

# Development and validation of a nomogram for predicting nonalcoholic fatty liver disease in the non-obese Chinese population

**Jiao Wang**

First Affiliated Hospital of Nanchang University

**Yunliang Tang**

First Affiliated Hospital of Nanchang University

**Kaili Peng**

First People's Hospital of Yuhang District of Hangzhou

**Honghong Liu**

First Affiliated Hospital of Nanchang University

**Jixiong Xu** (✉ [xujixiongtg@163.com](mailto:xujixiongtg@163.com))

Nanchang University

---

## Research

**Keywords:** Predictive modeling, Nonalcoholic fatty liver disease, Risk factor, Nomogram

**Posted Date:** April 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-24981/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, and can be found in non-obese individuals. The purpose of this study was to construct and validate a model for predicting NAFLD in the non-obese Chinese population.

## Methods

A total of 13240 NAFLD-free individuals at baseline from a 4-y longitudinal study were allocated to a training cohort (n = 8872) and a validation cohort (n = 4368). Cox proportional hazards regression analyses were used to select significant variables, and a nomogram was developed to predict NAFLD incidence in the training cohort. Concordance index (C-index), receiver operating characteristic (ROC) curves and calibration curves were applied to assess the predictive modeling of the nomogram in training and validation cohorts. Individuals were categorized into high- or low-risk groups based on median risk scores from the nomogram. Then, a decision curve analysis (DCA) was used to estimate the clinical usefulness of the nomogram in the combined training and validation cohorts.

## Results

The overall incidence of NAFLD was 13%. Nine significant predictors including age, gender, body mass index (BMI), fasting blood glucose (FBG), total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-c), uric acid (UA) and alanine aminotransferase (ALT) were identified and constructed for the nomogram. The C-index was 0.804 [95% CI, 0.792–0.852] and 0.802 [95% CI, 0.784–0.820] in the training and validation cohorts, respectively. In the training cohort, the area under the ROC curve (AUC) for 1-y, 2-y, 3-y and 4-y risk was 0.835, 0.825, 0.816 and 0.782, respectively. Likewise, in the validation cohort, the AUC for 1-y, 2-y, 3-y and 4-y risk was 0.817, 0.820, 0.814 and 0.813. The calibration curves for NAFLD risk showed excellent accuracy in the predictive modeling of the nomogram, internally and externally. The nomogram categorized individuals into high- and low-risk groups, and the DCA displayed the clinical usefulness of the nomogram for predicting NAFLD incidence.

## Conclusions

Our nomogram can predict a personalized risk of NAFLD in the non-obese Chinese population. This nomogram can serve as a simple and affordable tool for stratifying individuals at a high risk of NAFLD, and thus serve to expedite treatment of NAFLD.

## Background

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, affecting approximately 25.24% of the global population [1]. The prevalence of NAFLD in South America is 31%, 32% in the Middle East, 14% in Africa, and 27% in Asian, varying from 24.77–43.91% in the Chinese population [2]. NAFLD is increasingly identified as one of the major causes of liver-related morbidity, mortality, and liver transplantation. In addition, NAFLD has been found to be indirectly associated with cardiovascular disease, type 2 diabetes mellitus (T2DM) and chronic kidney disease [3].

The prevalence of NAFLD is higher in cohorts with obesity, T2DM, dyslipidemia and metabolic syndrome. However, NAFLD can be found in non-obese individuals, and it seems to be higher in Asians, even when strict body mass index (BMI) criteria ( $> 25 \text{ kg/m}^2$ ) is used to define obesity. The percentage of non-obese Asians with NAFLD is reported to 12.6% in Korea, 75% in India, and 7.3% in China [4]. In addition, research findings have proven that elevated low-density lipoprotein cholesterol (LDL-c) levels are positively associated with NAFLD incidence. Interestingly, Sun et al. [5]. reported that LDL-c levels within the normal range were positively associated with NAFLD incidence in non-obese individuals.

Nomograms have been widely developed for risk prediction models of various diseases [6]. Although several predictive models have been developed to evaluate the risk of NAFLD based on machine learning techniques [7, 8], a simple nomogram, to our knowledge, has not been established specifically to estimate the risk of NAFLD in a non-obese population with normal LDL-c levels. This study aimed to construct and validate a nomogram predicting NAFLD in a non-obese population, and provide a personalized prediction tool by cost-effective and accessible variables.

## Methods

### Study population

We downloaded the raw data from the Dryad Digital Repository (<http://www.datadryad.org/>), which were shared by Sun et al [9]. This longitudinal study consisted of 16173 NAFLD-free individuals at baseline, though complete data was only obtained for 13240 individuals at the 4-y follow-up. All eligible individuals were divided into two cohorts in a 2:1 ratio, using computer-generated random numbers. Therefore, 8872 individuals were assigned to the training cohort, whereas 4368 individuals were assigned to the validation cohort. A flow chart is depicted in Fig. 1. NAFLD was diagnosed by hepatic ultrasound examination, excluding alcoholic hepatitis, viral hepatitis, or other known causes of liver disease. Individuals were excluded if they had a BMI  $\geq 25 \text{ kg/m}^2$ , LDL-c  $> 3.12 \text{ mmol/L}$ , or with history of hypertension, diabetes, and hyperlipidemia.

The clinical data extracted included age, sex, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), fasting plasma glucose (FPG), total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-c), LDL-c, blood urea nitrogen (BUN), creatinine (Cr), uric acid (UA), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), total bilirubin (TBil), direct bilirubin (DBil), follow-up time and the diagnosis of NAFLD.

All data are available from the Dryad Digital Repository for secondary analysis. Ethical approval was not necessary because no data was directly collected from the participants.

## **Development of the nomogram**

A univariate Cox regression analysis was first performed to evaluate the clinical candidate variables at  $p < 0.10$ . Subsequently, significant variables were then included in a multivariate Cox regression analysis. Finally, a nomogram was constructed to estimate risk of developing NAFLD by incorporating selected predictors from the training cohort using a stepwise forward selection model [10].

## **Validation of the nomogram**

We validated the nomogram internally for the training cohort and externally for the validation cohort. Concordance index (C-index) was applied to evaluate the predictive modeling of the nomogram, and a C-index  $> 0.8$  suggested good discrimination. A receiver operating characteristic (ROC) curve was also applied to assess discriminating ability, and the area under the ROC curve (AUC) was  $> 0.7$ , indicating a good performance of the nomogram. A calibration curve was applied to measure the predictive accuracy of the nomogram. Bootstraps with 100 resamples were used for the calibration assessment.

## **Construction of risk signature**

Risk scores were calculated based on multivariate Cox coefficients and selected values. Individuals were categorized into high- and low-risk groups based on the median risk score. Risk curves were plotted using data from the training and validation cohorts.

## **Assessing the clinical usefulness of the nomogram**

The clinical usefulness of the nomogram was evaluated using a decision curve analysis (DCA) for the whole cohort. DCA is a method for evaluating and comparing predictive models and calculating the net benefits against threshold probabilities [11].

## **Statistical analyses**

Data analyses were performed using SPSS 22.0 software (SPSS Inc, Chicago, IL). Measurement data were first assessed for normal distribution, and then expressed as the means  $\pm$  standard deviation (SD) and medians (interquartile ranges), respectively. This data was compared between the two groups using t tests and Mann-Whitney U tests for normally and non-normally distributed data, respectively.

Enumeration data were expressed as ratios and analyzed by the  $\chi^2$  test. Age groups consisted of: 14–36, 37–45 and 46–93 y, programmed by X-tile software version 3.6.1 [12]. Cox regression analysis, nomogram model, C-index, ROC, calibration curve, DCA curves, and risk curves were conducted using R (<http://www.R-project.org>). All statistical tests were two-sided, and  $p < 0.05$  was considered statistically significant.

## **Results**

## ***Baseline characteristics and risk factors of NAFLD***

There were 1169 (13.2%) and 549 (12.6%) individuals developing NAFLD in the training and validation cohorts, respectively. The median follow-up time was 2.99 y (quartile: 1.96–3.95) for the training cohort and 2.97 y (quartile: 1.95–3.92) for the validation cohort. Additionally, no statistically significant differences in clinical characteristics were observed between the two cohorts. The baseline characteristics for the training and validation sets are summarized in Table 1.

Univariate Cox regression analysis identified 15 candidate clinical variables according to their p-values, including age, sex, blood pressure, BMI, FBG, TC, TG, HDL-c, LDL-c, Cr, BUN, UA, ALT, AST and GGT. Multivariate Cox regression analysis showed that age, sex, BMI, FBG, TC, TG, UA and ALT were positively correlated with the risk of NAFLD, whereas HDL-c was negatively correlated with the risk of NAFLD.

## ***Development and validation of a NAFLD-predicting nomogram***

The nomogram was constructed to predict the risk of NAFLD based on the significant predictors (age, sex, BMI, FBG, TC, TG, HDL-c, UA and ALT) in the training cohort (Fig. 2). Each value for the individuals was determined according to the top Points scale, and then the points for each variable were added. Finally, a personalized risk of NAFLD was obtained according to Total Points scale.

The C-index was 0.804 [95% CI, 0.792–0.852] and 0.802 [95% CI, 0.784–0.820] for the training and validation cohorts, respectively, which demonstrated good predictive modeling by the nomogram. In the training cohort, the AUC for 1-y, 2-y, 3-y and 4-y risk was 0.835, 0.825, 0.816 and 0.782, respectively (Fig. 3a). Likewise, in the validation cohort, the AUC for 1-y, 2-y, 3-y and 4-y risk was 0.817, 0.820, 0.814 and 0.813 (Fig. 3b), indicating an excellent performance of the nomogram. The calibration curves for the probability of NAFLD risk showed excellent accuracy in the predictive modeling of the nomogram for both the training and validation cohorts (Fig. 4a, b). Collectively, these results revealed that the nomogram could accurately predict NAFLD incidence in the non-obese Chinese population.

## ***Construction of risk signature***

Risk scores were calculated and are shown in *Additional file 1*. Individuals were categorized into high- and low-risk groups based on their median risk score. Significant differences were observed between high- and low-risk groups in the training cohort (Fig. 5a, b) and the validation cohort (Fig. 5c, d). Thus, this model could differentiate high- and low-risk individuals.

## ***Clinical usefulness of a NAFLD-predicting nomogram***

The DCA displayed a clinical usefulness for 1-y, 2-y, 3-y, and 4-y incidence of NAFLD using data from both the training and validation cohorts. When the threshold probability ranged from 0.15 to 0.82 at 1-y, 2-y, 3-y and 4-y, the NAFLD-predicting nomogram provided more of a net benefit than the “all individuals with NAFLD” or “no individuals with NAFLD” (Fig. 6a–d), which suggested that the nomogram was clinically useful. For example, in Fig. 6d with regard to the 40% risk probability, the net benefit was about 8%, which

could be interpreted that the model could be an alternative method for evaluating NAFLD by about 8 per 100 individuals without developing NAFLD over the next 4 years.

## Discussion

In this longitudinal study, we constructed an easy-to-use nomogram for predicting incidence NAFLD in the non-obese Chinese population. This nomogram exhibited excellent predictive modeling for the internal training and external validation cohorts. Moreover, the nomogram categorized individuals into high- and low-risk groups, with the high-risk group displayed a significant probability of suffering from NAFLD. The DCA displayed clinical usefulness of this nomogram for predicting NAFLD incidence. To the best of our knowledge, this study is the first to develop a NAFLD-predictive nomogram for a non-obese population using easily obtained clinical parameters.

NAFLD comprises non-alcoholic simple fatty liver, non-alcoholic steatohepatitis (NASH) and cirrhosis [13]. To improve early prevention and expedite intervention, several noninvasive panels have been developed to stratify high risk groups. Previous studies have assessed the diagnostic and predictive capabilities of some models including the fatty liver index (based on BMI, TG, GGT and waist circumference), the FIB-4 index (based on age, AST, ALT and platelet count), the NAFLD fibrosis score (based on age, BMI, AST/ALT ratio, diabetes status, albumin and platelet counts), the BARD score (based on BMI, AST/ALT ratio and diabetes status), the AST/platelet ratio index, the hepatic steatosis index (based on BMI, AST/ALT ratio and diabetes status), and the noninvasive Koeln-Essen-index (based on age, AST, AST/ALT ratio and total bilirubin) [14–20].

Consistent with these models, nine significant predictors (age, sex, BMI, FBG, TC, TG, HDL-c, UA, and ALT) in our study were identified. Unexpectedly, our study found that sex was associated with NAFLD and that males, in particular, were at a higher risk for NAFLD. Research has found NAFLD to be more prevalent in males over premenopausal females, and this is possibly due to growth and sex hormones, as well as genetic factors [21]. Among these metabolic parameters, UA was first included in our NAFLD-predictive model in the non-obese population because Zheng et al. [22] noted that elevated serum UA levels were positively correlated to NAFLD risk in lean Chinese adults, independent of any other metabolic parameters. Thus, these identified parameters in our NAFLD-predictive model were reliable and accurate.

The aforementioned models were primarily established in NAFLD patients with advanced fibrosis from Western countries, and thus exhibited higher positive predictivity in populations with progressive diseases. In addition, study populations were predominately overweight or obese, which limited the use of these models in non-obese populations. Furthermore, these studies were primarily cross-sectional designs and so cannot fully address causality or temporality between predictors and NAFLD incidence.

Consequently, we proposed a nomogram to predict the NAFLD incidence at 1-y, 2-y, 3-y and 4-y in a non-obese Chinese population based on data gathered from a longitudinal study. Our nomogram exhibited satisfactory discrimination and calibration capabilities using C-index, ROC and calibration plots both for the training and validation cohorts. Besides, this nomogram successfully classified individuals into high-

and low-risk groups, and the high-risk group displayed a significant probability for developing NAFLD. More importantly, the DCA revealed that when threshold probability ranged from 0.15 to 0.82 at 1-y, 2-y, 3-y and 4-y, the NAFLD-predicting nomogram provided more of a net benefit than the “all individuals with NAFLD” or “no individuals with NAFLD”.

A major strength in our study was that we developed a nomogram from a large sample in a longitudinal study, which provided excellent applicability to the general population, especially Asians. However, potential limitations of our study should also be considered. First, NAFLD was diagnosed by hepatic ultrasound examination, which is not the gold standard for diagnosing NAFLD because it is unable to evaluate the severity of NAFLD. Further biopsy results are needed to validate the predictive power of this nomogram. Second, some information about risk factors for NAFLD such as lifestyle variables and waist circumference were not collected, and cases with missing data were excluded from analysis, resulting in a possible selection bias. Third, this nomogram was constructed using data from a single center; multicenter studies should be performed to further validate the predictive capability of this nomogram.

## Conclusions

We constructed a nomogram based on nine risk predictors, including age, sex, BMI, FBG, HDL-c, TC, TG, UA and ALT to predict NAFLD incidence in 5 years. This nomogram can be deployed as a simple and affordable tool for stratifying individuals at high risk of NAFLD, and thus serve to expedite treatment of NAFLD.

## Abbreviations

NAFLD: Nonalcoholic Fatty Liver Disease; C-index: Concordance Index; ROC: Receiver Operating Characteristic; DCA: Decision Curve Analysis; BMI: Body Mass Index; FBG: Fasting Blood Glucose; TC: Total Cholesterol; TG: Triglycerides; HDL-c: High-Density Lipoprotein cholesterol; LDL-c: Low-Density Lipoprotein cholesterol; UA: Uric Acid; ALT: Alanine Aminotransferase; T2DM: Type 2 Diabetes Mellitus; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; FPG: Fasting Plasma Glucose; BUN: Blood Urea Nitrogen; Cr: Creatinine; AST: Aspartate Aminotransferase; GGT: Gamma-Glutamyl Transferase; TBil: Total Bilirubin; DBil: Direct Bilirubin.

## Declarations

### Acknowledges

We thank the Sun et al for data collection and sharing.

### Authors' contributions

JW and YT collected and analyzed the data; KP and HL analyzed and interpreted the data; JX conceived the study.

## Funding

This study was supported by grants from the National Natural Science Funds of China [grant no. 81760168].

## Availability of data and materials

The raw data including 16173 NAFLD-free individuals at baseline was downloaded from the Dryad Digital Repository (<http://www.datadryad.org/>), which were shared by Sun et al.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

There are no conflicts of interest to report.

## References

1. Younossi ZM, Marchesini G, Pinto-Cortez H, Petta S. Epidemiology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis: Implications for Liver Transplantation. *Transplantation*. 2019;103:22–7.
2. ZM Y, AB K, Y DA. F, L H, M W: Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64:73–84.
3. LA A, QM A, H T. G T: Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gut*. 2017;66:1138–53.
4. Kim D, Kim WR. Nonobese Fatty Liver Disease. *Clinical gastroenterology hepatology: the official clinical practice journal of the American Gastroenterological Association*. 2017;15:474–85.
5. DQ S, WY L, SJ W, GQ Z, DC MB, KQ Z, MH SDS. Z: Increased levels of low-density lipoprotein cholesterol within the normal range as a risk factor for nonalcoholic fatty liver disease. *Oncotarget*. 2016;7:5728–37.
6. Park SY. Nomogram: An analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg*. 2018;155:1793.
7. CF HM, CH XZS. Y, YM L: Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. *BioMed research international*. 2018;2018:4304376.

8. TC Y, AJ M, VW W, YK T, HL C, PC Y, GL W. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther.* 2017;46:447–56.
9. Sun D-Q, Wu S-J, Liu W-Y, Wang L-R, Chen Y-R, Zhang D-C, Braddock M, Shi K-Q, Song D, Zheng M-H. Association of low-density lipoprotein cholesterol within the normal range and NAFLD in the non-obese Chinese population: a cross-sectional and longitudinal study. *BMJ open.* 2016;6:e013781.
10. FE H, KL L, DB M. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine.* 1996;15:361–87.
11. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making: an international journal of the Society for Medical Decision Making.* 2006;26:565–74.
12. RL C, M D-F DLR. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research: an official journal of the American Association for Cancer Research.* 2004;10:7252–9.
13. E V-G NC. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *Journal of hepatology.* 2018;68:305–15.
14. 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.
15. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, Sulkowski S, Torriani M, Dieterich FJ, Thomas DT. DL, et al: Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology.* 2006;43:1317–25.
16. JM PA, GC HGMEBJG, JP FFESSADB. B, et al: The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology.* 2007;45:846–54.
17. SA H, HL DO, S G, BA N-T A. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut.* 2008;57:1441–7.
18. Lee J-H, Kim D, Kim HJ, Lee C-H, Yang JI, Kim W, Kim YJ, Yoon J-H, Cho S-H, Sung M-W, Lee H-S. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Digestive liver disease: official journal of the Italian Society of Gastroenterology the Italian Association for the Study of the Liver.* 2010;42:503–8.
19. Lin Z-H, Xin Y-N, Dong Q-J, Wang Q, Jiang X-J, Zhan S-H, Sun Y, Xuan S-Y. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology.* 2011;53:726–36.
20. U D, I W MDSLMS, I DN, A KJS, U C. T, HM S: **NIKEI: a new inexpensive and non-invasive scoring system to exclude advanced fibrosis in patients with NAFLD.** *PloS one.* 2013;8:e58360.
21. C S, Ž D. D R: Personalized therapy when tackling nonalcoholic fatty liver disease: a focus on sex, genes, and drugs. *Expert Opin Drug Metab Toxicol.* 2018;14:831–41.

## Tables

Table 1. Baseline characteristics of training and validation cohorts.

Characteristic	Training cohort (n=8872)	Validation cohort (n=4368)	P value
Age (years)	38 (30-48)	37 (30-48)	0.174
Sex, no. (%)			0.678
Male	4721 (53.2%)	2348 (53.8%)	
Female	4151 (46.8%)	2020 (46.2%)	
Hypertension			0.921
Without	7664 (86.4%)	3776 (86.4%)	
With	1208 (13.6%)	592 (13.6%)	
BMI (kg/m <sup>2</sup> )	21.13 (19.64-22.64)	21.11(19.55-22.64)	0.547
FPG (mmol/l)	5.04 (4.77-5.36)	5.04 (4.77-5.35)	0.244
TC (mmol/l)	4.30 (3.86-4.74)	4.29 (3.84-4.76)	0.992
TG (mmol/l)	0.93 (0.69-1.32)	0.93 (0.69-1.32)	0.365
HDL-c (mmol/l)	2.31 (1.99-2.63)	2.32 (1.97-2.65)	0.753
LDL-c (mmol/l)	1.41 (1.20-1.64)	1.40 (1.20-1.63)	0.478
BUN (mmol/l)	77.70 (66.20-91.50)	77.00 (66.00-92.00)	0.595
Cr (mmol/l)	4.17 (3.46-4.96)	4.22 (3.50-5.05)	0.144
UA (mmol/l)	266.00 (214.00-329.00)	265.00 (215.00-330.00)	0.649
ALT (U/L)	14.00 (11.00-20.00)	14.00 (11.00-19.00)	0.550
AST (U/L)	20.00 (17.00-23.00)	19.00 (17.00-23.00)	0.896
GGT (U/L)	16.00 (12.00-23.00)	16.00 (12.00-23.00)	0.990
TBil (μmol/l)	12.10 (9.60-15.30)	12.10 (9.60-15.30)	0.659
DBil (μmol/l)	1.60(1.30-2.00)	1.60 (1.30-2.00)	0.386
Follow-up years	2.99(1.96-3.95)	2.97 (1.95-3.92)	0.272
Identified NAFLD, no. (%)	1169 (13.2%)	549 (12.6%)	0.178

Data are shown as numbers (%) or medians (interquartile ranges).

NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FPG, fasting plasma glucose; TC, total cholesterol; TG, triglyceride; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; BUN, blood urea nitrogen; Cr, creatinine; UA, uric acid; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; TBil, total bilirubin; DBil, direct bilirubin.

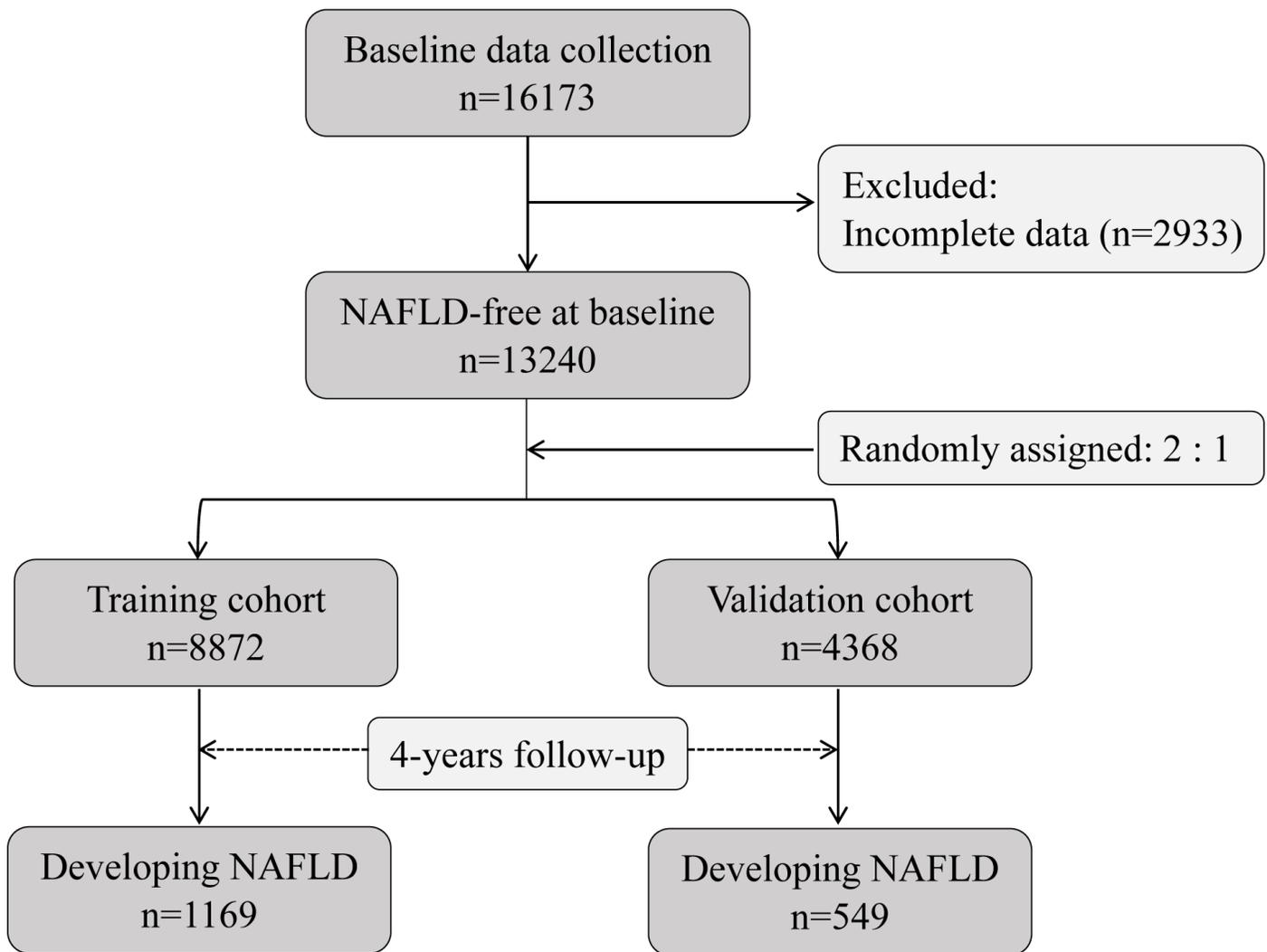
Table 2. Risk factors for NAFLD according to the Cox regression analyses.

Variable	Univariate analysis		Multivariate analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
Age (years)	1.596 [1.493, 1.706]	<0.001	1.276 [1.182, 1.378]	<0.001
Sex (Male vs. Female)	3.610 [3.166, 4.115]	<0.001	1.718 [1.416, 2.084]	<0.001
Hypertension	2.164 [1.897, 2.470]	<0.001	1.090 [0.940, 1.264]	0.193
Diabetes (Without vs. With)				
BMI (kg/m <sup>2</sup> )	1.616 [1.560, 1.674]	<0.001	1.417 [1.364, 1.473]	<0.001
SBP (mmol/l)	1.362 [1.305, 1.421]	<0.001	1.115 [1.047, 1.188]	0.001
DBP (mmol/l)	1.387 [1.270, 1.514]	<0.001	1.204 [1.085, 1.337]	0.001
FBG (mmol/l)	1.435 [1.401, 1.471]	<0.001	1.193 [1.143, 1.245]	<0.001
HDL-c (mmol/l)	1.909 [1.669, 2.184]	<0.001	0.909 [0.691, 1.195]	0.714
LDL-c (mmol/l)	0.183 [0.149, 0.224]	<0.001	0.533 [0.420, 0.677]	<0.001
Cr (mmol/l)	1.003 [1.003, 1.004]	<0.001	0.990 [0.985, 1.001]	0.051
BUN (mmol/l)	1.097 [1.060, 1.135]	<0.001	0.948 [0.898, 1.000]	0.053
UA (mmol/l)	1.007 [1.006, 1.008]	<0.001	1.003 [1.002, 1.004]	<0.001
ALT (U/L)	1.008 [1.007, 1.009]	<0.001	1.017 [1.011, 1.023]	<0.001
AST (U/L)	1.007 [1.005, 1.009]	<0.001	0.978 [0.966, 1.001]	0.058
GGT (U/L)	1.009 [1.008, 1.010]	<0.001	1.002 [1.000, 1.004]	0.069
TBil (μmol/l)	0.999 [0.988, 1.011]	0.910		
DBil (μmol/l)	0.940 [0.853, 1.037]	0.216		

Data are shown as odds ratio (95% CI), P value.

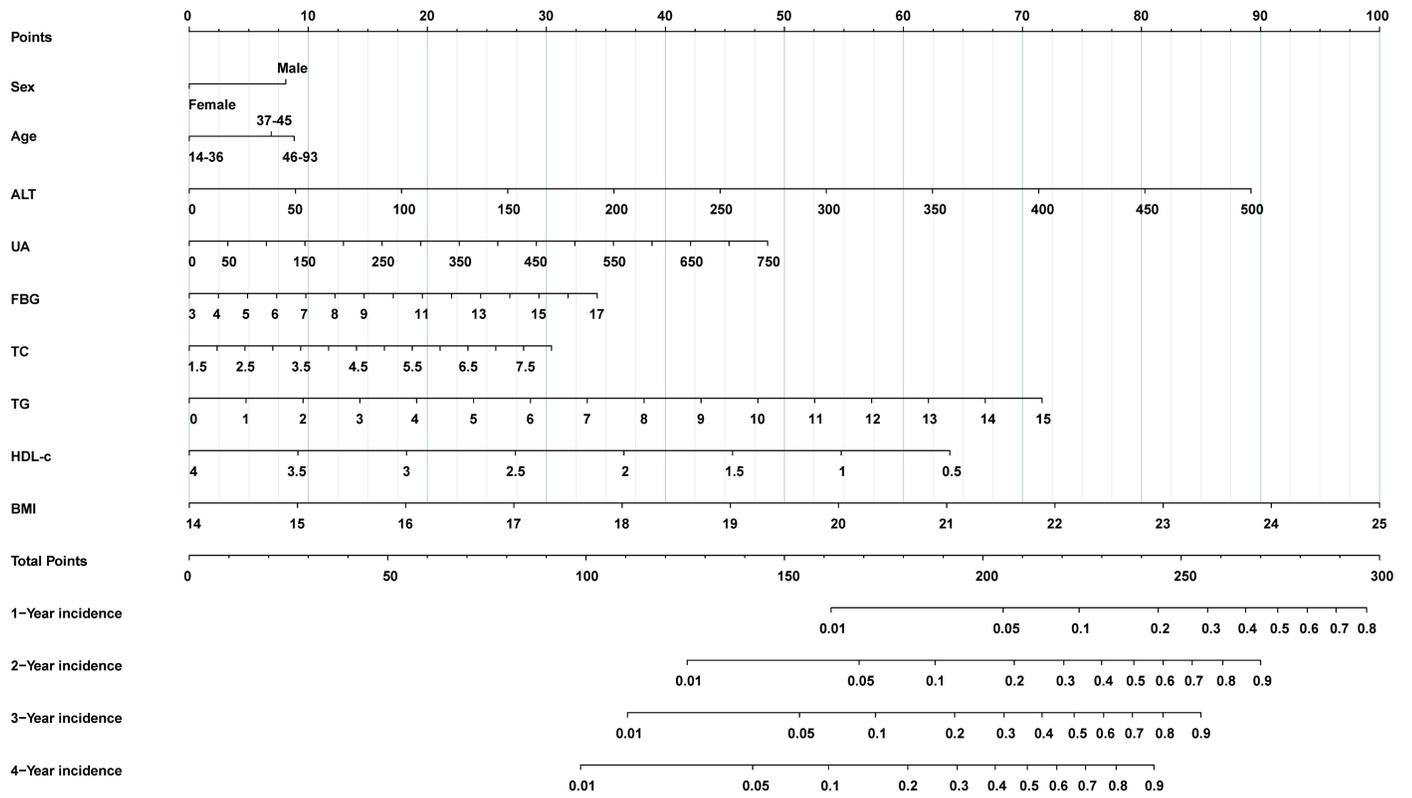
NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FPG, fasting plasma glucose; TC, total cholesterol; TG, triglyceride; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; BUN, blood urea nitrogen; Cr, creatinine; UA, uric acid; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; TBil, total bilirubin; DBil, direct bilirubin.

## Figures



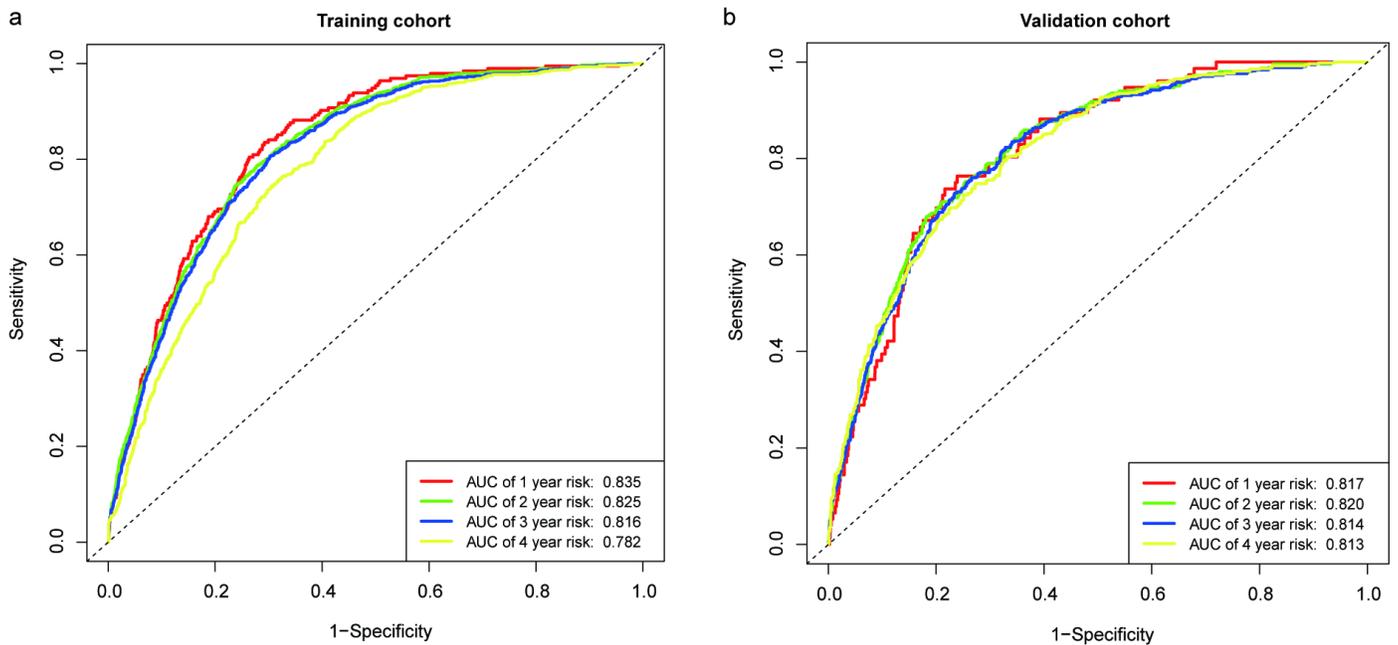
**Figure 2**

The flow diagram of individuals in the study. NAFLD: nonalcoholic fatty liver disease.



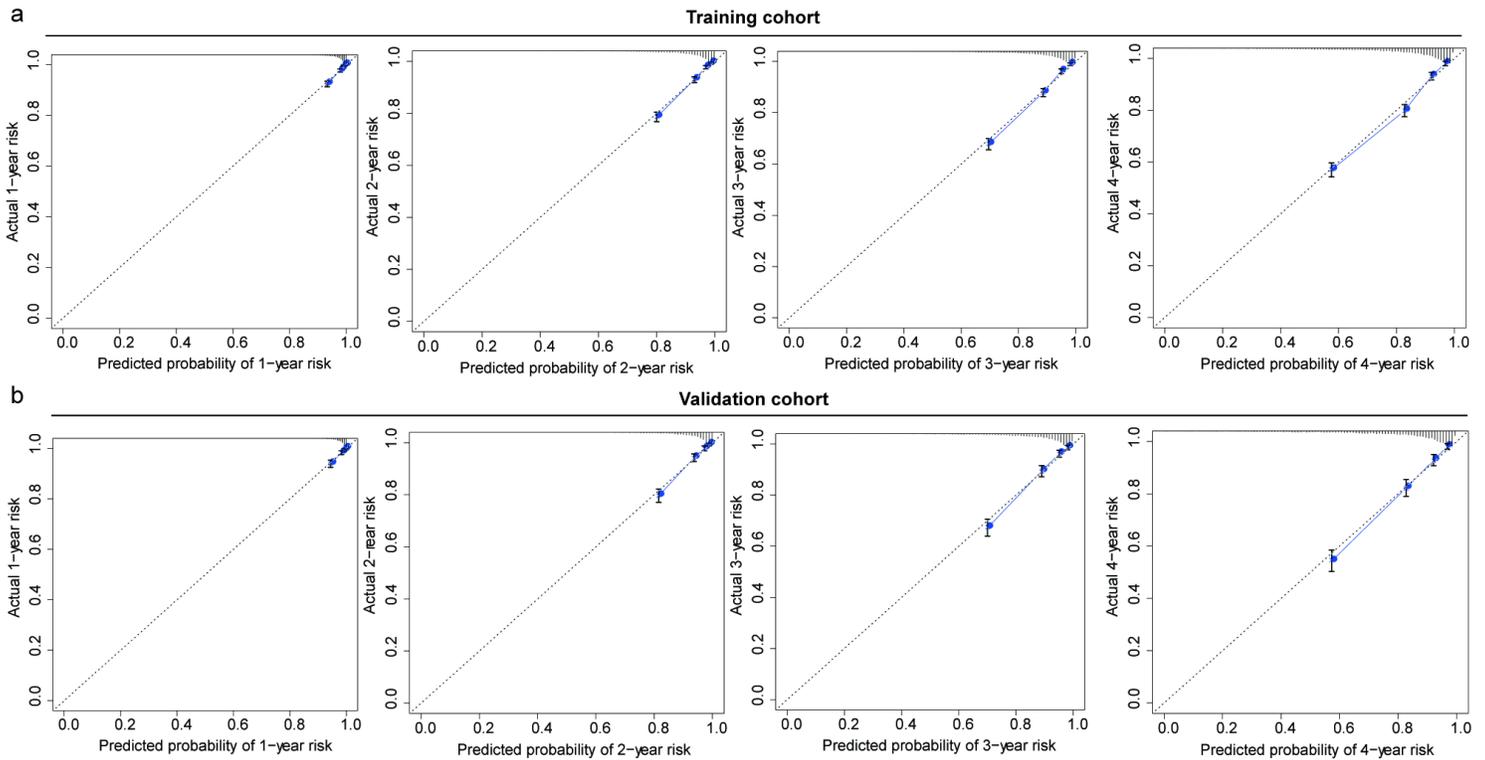
**Figure 4**

Nomogram for predicting NAFLD incidence in the non-obese Chinese population. NAFLD: nonalcoholic fatty liver disease.



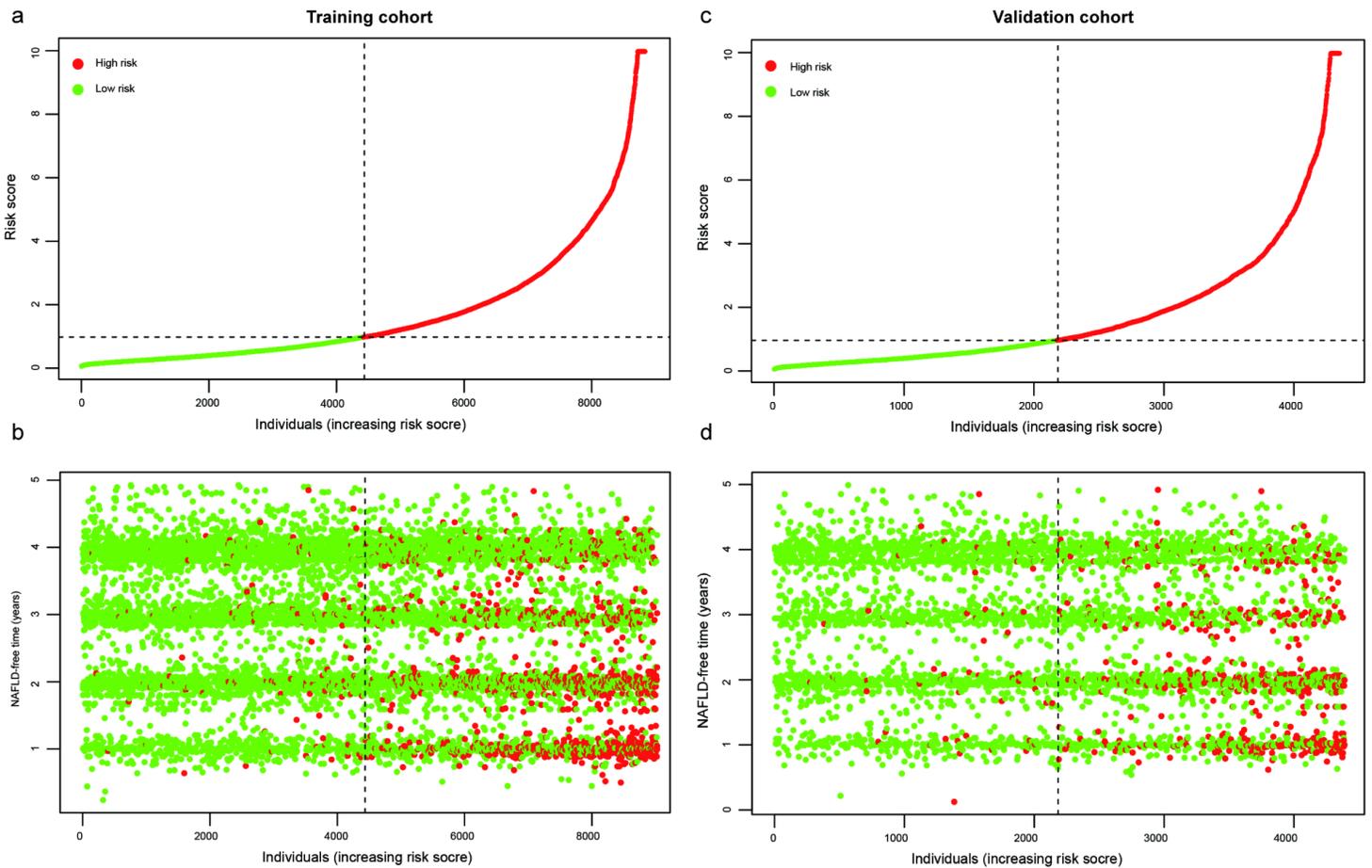
**Figure 6**

ROC curves of the nomogram in the training cohort (a) and the validation cohort (b). (a) The AUC for 1-y, 2-y, 3-y and 4-y risk in the training cohort. (b) The AUC for 1-y, 2-y, 3-y and 4-y risk in the validation cohort. ROC: receiver operating characteristics curves, AUC: area under the ROC curve.



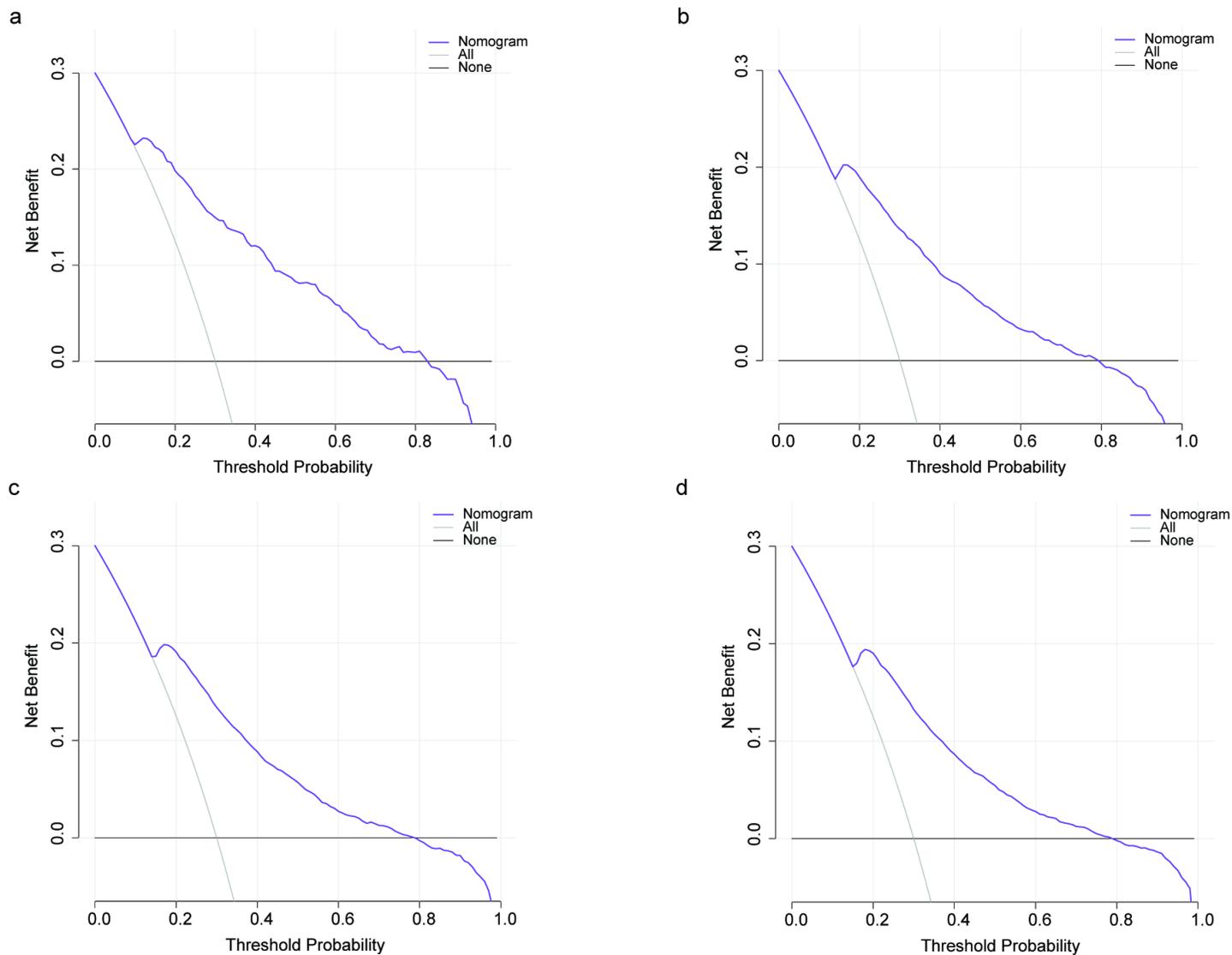
**Figure 7**

Calibration curves of the nomogram in the training cohort (a) and the validation cohort (b). The gray dashed lines indicate an ideal model, and the blue solid lines indicate the predictive performance of the nomogram.



**Figure 10**

Individuals were stratified into high- or low-risk groups based on the risk score. (a and b) Risk score distribution and NAFLD status in the training cohort. (c and d) Risk score distribution and NAFLD status in the validation cohort. Green dot represents individuals without NAFLD, and red dot represents individuals with NAFLD. NAFLD: nonalcoholic fatty liver disease.



**Figure 12**

The decision curve analysis of the nomogram for 1-y (a), 2-y (b), 3-y (c), and 4-y (d) NAFLD risk throughout the whole cohort. The black line indicates the net benefit when no individuals develop NAFLD, while the gray line indicates the net benefit when all individuals suffer from NAFLD. The area among the black line, gray line, and purple line indicates the clinical usefulness of the nomogram. NAFLD: nonalcoholic fatty liver disease.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [additionalfile1.docx](#)
- [additionalfile1.docx](#)