDL-C, TG, GLU, and ALT, as potential correlates of NAFLD risk. Our findings also suggest that serum 25(OH)D_3 may not serve as a predictive factor in NAFLD prediction models. Subsequently, we constructed a Nomogram model using commonly measured and easily accessible clinical and laboratory variables to discern the presence or absence of NAFLD.

Based on the study results, women were more likely to develop NAFLD, not aligning with previous research wherein women were reported to be less likely to develop NAFLD than men.[24,25] The small number of young women in our study could be an underlying factor. This gender disparity could be attributed to the protective impact of estrogen on the liver.[24,25] However, after the onset of NAFLD, particularly beyond the age of 50, women demonstrated a greater vulnerability to the emergence of advanced fibrosis compared to males, potentially attributable to the decline in estrogen

levels that expedited the progression of NAFLD.[24,25] Additionally, NAFLD might also differ by gender due to differences in obesity and metabolic risk factors.[26] Thus, gender differences in NAFLD and associated risk factors warrant exploration to enhance our understanding of disease risk and intervention effectiveness.

Metabolic Syndrome (MetS) emerges as a significant factor predisposing individuals to NAFLD and NASH.[1] The primary constituents of MetS encompass obesity and hypertension, and their manifestation increases the likelihood of cardiovascular disease, the predominant cause of mortality among patients with NAFLD.[27–29]

Our study underscores the strong association between NAFLD risk and BMI, highlighting obesity as a pivotal risk factor for the incidence of NAFLD. Visceral adipocyte alterations due to obesity play a crucial role in NAFLD pathogenesis, with increased adipocyte size correlating with severe liver histopathology, including but not limited to severe steatosis, activity, and fibrosis.[30] A meta-analysis demonstrated that obesity increases NAFLD risk 3.5-fold.[31] As a measure of obesity, BMI focuses on the relationship between weight and height, with a dose-dependent association with NAFLD risk.[31] Furthermore, obesity, characterized by higher BMI, significantly increases the genetic risk of NAFLD.[32] In both humans and animals, obesity or eating an obesogenic diet during pregnancy increases the risk of NAFLD and NASH in the newborn. Notably, it has become common in epidemiology to quantify central obesity using WC. Researchers reported that despite a lean BMI individuals with visceral obesity may be at greater risk of death from NAFLD.[33] In studies of middle-aged and older Koreans, WC was identified as the strongest risk factor for predicting NAFLD events.[34] Thus, the inclusion of BMI and WC in this investigation allows for the comprehensive measurement of the impact of obesity on the risk of NAFLD.

Hypertension, a complex interplay of environmental risk factors and genetic predisposition, shows a robust correlation with NAFLD.[35] For instance, Francesco et al reported hepatic steatosis in 57.5% of individuals diagnosed with hypertension.[36,37] Notably, previous research has indicated the presence of a potential reciprocal causal association between NAFLD and hypertension.[38] There are 13 genes that are shared between NAFLD and hypertension, with the gene associated with hypertension being in closer proximity to the NAFLD gene compared to a randomly selected gene.[38] This suggests a common underlying mechanism for the development of both NAFLD and hypertension, necessitating a multi-targeted treatment approach rather than a singular intervention. The reason for including a history of hypertension, rather than specific SBP or DBP levels, in the model is that healthcare providers remind patients to take their medication on time to avoid cardiovascular risk events due to high blood pressure levels during their physical examinations. As a result, most patients with hypertension are likely to have blood pressure levels within the normal range at the time of the physical examination. Thus, incorporating the historical context of hypertension into the model holds greater significance compared to the inclusion of SBP and DBP measurements. However, it is important to consider that a history of hypertension based on self-report may lead to an underestimate of the prevalence.

Elevated glucose levels pose a multifaceted risk for NAFLD through various mechanisms. Notably, enhanced intrahepatic triglyceride levels are one of the hallmarks of NAFLD. De novo lipogenesis (DNL) is considered an important modulator of intrahepatic triglyceride (IHTG) content and promotes steatosis in patients with NAFLD.[39] Hyperglycemia induces hepatic DNL via the transcription factor carbohydrate response-element binding protein (ChREBP).[39,40] Moreover, a loss of Nuclear factor erythroid-2-related factor 2 (Nrf2) activity caused liver lipid accumulation.[41] NAFLD and impaired hepatic lipid metabolism may be associated with hyperactivation of the angiotensin receptor type 1 (AT1). An animal experiment found that AT1 overactivation associated with hyperglycemia enhanced hepatic oxidative damage by chronically reducing nuclear levels of Nrf2.[42] However, in this study, specific blood glucose levels were included in the model rather than a history of diabetes as the undiagnosed rate of diabetes is high in China.[43]

NAFLD is linked to dyslipidemia, including hypertriglyceridaemia and low levels of HDL in this study. The liver plays a pivotal role in the metabolism of lipids and the synthesis of plasma lipoproteins.[44] Extensive dysregulation of cholesterol homeostasis in NAFLD has been reported to lead to increased cholesterol synthesis and uptake, alongside diminished cholesterol elimination, thereby culminating in elevated hepatic cholesterol levels.[44] A cohort study conducted in China revealed that patients with NAFLD exhibited a considerably higher incidence of dyslipidemia compared to those free of the disease.[45] Insulin resistance (IR) has a negative impact on lipid and lipoprotein metabolism and has been significantly linked to dyslipidemia. In NAFLD, IR is the primary causative factor for lipoprotein abnormalities,

which leads to an augmented accumulation of triacylglycerols in adipose tissue and the release of non-esterified fatty acids from adipose tissue.[46,47] One of the mechanisms by which IR lowers HDL-C is by increasing the exchange of TG and very low density lipoprotein cholesterol (VLDL) in celiac particles with cholesterol esters in HDL particles.[47] A correlation between heightened consumption of dietary cholesterol and the likelihood or intensity of NAFLD has been reported,[44] emphasizing the significance of restricting excessive cholesterol intake among individuals susceptible to NAFLD as a preventive measure against the disease.

Platelet count emerged as a significant predictor of NAFLD in the present study. Existing literature presents controversial speculations surrounding the relationship between NAFLD and platelet count. Certain investigations have reported a negative correlation between platelet count and NAFLD severity.[12,48] The potential mechanisms may involve the coexistence of cirrhosis accompanied by splenomegaly, as well as impaired platelet production in individuals with advanced liver disease.[48] Whereas others have demonstrated that platelet count serves as an independent predictor for NASH in individuals with NAFLD.[13] Platelet count has also been independently correlated with advanced liver fibrosis.[49] Platelets function as inflammatory agents, facilitating the creation of neutrophil extracellular traps (NET) within the liver.[50] This mechanism establishes a connection between infection, inflammation, and platelet response. Patients with NAFLD exhibit a pro-inflammatory phenotype in their circulating platelets, along with heightened platelet accumulation in the hepatic sinusoids, which is associated with localized NET deposition.[50] A prior study demonstrated the significant involvement of transforming growth factor-β1 (TGF-β1) signaling in hepatic stellate cells (HSC) during the progression of liver fibrosis.[51] Notably, platelets possess substantial levels of TGF-β1, and the TGF-β1 fraction within platelets appears to be the primary driver of liver fibrosis, functioning by initiating pro-fibrotic signaling pathways and stimulating collagen synthesis in HSC.[51] These findings provide compelling evidence that platelets exhibit considerable potential as a viable therapeutic target in NAFLD treatment.

In our previous study, low serum $25(OH)D_3$ levels were identified as a risk factor in patients with NAFLD with steatosis but without advanced fibrosis.[11] In the present study, after variable screening, serum $25(OH)D_3$ was not found to be a potential predictor of NAFLD. This inconsistency could be because our sample not only included patients with NAFLD who had steatosis and no fibrosis, but also included patients with NAFLD who had progressed to other disease stages. However, it is not possible to determine the disease state of NAFLD based on ultrasound diagnosis alone, necessitating the need for further evaluation of the potential predictive value of serum $25(OH)D_3$ in NAFLD in the future.

In most cases, NAFLD leads to increased liver enzyme levels, namely ALT, AST, and GGT, without any related symptoms.[52] Among these enzymes, ALT exhibits the strongest correlation with liver steatosis and serves as an independent risk factor for NAFLD.[53,54] Consequently, ALT is frequently employed as a predictor of NAFLD.[53] In humans, there exist two isoforms of ALT, namely ALT1 and ALT2, with ALT2 potentially accounting for the heightened ALT activity observed in hepatic steatosis.[55] Notably, the ALT2 isoform predominantly originates from adipose tissue in cases of IR and obesity.[55] Hence, the management of body weight may exert a positive influence on ALT levels in individuals with NAFLD. Nonetheless, it has been observed that NAFLD can also manifest in individuals exhibiting normal ALT levels,[56] indicating that ALT levels alone do not adequately predict NASH or advanced fibrosis.[57] Consequently, the integration of ALT with other indicators in this study model significantly enhanced the accuracy of predictions.

In a study by Cen et al, fewer variables were included in the nomogram and a larger sample size was used, but AUC values were lower.[14] Danting Li et al constructed a prognostic model for NAFLD in a population in northwestern China, but unlike our study, its clinical applicability was not evaluated.[15] In terms of model calibration, the P-value of the calibration curves in our training set is 0.972, which exceeds the P-value of 0.812 for the model training set of Li et al. Furthermore, the P-value of the calibration curve in our validation set was 0.370, also exceeding that of their validation set (0.109).[15] From a theoretical perspective, both the models we created and those developed by Li et al demonstrate noteworthy calibration capabilities. While our clinical applicability evaluation laid the theoretical groundwork, further studies are required to determine its practical utility.

A notable strength of this study lies in its pioneering development of a personalized NAFLD prediction model, specifically tailored for the population attending medical check-up centers in the southwest region. We observed that

platelet count has potential as an indicator for NAFLD predictive model but not VD. This population, often in a healthy or subhealthy state, tends to overlook timely screening for NAFLD, consequently leading to disease progression. Thus, this model may be used as a convenient screening tool for individuals susceptible to NAFLD.

Despite these strengths, limitations exist. Firstly, the identification of patients with NAFLD was conducted solely through ultrasound, which, while widely utilized for NAFLD screening in China, cannot assess inflammation or fibrosis when steatosis is below 20–30%. Consequently, this may result in an underestimation of the prevalence of NAFLD. Secondly, this study was initially conducted at one center, and the model's calibration and identification relied heavily on internal validation. The performance of the predictors could be influenced by the heterogeneity of different populations and ethnicities. Thus, multi-center collaborations and Cohort Study should be conducted for external validation to promote the degree of generalizability of the model. Thirdly, the study failed to encompass various influential factors related to NAFLD, including clinical indicators like IR assessment and lifestyle factors such as dietary patterns and physical activity. This omission impedes a comprehensive evaluation of the determinants contributing to the development of NAFLD.

Conclusion

In this investigation, we developed a predictive model for NAFLD based on routine physical examination indices and identified platelets as a potential predictor, while serum $25(OH)D_3$ lacked predictive capability in this context. It is worth noting that the use of ultrasonography may have underestimated the prevalence of NAFLD in this study. Furthermore, as this remains a single-center study, further external validation and practical application of the model are imperative to increase its generalizability and clinical utility.

Institutional Review Board Statement

This research received funding from a grant from the Department of Science and Technology of Sichuan Province (2019YJ0483). Furthermore, the study underwent a thorough review and received approval from the ethics research committee of The Affiliated Hospital, Southwest Medical University. Participants provided informed consent.

Acknowledgments

The authors express gratitude to the Sichuan Provincial Department of Science and Technology for the support accorded to the project. We thank all the study participants and colleagues who contributed to this analysis.

Disclosure

The authors of the study have no commercial or financial affiliations that could present a potential conflict of interest.

References

1. Friedman SL, Neuschwander-Tetri BA, Rinella M, et al. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med*. 2018;24(7):908–922. doi:10.1038/s41591-018-0104-9
2. Le MH, Le DM, Baez TC, et al. Global incidence of non-alcoholic fatty liver disease: a systematic review and meta-analysis of 63 studies and 1,201,807 persons. *J Hepatol*. 2023;79(2):287–295. doi:10.1016/j.jhep.2023.03.040
3. Zhou J, Zhou F, Wang W, et al. Epidemiological features of NAFLD from 1999 to 2018 in China. *Hepatology*. 2020;71(5):1851–1864. doi:10.1002/hep.31150
4. Dasarathy S, Dasarathy J, Khiyami A, et al. Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. *J Hepatol*. 2009;51(6):1061–1067. doi:10.1016/j.jhep.2009.09.001
5. Selvaraj EA, Mózes FE, Jayaswal ANA, et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis. *J Hepatol*. 2021;75(4):770–785. doi:10.1016/j.jhep.2021.04.044
6. Woo Baidal JA, Lavine JE. The intersection of nonalcoholic fatty liver disease and obesity. *Sci Transl Med*. 2016;8(323):323rv1. doi:10.1126/scitranslmed.aad8390
7. Zhou ZR, Wang WW, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann Transl Med*. 2019;7(23):796. doi:10.21037/atm.2019.08.63
8. Wang X, Lu J, Song Z, et al. From past to future: bibliometric analysis of global research productivity on nomogram (2000–2021). *Front Public Health*. 2022;10:997713. doi:10.3389/fpubh.2022.997713
9. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173–e180. doi:10.1016/S1470-2045(14)71116-7
10. Dominguez LJ, Farruggia M, Veronese N, Barbagallo M. Vitamin D sources, metabolism, and deficiency: available compounds and guidelines for its treatment. *Metabolites*. 2021;11(4):255. doi:10.3390/metabo11040255

11. Du T, Xiang L, Zhang J, et al. Vitamin D improves hepatic steatosis in NAFLD *via* regulation of fatty acid uptake and β-oxidation. *Front Endocrinol*. 2023;14:1138078. doi:10.3389/fendo.2023.1138078
12. Campos GM, Bambha K, Vittinghoff E, et al. A clinical scoring system for predicting nonalcoholic steatohepatitis in morbidly obese patients. *Hepatology*. 2008;47(6):1916–1923. doi:10.1002/hep.22241
13. Kozumi K, Kodama T, Murai H, et al. Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology*. 2021;74(5):2452–2466. doi:10.1002/hep.31995
14. Cen C, Wang W, Yu S, et al. Development and validation of a clinical and laboratory-based nomogram to predict nonalcoholic fatty liver disease. *Hepatol Int*. 2020;14(5):808–816. doi:10.1007/s12072-020-10065-7
15. Li D, Zhang M, Wu S, et al. Risk factors and prediction model for nonalcoholic fatty liver disease in northwest China. *Sci Rep*. 2022;12(1):13877. doi:10.1038/s41598-022-17511-6
16. Fatty Liver Expert Committee CMDA. Guidelines of prevention and treatment for nonalcoholic fatty liver disease: a 2018 update. *Infect Dis Info*. 2018;31(5):393–420. doi:10.3969/j.issn.1007-8134.2018.05.002
17. Zhu B, Li Y, Shi Y, et al. Long-term drinking behavior change patterns and its association with hyperuricemia in Chinese adults: evidence from China Health and Nutrition Survey. *BMC Public Health*. 2022;22(1):1230. doi:10.1186/s12889-022-13637-4
18. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373–1379. doi:10.1016/s0895-4356(96)00236-3
19. López-Gil JF, Smith L, Victoria-Montesinos D, et al. Mediterranean dietary patterns related to sleep duration and sleep-related problems among adolescents: the EHDLA study. *Nutrients*. 2023;15(3):665. doi:10.3390/nu15030665
20. Gao M, Wei YX, Lyu J, et al. The cut-off points of body mass index and waist circumference for predicting metabolic risk factors in Chinese adults. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2019;40(12):1533–1540. doi:10.3760/cma.j.issn.0254-6450.2019.12.006
21. Joint Committee for Developing Chinese guidelines on Prevention and Treatment of Dyslipi-demia in Adults. Guidelines for the prevention and treatment of dyslipidemia in Chinese adults (revised in 2016). *Chin Circ J*. 2016;31(10):937–950. doi:10.3969/j.issn.1000-3614.2016.10.001
22. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension-A report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. *J Geriatr Cardiol*. 2019;16(3):182–241. doi:10.11909/j.issn.1671-5411.2019.03.014
23. Zhan C, Shi M, Yang Y, et al. Prevalence and risk factors of carotid plaque among middle-aged and elderly adults in Rural Tianjin, China. *Sci Rep*. 2016;6(1):23870. doi:10.1038/srep23870
24. Balakrishnan M, Patel P, Dunn-Valadez S, et al. Women have a lower risk of nonalcoholic fatty liver disease but a higher risk of progression vs men: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2021;19(1):61–71. doi:10.1016/j.cgh.2020.04.067
25. Im HJ, Ahn YC, Wang J-H, et al. Systematic review on the prevalence of nonalcoholic fatty liver disease in South Korea. *Clin Res Hepatol Gastroenterol*. 2021;45(4):101526. doi:10.1016/j.clinre.2020.06.022
26. Lonardo A, Nascimbeni F, Ballestri S, et al. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. *Hepatology*. 2019;70(4):1457–1469. doi:10.1002/hep.30626
27. Hsu CN, Hou CY, Hsu WH, et al. Early-life origins of metabolic syndrome: mechanisms and preventive aspects. *Int J Mol Sci*. 2021;22(21):11872. doi:10.3390/ijms222111872
28. Marjot T, Moolla A, Cobbold JF, et al. Nonalcoholic fatty liver disease in adults: current concepts in etiology, outcomes, and management. *Endocr Rev*. 2020;41(1):bnz009. doi:10.1210/endrev/bnz009
29. Zhang QQ, Lu LG. Nonalcoholic fatty liver disease: dyslipidemia, risk for cardiovascular complications, and treatment strategy. *J Clin Transl Hepatol*. 2015;3(1):78–84. doi:10.14218/JCTH.2014.00037
30. Sun H, Fang D, Wang H, et al. The association between visceral adipocyte hypertrophy and NAFLD in subjects with different degrees of adiposity. *Hepatol Int*. 2023;17(1):215–224. doi:10.1007/s12072-022-10409-5
31. Li L, Liu DW, Yan HY, et al. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. *Obes Rev*. 2016;17(6):510–519. doi:10.1111/obr.12407
32. Schnurr TM, Katz SF, Justesen JM, et al. Interactions of physical activity, muscular fitness, adiposity, and genetic risk for NAFLD. *Hepatol Commun*. 2022;6(7):1516–1526. doi:10.1002/hep4.1932
33. Golabi P, Paik JM, Arshad T, et al. Mortality of NAFLD according to the body composition and presence of metabolic abnormalities. *Hepatol Commun*. 2020;4(8):1136–1148. doi:10.1002/hep4.1534
34. Lee JH, Jeon S, Lee HS, Kwon YJ. Cutoff points of waist circumference for predicting incident non-alcoholic fatty liver disease in middle-aged and older Korean adults. *Nutrients*. 2022;14(14):2994. doi:10.3390/nu14142994
35. Zhao YC, Zhao GJ, Chen Z, She ZG, Cai J, Li H. Nonalcoholic fatty liver disease: an emerging driver of hypertension. *Hypertension*. 2020;75(2):275–284. doi:10.1161/HYPERTENSIONAHA.119.13419
36. Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73–84. doi:10.1002/hep.28431
37. Fallo F, Dalla Pozza A, Sonino N, et al. Nonalcoholic fatty liver disease, adiponectin and insulin resistance in dipper and nondipper essential hypertensive patients. *J Hypertens*. 2008;26(11):2191–2197. doi:10.1097/HJH.0b013e32830dfe4b
38. Ma C, Yan K, Wang Z, et al. The association between hypertension and nonalcoholic fatty liver disease (NAFLD): literature evidence and systems biology analysis. *Bioengineered*. 2021;12(1):2187–2202. doi:10.1080/21655979.2021.1933302
39. Smith GI, Shankaran M, Yoshino M, et al. Insulin resistance drives hepatic de novo lipogenesis in nonalcoholic fatty liver disease. *J Clin Invest*. 2020;130(3):1453–1460. doi:10.1172/JCI134165
40. Stefan N, Cusi K. A global view of the interplay between non-alcoholic fatty liver disease and diabetes. *Lancet Diabetes Endocrinol*. 2022;10(4):284–296. doi:10.1016/S2213-8587(22)00003-1
41. Ramadori P, Drescher H, Erschfeld S, et al. Hepatocyte-specific Keap1 deletion reduces liver steatosis but not inflammation during non-alcoholic steatohepatitis development. *Free Radic Biol Med*. 2016;91:114–126. doi:10.1016/j.freeradbiomed.2015.12.014
42. Godoy-Lugo JA, Thorwald MA, Mendez DA, et al. Glucose increases hepatic mitochondrial antioxidant enzyme activities in insulin resistant rats following chronic angiotensin receptor blockade. *Int J Mol Sci*. 2022;23(18):10897. doi:10.3390/ijms231810897

43. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabet Res Clin Pract*. 2022;183:109119. doi:10.1016/j.diabres.2021.109119
44. Ioannou GN. The role of cholesterol in the pathogenesis of NASH. *Trends Endocrinol Metab*. 2016;27(2):84–95. doi:10.1016/j.tem.2015.11.008
45. Ren XY, Shi D, Ding J, et al. Total cholesterol to high-density lipoprotein cholesterol ratio is a significant predictor of nonalcoholic fatty liver: jinchang cohort study. *Lipids Health Dis*. 2019;18(1):47. doi:10.1186/s12944-019-0984-9
46. Meex RCR, Watt MJ. Hepatokines: linking nonalcoholic fatty liver disease and insulin resistance. *Nat Rev Endocrinol*. 2017;13(9):509–520. doi:10.1038/nrendo.2017.56
47. Bjornstad P, Eckel RH. Pathogenesis of lipid disorders in insulin resistance: a brief review. *Curr Diab Rep*. 2018;18(12):127. doi:10.1007/s11892-018-1101-6
48. Yoneda M, Fujii H, Sumida Y, et al. Platelet count for predicting fibrosis in nonalcoholic fatty liver disease. *J Gastroenterol*. 2011;46(11):1300–1306. doi:10.1007/s00535-011-0436-4
49. McPherson S, Henderson E, Burt AD, Day CP, Anstee QM. Serum immunoglobulin levels predict fibrosis in patients with non-alcoholic fatty liver disease. *J Hepatol*. 2014;60(5):1055–1062. doi:10.1016/j.jhep.2014.01.010
50. Miele L, Alberelli MA, Martini M, et al. Nonalcoholic fatty liver disease (NAFLD) severity is associated to a nonhemostatic contribution and proinflammatory phenotype of platelets. *Transl Res*. 2021;231:24–38. doi:10.1016/j.trsl.2020.11.003
51. Ghafoory S, Varshney R, Robison T, et al. Platelet TGF-β1 deficiency decreases liver fibrosis in a mouse model of liver injury. *Blood Adv*. 2018;2(5):470–480. doi:10.1182/bloodadvances.2017010868
52. Clark JM, Brancati FL, Diehl AM. The prevalence and etiology of elevated aminotransferase levels in the United States. *Am J Gastroenterol*. 2003;98(5):960–967. doi:10.1111/j.1572-0241.2003.07486.x
53. Westerbacka J, Cornér A, Tiikkainen M, et al. Women and men have similar amounts of liver and intra-abdominal fat, despite more subcutaneous fat in women: implications for sex differences in markers of cardiovascular risk. *Diabetologia*. 2004;47(8):1360–1369. doi:10.1007/s00125-004-1460-1
54. Lu ZY, Shao Z, Li YL, et al. Prevalence of and risk factors for non-alcoholic fatty liver disease in a Chinese population: an 8-year follow-up study. *World J Gastroenterol*. 2016;22(13):3663–3669. doi:10.3748/wjg.v22.i13.3663
55. Schindhelm RK, Diamant M, Dekker JM, et al. Alanine aminotransferase as a marker of non-alcoholic fatty liver disease in relation to type 2 diabetes mellitus and cardiovascular disease. *Diabetes Metab Res Rev*. 2006;22(6):437–443. doi:10.1002/dmrr.666
56. Verma S, Jensen D, Hart J, et al. Predictive value of ALT levels for non-alcoholic steatohepatitis (NASH) and advanced fibrosis in non-alcoholic fatty liver disease (NAFLD). *Liver Int*. 2013;33(9):1398–1405. doi:10.1111/liv.12226
57. Torres DM, Harrison SA. NAFLD: predictive value of ALT levels for NASH and advanced fibrosis. *Nat Rev Gastroenterol Hepatol*. 2013;10(9):510–511. doi:10.1038/nrgastro.2013.138