ORIGINAL RESEARCH

Establishment of a Risk Prediction Model for Non-alcoholic Fatty Liver Disease in Type 2 Diabetes

Yali Zhang · Rong Shi · Liang Yu · Liping Ji · Min Li ·
Fan Hu

ABSTRACT

Introduction: Non-alcoholic fatty liver disease (NAFLD) is becoming more prevalent in patients with type 2 diabetes mellitus (T2DM) and can contribute to serious liver damage in this patient population. The aim of this study was to develop a risk nomogram for NAFLD in a Chinese population with T2DM.

Methods: A questionnaire survey, physical examination and biochemical indicator testing were performed on 874 patients with T2DM, and the collected data were used to evaluate the risk to develop NAFLD in T2DM patients. The least absolute shrinkage and selection operator (LASSO) regression analysis method was used to optimize variable selection by running cyclic coordinate descent with *k*-fold (tenfold in this case) cross-validation. Multivariable logistic regression analysis was applied to build a predictive model by introducing the predictors selected from the LASSO regression analysis. The nomogram was developed based on the selected variables visually. A calibration plot, receiver operating characteristic curve (ROC) and decision curve analysis (DCA) were used to validate the model, with further assessment by external validation.

Results: A total of nine predictors, namely sex, age, total cholesterol (TC), body mass index (BMI), waistline, diastolic blood pressure (DBP), serum uric acid (SUA), course of disease and high-density lipoprotein-cholesterol (HDL-C), were identified by LASSO regression analysis from a total of 24 variables studied. The model constructed using these nine predictors displayed medium prediction ability, with an area under the ROC of 0.848 in the training set and 0.809 in the validation set. The DCA curve showed that the nomogram could be applied clinically if the risk threshold was between 48 and 91%, which was found to be between 44 and 82% in the external validation.

Conclusion: Introducing sex, age, TC, BMI, waistline, DBP, SUA, course of disease and HDL-C into the risk nomogram increased its usefulness for predicting NAFLD risk in patients with T2DM.

Y. Zhang · R. Shi · L. Yu · L. Ji · M. Li · F. Hu (✉)
School of Public Health, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
e-mail: joyking2003@163.com

Key Summary Points

Why carry out the study?

The prevalence of diabetes in the Chinese adult population is estimated to be as high as 9.7%.

About one- to two-thirds of people with type 2 diabetes mellitus (T2DM) may develop non-alcoholic fatty liver disease (NAFLD), placing a serious economic burden on healthcare systems and posing health risks to individuals and society.

In this study, factors potentially influencing the development of NAFLD in T2DM were assessed by developing a nomogram risk predictive model, with the aim to identify important risk factors for the development of NAFLD in patients with T2DM.

What was learned from the study?

The results showed that nine indicators, namely sex, age, total cholesterol, body mass index, waistline, diastolic blood pressure, serum uric acid, course of disease and high-density lipoprotein-cholesterol, are effective risk predictors of NAFLD in T2DM.

The risk nomogram is useful for prediction of NAFLD risk in persons with T2DM, and the nomogram method can also be extrapolated to other complications and study-related risk factors.

The prediction model was established using LASSO regression, logistic regression and a newly developed risk nomogram. This nomogram of NAFLD was validated, and the results were also verified by external verification methods (including ROC, C-INDEX, DCA), with good results.

INTRODUCTION

Type 2 diabetes (T2DM) is a complex metabolic disorder that has become a major public health problem worldwide in the past three decades, with diabetes mellitus the ninth major cause of death. About one in 11 adults worldwide now have diabetes mellitus, of whom 90% have T2DM [1]. T2DM can cause injury not only to the pancreas and insulin system but also the liver, cardiac, kidneys, gut, muscle, fat cells and even the brain [2–6]. Diabetes can affect many different organ systems in the body and, over time, can lead to serious complications, such as non-alcoholic fatty liver disease (NAFLD).

NAFLD is gradually becoming more prevalent, with a recent estimate of global prevalence of 25% [7]. There is a very high prevalence of NAFLD in individuals with T2DM, with the risk of NAFLD in this patient population affected by age, gender, body mass index (BMI), and location (countries). Studies have shown that NAFLD may be present in up to 70% of patients with diabetes [8]. A recent meta-analysis found that the global prevalence of NAFLD among patients with T2DM is about 55%, with studies from the European region reporting the highest prevalence (68%), increasing to 98% in patients with morbid obesity [7, 9]. Patients with NAFLD are usually insulin resistant. A large number of patients with T2DM develop NAFLD due to an inflammatory complication known as non-alcoholic steatohepatitis (NASH). It is increasingly recognized that NASH in patients with T2DM may lead to further complications, such as cirrhosis and hepatocellular carcinoma [10,11]. In the Chinese adult population, diabetes among those without known diabetes is associated with higher risks of liver cancer and major chronic liver diseases.

The pathological damage caused by diabetes itself is exacerbated by NAFLD, which in turn negatively affects the occurrence and development of diabetes[12] while possibly also introducing various complications, such as metabolic syndrome, obesity and cardiovascular and cerebrovascular diseases. It is also important to note the importance of the bidirectional association between NAFLD and T2DM. In this

context, T2DM and NAFLD can develop almost simultaneously in a patient, which confounds the prevalence of NAFLD in patients with T2DM or the prevalence of T2DM in patients with NAFLD. There is a strong association between NAFLD and diabetes risk, with an individual's risk of developing diabetes increased by approximately fivefold in the presence of NAFLD, although this development is dependent on the population studied, course of follow-up and methodology used to diagnose NAFLD [8]. Nevertheless, this association and its bidirectional causal relationship require additional investigation [13].

Overall, as patients with T2DM have a higher prevalence of NAFLD, there should be systematic screening for NAFLD to increase early detection and prevent complications. It is also extremely important to control complications in persons with diabetes. The World health organization (WHO) reported that the number of people with diabetes is expected to increase to 592 million by 2035 [1], with people in developing countries experiencing a larger decrease in in health than those in developed countries by 2035.

Based on screening data on a T2DM patient cohort and the results of clinical examinations, our team is evaluating and modeling the risk factors of T2DM with NAFLD one by one, with the aim to establish a scientific risk prediction model suitable for family doctors to conduct early screening of NAFLD patients with diabetes.

METHODS

Subjects

The study is a retrospective cohort study. Based on screening of the electronic medical records system of the Sanlin and Jinyang community health service centers in Shanghai, we identified 24 relevant indicators of T2DM. In 2016–2017, data on 1013 people with T2DM were collected, and these data were screened and followed up in these two communities from September 2018 to August 2019. Of the 1013 patients with T2DM originally identified, 139 were not included in the final analysis due to loss to follow-up, missing or incorrect data and/or missing samples (loss rate 13.7%). The final patient population for assessment thus included 874 people with T2DM. A flow diagram of the study design is shown in Fig. 1.

The patients with T2DM eligible for entry into the study had been admitted to the Shanghai University of Traditional Chinese Medicine, School of Public Health between September 2018 and August 2019. Inclusion criteria were:

1) Confirmed diagnosis of T2DM based on WHO and American Diabetes Association (ADA) criteria as follows: fasting plasma glucose concentration of $\geq$ 7.0 mmol/L (126 mg/dL), or 2-h post-glucose load venous plasma glucose of $\geq$ 11.1 mmol/L (200 mg/dL), confirmed on two occasions. The exclusion of type 1 diabetes mellitus was based on clinical distinction by the respective attending physicians, including young age of onset, positive islet autoantibody and history of insulin-dependent glycemic control.
2) Adult, with a stable residence in Shanghai for no less than 6 months, or permanent residency.
3) Ongoing or recent alcohol consumption of < 21 standard drinks on average per week for men and < 14 standard drinks on average per week for women. These were considered to be reasonable thresholds for significant alcohol consumption when evaluating patients with suspected NAFLD [9].

Exclusion criteria were: (1) presence of cognitive impairment and/or inability to participate in the cohort independently; (2) presence of serious organic disease, such as a tumor, major surgery, etc.

All analyses were approved by the School of Public Health, Shanghai University of Traditional Chinese Medicine and Sanlin/Jinyang Community Health Centers. All procedures were carried out following international guidelines and regulations by trained researchers. All researchers are doctors, nurses or researchers with professional medical training.
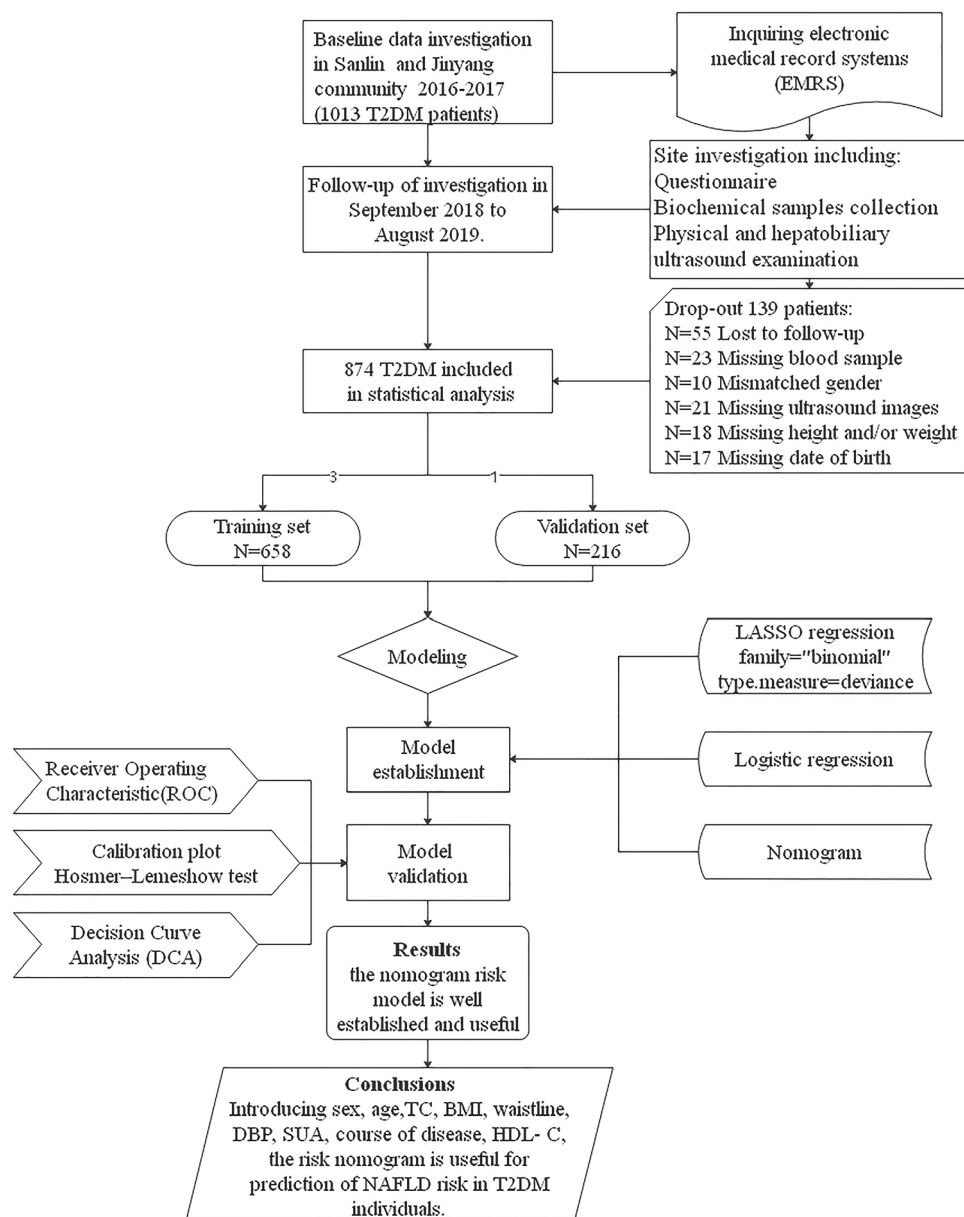
Fig. 1 Flow diagram of study design. *BMI* Body mass index, *DBP* diastolic blood pressure, *HDL-C* high-density lipoprotein-cholesterol, *NAFLD* non-alcoholic fatty liver disease, *SUA* serum uric acid, *TC* total cholesterol, *T2DM* type 2 diabetes mellitus

Measurements

The testing process consisted of:

1) Confirming the time and place with the participating community health center 7 days before start of the study, so that the doctors and nurses in the community will be able to notify patients with T2DM and recruit them to the study.

2) Participants were asked to fast for > 10 h before testing, with screening starting at 7 a.m.

3) Venous blood was drawn first, following with the patient had breakfast; 2 h after the meal, venous blood was drawn again.

4) During the period between taking the samples of fasting blood and postprandial blood, patients completed the fundus examination, physical examination and questionnaire survey according to the medical staff's instructions. The contents related to clinical indicators are completed by specialists.

5) Finally, patients underwent an abdominal ultrasound examination performed by trained professionals to screen for a fatty liver.

Patients' demographic characteristics, physical data, serum biomarkers and transabdominal ultrasound data were collected, and all data were then submitted electronically to the Diabetes Data Management System of our group. All patients completed a questionnaire survey to provide basic demographic variables including age, sex, course of T2DM, drugs, smoking status, alcohol drinking status, among others. The questionnaire was given as an electronic questionnaire on a tablet, with the staff entering the questionnaire into the questionnaire system in advance and helping each patient at one-on-one to obtain the questionnaire data.

A physical examination was also conducted to collect data on body weight (kg), height (m), waistline (cm), systolic blood pressure (SBP) and diastolic blood pressure (DBP) (SBP and DBP measurements [in mmHg] were obtained using an automatic electronic blood pressure monitor [model HBP-9021; Omron Corp., Kyoto, Japan]). BMI was calculated as weight (kg)/ (height [m])^2. An abdominal ultrasound examination was carried by a specialist on the iU22 xMATRIX Ultrasound system (Philips N.V., Amsterdam, the Netherlands). Four technicians were involved in the ultrasound examinations, all of whom were trained to diagnose NAFLD using the same criteria used prior to the study to ensure quality control of the diagnostic images and results. One physician and one nurse also participated in the scan.

Blood samples were collected to obtain essential biochemical variables, including fasting blood glucose (FBP), postprandial blood glucose (PBG), hemoglobin A1c (HbA1c), total cholesterol (TC), triglyceride (TG), low-density lipoprotein-cholesterol (LDL-C), high-density lipoprotein-cholesterol (HDL-C), blood urea nitrogen (BUN), serum creatinine (SCR), serum uric acid (SUA), urinary creatinine (UCR), urine microalbumin (UMA), urine albumin and creatinine ratio (ACR). Blood samples were collected through standardized processes and stored under standard conditions (refrigeration at 4 °C) until sent to the laboratory of Ruijin Hospital in Luwan District for measurement (within 2 h of collection). An automatic biochemical analyzer was used for biochemical testing (AU5800 clinical chemistry analyzer; Beckman Coulter Inc., Brea, CA, USA).

Compliance with Ethics Guidelines

The previous study and data collection were approved by the Ethics Committee of Shanghai Oriental Hospital affiliated to Tongji University (Batch number: [2017] Research Review No. 20). All patients gave informed consent and signed paper informed consent. The study was conducted in accordance with the 1964 Declaration of Helsinki and its subsequent amendments.

Statistical Analyses

Statistical analysis was performed using R software (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria), First, the R *caret* package was used to randomly divide the 874 patients with T2DM into a training set with 658 participants and a validation set with 216 participants for external validation, conforming to the theoretical ratio of 3:1 [14]. The *glmnet* package was used to run LASSO (least absolute shrinkage and selection operator) regression analysis, which is a shrinkage and variable selection method for linear regression models. To obtain the subset of predictors, the LASSO regression analysis minimizes prediction error for a quantitative response variable by imposing a constraint on the model parameters that cause the regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model while

variables with non-zero regression coefficients variables are most strongly associated with the response variable. We set family = "binomial"; this parameter is suitable for the binary discrete dependent variable (binary) because the included dependent variable is whether NAFLD is present or not. Then we set type.measure = deviance to use deviance, that is − 2log-likelihood. Based on the type measure of − 2log-likelihood and the binomial family, the LASSO regression analysis running in R software runs a k-fold (tenfold in this case) cross-validation for centralization and normalization of the included variables and then picks the best lambda value. Lambda.lse gives a model with good performance but the least number of independent variables. So the LASSO method was used to analyze the data in the training set to select the optimal predictors of the present risk factors, including sex, age, course of disease, hypertension, diabetic retinopathy (DR), diabetic nephropathy (DN), BMI, waistline, SBP, DBP, FBG, PBG, HbA1c, TC, TG, LDL-C, HDL-C, BUN, SCR, UA, UCR, UMA and ACR. The risk factor variables were initially screened using these above-mentioned inclusion variables.

Then we used the R language rms package to do logistic regression; a multivariable logistic regression analysis was used to construct a predictive model by introducing the features selected in the LASSO regression model [15]. The features were considered to be odds ratios and a P value with 95% confidence interval. The statistical significance levels were all two-sided. Introducing all the selected features and analyzing statistical significance levels of the features, we applied the statistically significant predictors to develop a predictive model for NAFLD risk. In our study, all of the selected features had statistical significance and were applied to develop the nomogram prediction models. The rms package used to develop the nomogram diagram also uses the R language.

Further, several kinds of validation methods were used to estimate the accuracy of the risk prediction model by using the data of the training set and the validation set respectively. We used the R language pROC package for receiver characteristic curve (ROC) operation. The area under the ROC was used to provide good discrimination for the quality of the risk nomogram to separate true positives from false positives [16]. We used the rms package to draw and calculate the calibration curves, which were used to evaluate the calibration of the NAFLD risk nomogram, accompanied by the Hosmer–Lemeshow test (HLtest.R). We used the nricens package for the decision curve analysis (DCA), which is used to determine the clinical practicability of nomograms based on the net benefit under different threshold probability in the T2DM cohort [17].

RESULTS

Characteristics of the Study Cohort

Of the 874 participants with T2DM in our analysis, of whom 394 were men and 480 were women, 415 had NAFLD and 459 did not have NAFLD. The prevalence of NAFLD in this cohort was thus 47.5%. These patients were randomly placed into the training set or validation set at a ratio of 3:1 (3 in the training set and 1 in the validation set), such that 658 and 216 patients were included in the training set and validation set. All patients completed the related examinations. The basic characteristics of the patients in the two sets are given in Table 1.

Among the patients with NAFLD and without NAFLD, 38.1 and 51.4% were men, respectively. In patients with NAFLD, 27.5% had hypertension, 18.3% had DR and 41.2% had DN; in patients without NAFLD, 52.5% had hypertension, 16.1% had DR and 30.9% had DN.

Independent Risk Factors in the Training Set

Multivariate logistic regression analysis demonstrated that among our patients with T2DM, sex, age, course of disease, BMI, waistline, DBP, TG, HDL-C and SUA were independent risk factors for NAFLD (Fig. 2).

Table 1 Characteristics of the 874 patients with type 2 diabetes mellitus enrolled in the study according to presence/absence of non-alcoholic fatty liver disease and randomization to training set and validation set

Items	Total patient cohort ($n = 874$)	Patients with T2DM with NAFLD ($n = 415$)	Patients with T2DM without NAFLD ($n = 459$)	Training set ($n = 658$)	Validation set ($n = 216$)	P value
Age (years)	64.32 ± 5.92	63.65 ± 5.80	64.92 ± 5.98	64.17 ± 5.94	64.76 ± 5.86	0.001
Sex, n (%)						
Male	394 (45.1%)	158 (38.1%)	236 (51.4%)	300 (45.6%)	94 (43.5%)	0.326
Female	480 (54.9%)	257 (61.9%)	223 (48.6%)	358 (54.4%)	122 (56.5%)	0.000
Diabetic retinopathy, n (%)	150 (17.2%)	76 (18.3%)	74 (16.1%)	118 (17.9%)	32 (14.8%)	0.391
Diabetic nephropathy, n (%)	313 (35.8%)	171 (41.2%)	142 (30.9%)	238 (36.2%)	75 (34.7%)	0.002
Hypertension, n (%)	355 (40.6%)	114 (27.5%)	241 (52.5%)	388 (59.0%)	131 (60.6%)	0.000
Course of disease (years)	9.28 ± 6.31	8.36 ± 6.05	10.11 ± 6.42	9.02 ± 6.16	10.08 ± 6.68	0.000
BMI (kg/m^2)	25.26 ± 3.47	26.87 ± 3.30	23.81 ± 2.94	25.28 ± 3.48	25.21 ± 3.44	0.000
Waistline (cm)	87.24 ± 9.09	90.86 ± 8.58	83.96 ± 8.27	87.34 ± 9.15	86.94 ± 8.92	0.000
SBP (mmHg)	144.95 ± 20.05	147.56 ± 19.98	142.60 ± 19.85	144.4 ± 19.47	146.66 ± 21.69	0.000
DBP (mmHg)	80.26 ± 10.46	82.44 ± 10.42	78.29 ± 10.11	79.9 ± 10.32	81.37 ± 10.82	0.000
FBG (mmol/L)	7.77 ± 2.52	7.80 ± 2.35	7.75 ± 2.66	7.72 ± 2.42	7.94 ± 2.78	0.771
PBG (mmol/L)	12.32 ± 4.74	12.09 ± 4.49	12.53 ± 4.94	12.21 ± 4.55	12.68 ± 5.27	0.166
HbA1c (%)	7.15 ± 1.39	7.20 ± 1.33	7.10 ± 1.44	7.14 ± 1.37	7.17 ± 1.46	0.311
TC (mmol/L)	4.98 ± 1.11	5.07 ± 1.19	4.90 ± 1.02	4.98 ± 1.13	5.00 ± 1.05	0.030
TG (mmol/L)	1.92 ± 1.10	2.32 ± 1.26	1.55 ± 0.76	1.90 ± 1.11	1.97 ± 1.07	0.000
LDL-C (mmol/L)	1.62 ± 0.47	1.66 ± 0.50	1.59 ± 0.43	1.62 ± 0.48	1.64 ± 0.44	0.014
HDL-C (mmol/L)	1.60 ± 0.39	1.50 ± 0.33	1.69 ± 0.42	1.59 ± 0.39	1.62 ± 0.39	0.000
BUN (mmol/L)	5.55 ± 1.52	5.50 ± 1.51	5.60 ± 1.53	5.47 ± 1.50	5.80 ± 1.557	0.371
SCR (µmoI/L)	65.25 ± 18.36	64.53 ± 19.13	65.91 ± 17.62	65.51 ± 18.98	64.47 ± 16.337	0.269
SUA (µmol/L)	309.66 ± 79.31	328.51 ± 82.31	292.62 ± 72.47	307.88 ± 79.96	315.07 ± 77.22	0.000
UCR (µmoI/L)	9.02 ± 4.13	8.99 ± 4.10	9.04 ± 4.16	9.13 ± 4.15	8.65 ± 4.05	0.880

Table 1 continued

Items	Total patient cohort (n = 874)	Patients with T2DM with NAFLD (n = 415)	Patients with T2DM without NAFLD (n = 459)	Training set (n = 658)	Validation set (n = 216)	P value
UMA (mg/mmol)	49.1 ± 72.92	57.29 ± 78.57	41.69 ± 66.62	48.39 ± 71.61	51.25 ± 76.89	0.002
ACR	55.81 ± 102.87	68.78 ± 122.88	44.08 ± 78.96	55.52 ± 103.25	56.69 ± 101.96	0.001

Values in table are presented as the mean with the standard deviation in parenthesis, unless indicated otherwise

BMI Body mass index, *SBP/DBP* systolic/diastolic blood pressure, *FBG* fasting blood glucose, *PBG* postprandial blood glucose, *HbA1c* hemoglobin A1c, *TC* total cholesterol, *TG* triglyceride, *LDL-C* low-density lipoprotein-cholesterol, *HDL-C* high-density lipoprotein-cholesterol, *BUN* blood urea nitrogen, *SCR* serum creatinine, *SUA* serum uric acid, *UCR* urinary creatinine, *UMA* urine microalbumin, *ACR* urine albumin:creatinine ratio
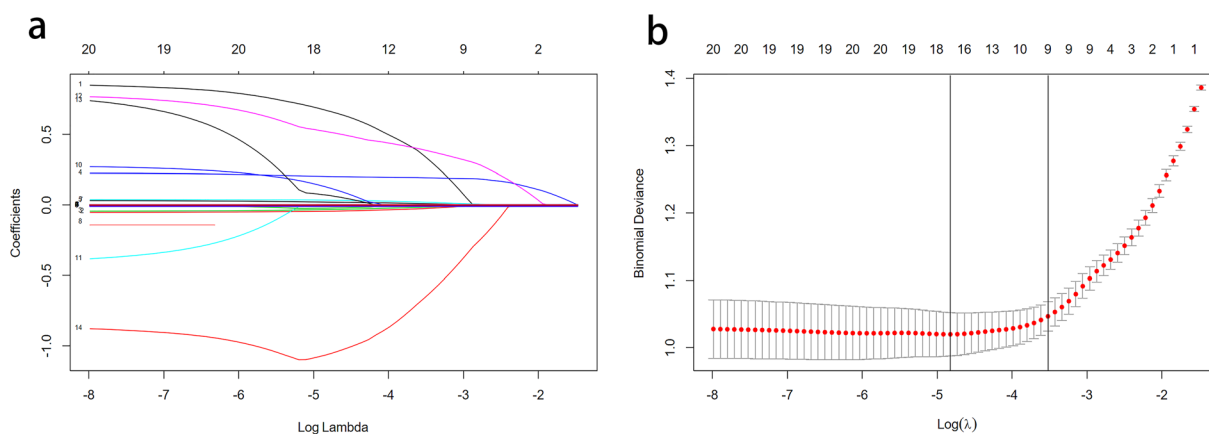
Fig. 2 Variable selection by the LASSO binary logistic regression model. A coefficient profile plot was constructed against the log(lambda) sequence. **a** Ten variables with nonzero coefficients were selected by deriving the optimal lambda. **b** Following verification of the optimal parameter (lambda) in the LASSO model, we plotted the partial likelihood deviance (binomial deviance) curve versus log(lambda) and drew dotted vertical lines based on 1 standard error criteria. *LASSO* Least absolute shrinkage and selection operator

Predictive Model Construction

LASSO regression analysis was used to select predictive variables from those shown in Table 1, and the multivariate logistic regression was used to establish the predictive model. Nine of the original 24 variables were included in the predictive model, namely sex, TG, BMI, waistline, DBP, SUA, course of disease, age and HDL-C, as predictors (Fig. 2). These nine variables had nonzero coefficients in the LASSO regression model. The predictive model was presented as a nomogram, which was used to quantitatively predict the risk probability of NAFLD in patients with T2DM (Fig. 2).

The results of the logistic regression analysis of these nine variables are given in Table 2. As all nine predictors showed significant statistical differences, i.e. they were independent of each other, they were introduced into the predictive model to develop a NAFLD risk nomogram (Fig. 3a). For example, using the nomogram model, a male patient with T2DM of age 62 years, a duration of diabetes for 15 years, a

Table 2 Logistic regression analysis of the predictors for the risk of non-alcoholic fatty liver disease in patients with T2DM identified

Intercept and variables	Estimate	z value	Prediction model			
			P value	Odds ratio	Confidence interval (2.5%)	Confidence interval (97.5%)
Intercept	− 7.350	− 4.147	< 0.001	0.000	1.865	0.020
Sex	0.902	4.279	< 0.001	2.464	1.638	3.746
Age	0.058	− 3.365	< 0.001	0.944	9.123	0.976
Course of disease	− 0.038	− 2.281	0.022	0.963	9.314	0.994
BMI	0.216	4.610	< 0.001	1.241	1.134	1.363
Waistline	0.041	2.398	0.016	1.042	1.008	1.077
DBP	0.023	2.468	0.014	1.024	1.005	1.044
TG	0.565	4.599	< 0.001	1.760	1.397	2.260
HDL-C	− 1.283	− 4.322	< 0.001	0.277	1.530	0.491
SUA	0.003	2.169	0.030	1.003	1.000	1.006

BMI of 28.16 kg/m^2, waistline measurement of 84 cm, a DBP of 94 mmHg, TG of 1.14 mmol/L, HDL-C of 1.07 mmol/L and SUA of 269 μmol/L has an estimated probability of NAFLD of 28.4% (Fig. 3b).

Predictive Model Validation

The ROC curve was used to evaluate the discriminatory capacity of the predictive model. For the predictive model, the pooled area under the ROC of the nomogram is 0.848 in the
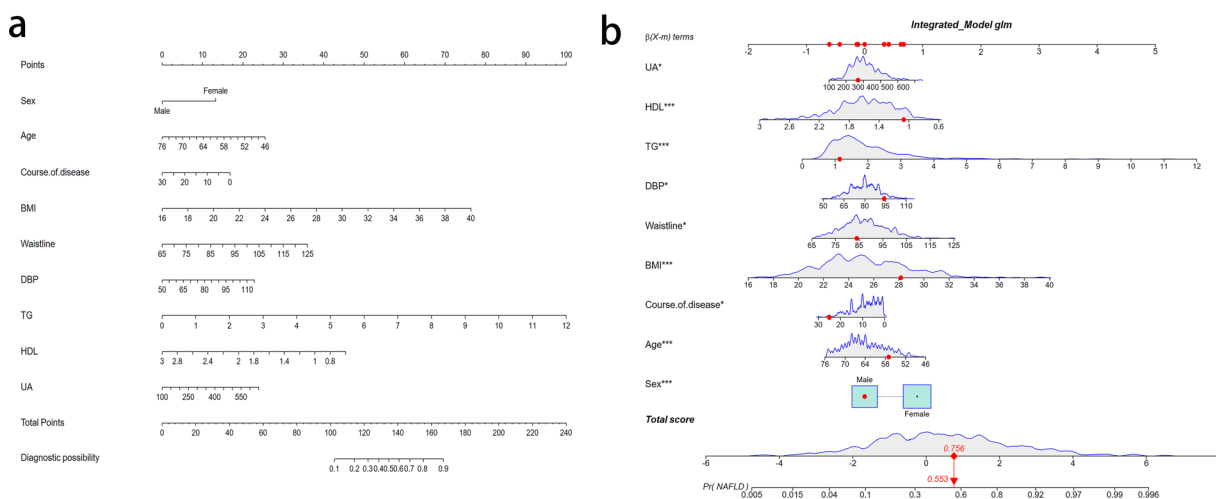
Fig. 3 a Risk factors of sex, age, course of disease, BMI, waistline, DBP, TG,HDL and UA for nomogram prediction model. **b** Dynamic nomogram used as an example.

The significance of the asterisks beside each variable in part **b** represent importance of all the risk factors

training set and 0.809 in the validation set (Fig. 4), which indicates moderately good performance (Fig. 4).

A calibration plot and Hosmer–Lemeshow test were used to calibrate the predictive model. From the calibration curves, the predictive model and the validation set showed a very good degree of fit. As shown by the Hosmer–Lemeshow test, the predicted and actual probability were highly consistent (training set, $P = 0.906$; validation set, $P = 0.049$) (Fig. 5).

The DCA demonstrated that the threshold probability of the prediction model in the training set and validation set is 48–91% and 44–82%, respectively (Fig. 6).

DISCUSSION

According to the predictive model we established of NAFLD in T2DM, nine indicators, namely sex, TG, BMI, waistline, DBP, UA, course of disease, age and HDL-C, are independent risk factors for NAFLD in patients with T2DM.

Although most of the risk factors related to the occurrence of NAFLD in this patient population are consistent and similar to those reported in previous studies, this study differs from the perspective of statistical research methods in that the previous studies used only logistic regression analysis. Our aim was to use

novel statistical methods to identify the risk factors for NAFLD based on previous studies and explain the problem from different perspectives. Previous studies used a single factor analysis to validate a multivariate analysis or performed stepwise regression and recycling logistic regression analyses of the process to obtain the results. For example, Zou and colleagues used multivariate logistic regression analysis to analyze the risk factors of NAFLD in T2DM, with their results showing that the high TG and age were risk factors [18]. After adjusting for all other potential confounding factors, Pawar et al. reported that only systolic hypertension was an independent risk factor for NAFLD on binary logistic regression [19]. However, in this process, various confounding factors must be considered as variables along with the problem of multicollinearity. In the present study, we determined that the LASSO regression analysis provided a better solution to this problem and provided more accurate results. The greatest difference between our approach and those of previous studies using a logistic regression analysis is that we randomly divided the study population into two groups at a 3:1 ratio for external verification. Variables were screened using the LASSO regression analysis, and a traditional logistic regression analysis was also performed. ROC, calibration and DCA
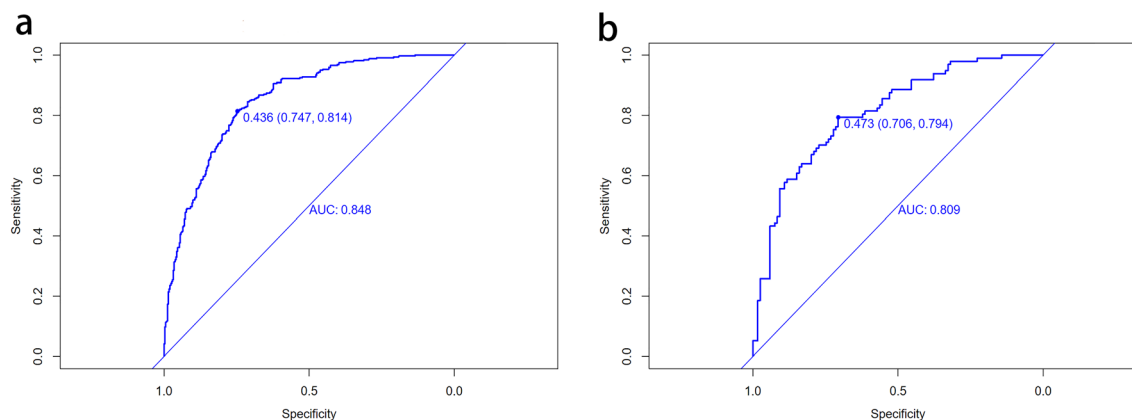

Fig. 4 Receiver operating characteristic curve (ROC) validation of the NAFLD risk nomogram prediction. The y-axis represents the true positive rate of the risk prediction, the x-axis represents the false positive rate of

the risk prediction. The thick blue line represents the performance of the nomogram in the training set (**a**) and validation set (**b**)
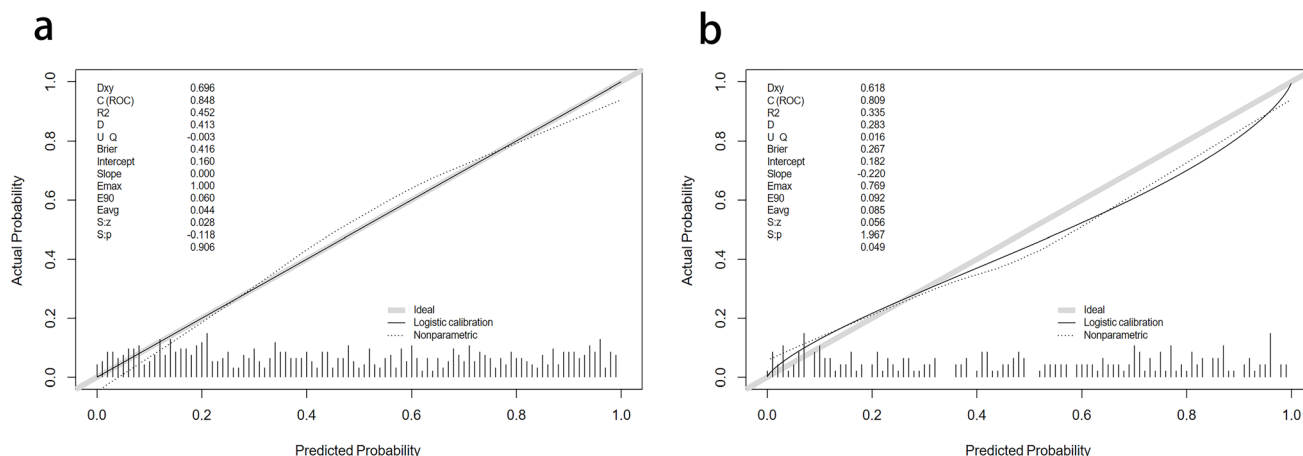
a

b

Fig. 5 Calibration curves of the predictive NAFLD risk nomogram. The *y*-axis represents actual diagnosed cases of NAFLD, the *x*-axis represents the predicted risk of NAFLD. The diagonal dotted line represents a perfect prediction by an

ideal model, the solid line represents the performance of the training set (**a**) and validation set (**b**), with the results indicating that a closer fit to the diagonal dotted line represents a better prediction
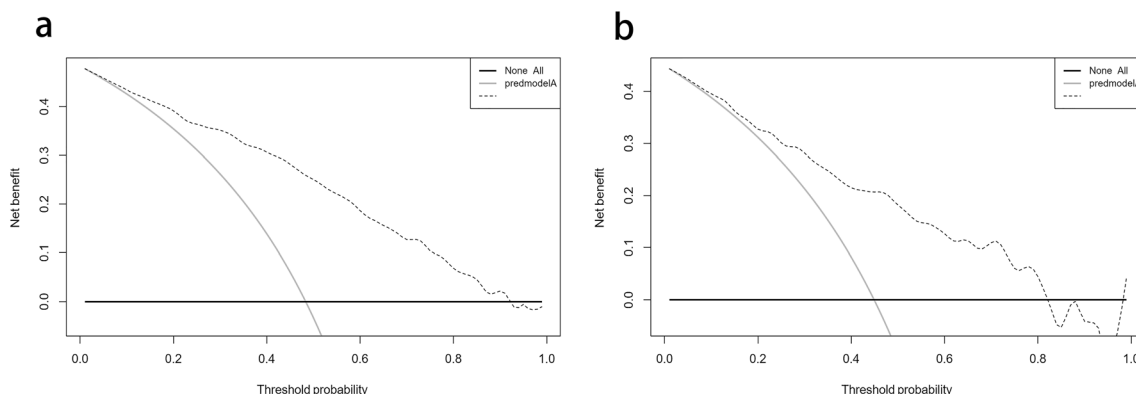
a

b

Fig. 6 Decision curve analysis for the NAFLD risk nomogram. The *y*-axis measures the net benefit. The thick solid line represents the assumption that all patients have no NAFLD, the thin solid line represents the assumption

that all patients have NAFLD, the dotted line represents the risk nomogram. **a** From the training set, **b** from the validation set

curves were constructed for the training and validation sets to verify the accuracy and stability the model.

The present study does lack a review of other factors contributing to NAFLD in patients with T2DM, including smoking, alcohol consumption, diet, other lifestyle factors, some biochemical parameters and some pharmacological parameters. For example, a cross-sectional survey showed a strong relationship between a

family/personal history of metabolic traits and the development of NAFLD [20]. Another survey of 340 participants showed that the proportion of subjects with NAFLD increased according to glucose tolerance status, with subjects with T2DM having a higher prevalence of severe NAFLD than those with normal glucose tolerance. Increased hepatic steatosis has been significantly associated with the presence of T2DM [21]. Another study showed that

natural killer cells and innate lymphoid cells make a large contribution to metabolism, with a focus on their contribution to the development of NAFLD [22]. The aforementioned factors, including smoking, alcohol consumption and diet, were not included in the initial phase of this study. Therefore, we were unable to determine whether these factors would cause NAFLD in patients with T2DM. Future studies will include a more detailed investigation by collecting samples and analyzing indicators in the population of patients with T2DM, including further analysis of the factors influencing NAFLD in Chinese patients with T2DM, by involving a larger sample of people with T2DM and a more comprehensive list of factors.

Association of Sex with NAFLD

The impact of various risk factors for complications in T2DM may differ by sex. Men with T2DM appear to suffer more microvascular complications, while women have higher morbidity and mortality associated with macroangiopathy disease and also fare worse psychologically [23]. Some studies have shown that sex distinctions do exist in T2DM patients and NAFLD patients. Arnetz et al. reported that the presence of diabetes confers the potential for a greater risk for vascular complications in women compared with men due to a number of factors, including the contribution of sex hormones and sex-specific risk factors [23]. At any given age, women have about half the risk of cardiovascular disease relative to men, and estrogens can collectively influence obesity, inflammatory disease, diabetes, cancers, metabolic syndrome and NAFLD [24]. Interestingly, our results show that men with diabetes are more likely to have NAFLD than women. More epidemiological research needs to be done to confirm this association.

Association of TG and HDL-C with NAFLD

High TG levels and low HDL-C levels are common in patients with NAFLD [9]. Patients with higher lipid ratios (TC:HDL-C or TG:HDL-C ratios) have been found to have a significantly greater risk for advanced NAFLD and diabetes [25]. In a large, cross-sectional study conducted among 44,767 Taiwanese patients, the overall prevalence rate of NAFLD in the patient cohort was 53.76%; however, the NAFLD prevalence rate for those with the lowest TC:HDL-C and TG:HDL-C ratios was 33.41% in comparison to the prevalence rate of 78.04% in the group with the highest ratios [9]. It has also been reported that HDL-C may directly alter glucose metabolism by promoting pancreatic β-cell insulin secretion [26], and epidemiological studies and laboratory analyses have demonstrated that low levels of HDL-C are associated with a higher risk of T2DM [27, 28].

Association of BMI and Waistline Measurement with NAFLD

In a meta-analysis study of global epidemiology, obesity prevalence among patients with NAFLD was estimated at 51.3% [29], and statistics in the National Diabetes Statistics Report indicate that the prevalence of obesity/overweight in diabetes is as high as 87.5% in adults with diabetes in the USA [30]. Obesity has been reported to be an independent risk factor for both T2DM [31] and NAFLD [9]. An excessive BMI and visceral obesity are the most common and well-documented risk factors for NAFLD, also contributing to metabolic syndrome and other complications [32]. In diabetes, individuals with excessive fat storage have been noted to have higher insulin levels and are said to have insulin resistance, which is described as the cause of T2DM [33]. The entire spectrum of obesity, ranging from overweight to obese and severely obese, is associated with NAFLD. In our study, the proportion of patients with an excessive BMI was significantly larger in patients with T2DM and NAFLD, compared to T2DM patients without NAFLD. Our results agree with those from studies reporting that an excessive BMI and large waistline can promote the development of diabetes and NAFLD.

Association of DBP with NAFLD

Interestingly, it has been demonstrated that patients with uncontrolled blood pressure

despite anti-hypertensive therapy are at increased risk for developing diabetes mellitus [34]. Also, the average blood pressure level of diabetic patients is higher than that of non-diabetic patients, and elevated blood pressure levels have also been recognized as a risk factor in diabetic patients [35,36]. A systematic review based on 100,354 participants in 40 trials indicated that, among patients with T2DM, lower blood pressure is associated with improved mortality and other clinical outcomes and that patients with a baseline blood pressure of ≥ 140 mmHg have lower risk rate [37]. NAFLD has a strong association with metabolic syndrome, and blood pressure is one of the risk factors for the occurrence of metabolic syndrome [37]. Our research also confirmed that there is a close relationship between blood pressure and NAFLD in the diabetic patient population.

Association of SUA with NAFLD

The level of SUA indicates the degree of metabolism of uric acid in an individual. Persons with increased SUA are at higher future risk of metabolic syndrome, T2DM and cardiovascular diseases, independent of other known risk factors, suggesting a potential role of SUA in the pathogenesis of these diseases. As reported in previous studies, SUA has emerged as an independent predictor for severity of liver damage in NAFLD development [38]. Patients with T2DM and NAFLD have a higher possibility of developing chronic kidney disease [39]. In a study of a US population, the likelihood of NAFLD increased with increasing SUA levels among the total study population, with individuals with NAFLD having higher SUA levels than those without NAFLD and the prevalence rate of NAFLD progressively increasing in parallel to increasing SUA [40]. In prospective studies, the elevation of SUA has been shown to be independently associated with a higher incidence of NAFLD after adjustment for potential confounders [41]. However, diabetic patients with persistent hyperuricemia have an increased risk of diabetic complications, especially renal disease [42]. Consequently, SUA is a major predictor of NAFLD.

Association of Course of Disease and Age with NAFLD

Our study also confirmed that the course of disease and age are protective factors for T2DM with NAFLD in the Chinese population. Age at the time of diabetes onset and the course of diabetes are closely related to the development of NAFLD. We speculate that the earlier the onset of diabetes, the longer the course of T2DM, the higher the likelihood of developing NAFLD. The course of T2DM also shows the same trend as age—that is, the earlier the onset of diabetes, the greater the probability of NAFLD. Our results also fill in the shortcomings of published reports in this field.

Several studies have confirmed that a nomogram is a new, non-invasive, risk prediction model that is important for screening and clinic practice [38–43]. The nomogram is now widely used in risk prediction for various diseases, including hypertension, stroke, among others [44, 45]. Application of this new model can accurately screen related variables and indicators and identify the most suitable risk factors; this model is of great significance in the primary and secondary prevention of diabetes associated with chronic diseases [15,46].

Once T2DM occurs, patients will face the burden of lifelong illness and development-related complications [47], which may lead to a serious decline in quality of life and an increased risk of death. Diabetes with NAFLD causes even more irreversible effects on the patient's metabolic system [27,48]. The current situation in China is such that elderly diabetic patients are unable to go to the hospital for a timely regular liver pathological examination due to inconvenience, difficulty in getting medical treatment, among other factors. The system of door-to-door examination by the family doctor currently promoted in China is a godsend for these patients, and the conclusions drawn from these examinations provide family doctors with reliable methods for universal screening, as well as a more scientific reference basis for treatment. Consequently, relevant indicators, such as those reported here, are of great value in practical applications [49].

Limitations

First, this study was limited by funding, and the authors were unable to collect data on liver fibrosis and cirrhosis for inclusion in the analysis; application of such data will be the focus of future research. Second, based on the existing screening results, the authors identified 24 conventional indicators. In future research, we will assess some of these indicators to enrichen the established models as these may be related to the occurrence and development of NAFLD, although current information is insufficient. Finally, if the population sample size can be further expanded to the entire T2DM population in the Pudong district or to that in all of Shanghai, the established model could be applied well.

CONCLUSION

The nine indicators verified by nomogram in this analysis, including age, course of disease, PBG, HbA1c, UCR, UMA and SBP, are very meaningful in terms of identifying risk for NAFLD in T2DM. Also, these indicators can have a certain effect on the early screening and the timely prevention of the progress of related complications. As a result, introducing these nine indicators in the risk nomogram is useful for the prediction of NAFLD risk in T2DM individuals.

ACKNOWLEDGEMENTS

with professional medical training. The previous study and data collection were approved by the Ethics Committee of Shanghai Oriental Hospital affiliated to Tongji University (Batch number: [2017] Research Review No. 20). All patients provided written informed consent for anonymized publication of their data. The study was carried out in accordance with the Helsinki Declaration of 1964, and its later amendments.

Data Availability. Considering the privacy of patients, if readers have similar research and want to obtain data related to the manuscript, they can contact the corresponding author.

REFERENCES

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14:88–98.

2. Kramer CK, Retnakaran R. Liver enzymes and type 2 diabetes: a complex two-way relationship. J Diabetes Complicat. 2013;27:301–2.

3. Lee WS, Kim J. Diabetic cardiomyopathy: where we are and where we are going. Korean J Intern Med. 2017;32:404–21.

4. Zelnick LR, Weiss NS, Kestenbaum BR, et al. Diabetes and CKD in the United States population, 2009–2014. Clin J Am Soc Nephrol. 2017;12:1984–90.

5. Brunton S. Pathophysiology of type 2 diabetes: the evolution of our understanding. J Fam Pract. 2016;65(4 Suppl):supp_az_0416.

6. Sarparanta J, Garcia-Macia M, Singh R. Autophagy and mitochondria in obesity and type 2 diabetes. Curr Diabetes Rev. 2017;13:352–69.

7. Araujo AR, Rosso N, Bedogni G, Tiribelli C, Bellentani S. Global epidemiology of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: what we need in the future. Liver Int. 2018;38(Suppl 1):47–51.

8. Hazlehurst JM, Woods C, Marjot T, Cobbold JF, Tomlinson JW. Non-alcoholic fatty liver disease and diabetes. Metabolism. 2016;65:1096–108.

9. Younossi ZM, Golabi P, de Avila L, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. J Hepatol. 2019;71:793–801.

10. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. Nat Rev Gastroenterol Hepatol. 2017;14:32–42.

11. Pang Y, Kartsonaki C, Turnbull I, et al. Diabetes, plasma glucose, and incidence of fatty liver, cirrhosis, and liver cancer: a prospective study of 0.5 million people. Hepatology. 2018;68:1308–18.

12. Lee YH, Cho Y, Lee BW, et al. Nonalcoholic fatty liver disease in diabetes. Part I: Epidemiology and diagnosis. Diabetes Metab J. 2019;43:31–45.

13. Fruci B, Giuliano S, Mazza A, Malaguarnera R, Belfiore A. Nonalcoholic fatty liver: a possible new target for type 2 diabetes prevention and treatment. Int J Mol Sci. 2013;14:22933–66.

14. Li W, Xie B, Qiu S, et al. Non-lab and semi-lab algorithms for screening undiagnosed diabetes: A cross-sectional study. EBioMedicine. 2018;35:307–16.

15. Wang H, Zhang L, Liu Z, et al. Predicting medication nonadherence risk in a Chinese inflammatory rheumatic disease population: development and assessment of a new predictive nomogram. Patient Prefer Adherence. 2018;12:1757–65.

16. Latti S, Niinivehmas S, Pentikainen OT. Rocker: open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. J Cheminform. 2016;8:45.

17. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.

18. Zou D, Ye Y, Zou N, Yu J. Analysis of risk factors and their interactions in type 2 diabetes mellitus: a cross-sectional survey in Guilin. China J Diabetes Investig. 2017;8:188–94.

19. Pawar SV, Zanwar VG, Choksey AS, et al. Most overweight and obese Indian children have nonalcoholic fatty liver disease. Ann Hepatol. 2016;15: 853–61.

20. Bhadoria AS, Kedarisetty CK, Bihari C, et al. Impact of family history of metabolic traits on severity of non-alcoholic steatohepatitis related cirrhosis: a cross-sectional study. Liver Int. 2017;37:1397–404.

21. Chon YE, Kim KJ, Jung KS, et al. The relationship between type 2 diabetes mellitus and non-alcoholic fatty liver disease measured by controlled attenuation parameter. Yonsei Med J. 2016;57:885–92.

22. Luci C, Vieira E, Perchet T, Gual P, Golub R. Natural killer cells and type 1 innate lymphoid cells are new actors in non-alcoholic fatty liver disease. Front Immunol. 2019;10:1192. https://doi.org/10.3389/fimmu.2019.01192.

23. Arnetz L, Ekberg NR, Alvarsson M. Sex differences in type 2 diabetes: focus on disease course and outcomes. Diabetes Metab Syndr Obes. 2014;7: 409–20.

24. Chen KL, Madak-Erdogan Z. Estrogens and female liver health. Steroids. 2018;133:38–433.

25. Wu KT, Kuo PL, Su SB, et al. Nonalcoholic fatty liver disease severity is associated with the ratios of total cholesterol and triglycerides to high-density lipoprotein cholesterol. J Clin Lipidol. 2016;10(420–5):e1.

26. Femlak M, Gluba-Brzozka A, Cialkowska-Rysz A, Rysz J. The role and function of HDL in patients with diabetes mellitus and the related cardiovascular risk. Lipids Health Dis. 2017;16:207.

27. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: the Atherosclerosis Risk in Communities study. Diabetes Care. 2005;28:2013–8.

28. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med. 2007;167:1068–74.

29. Saitta C, Pollicino T, Raimondo G. Obesity and liver cancer. Ann Hepatol. 2019;18:810–5.

30. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United States, 2011–2014. NCHS Data Brief, no219. Hyattsville: National Center for Health Statistics; 2015. p. 1–8.

31. Li S, Xiao J, Ji L, et al. BMI and waist circumference are associated with impaired glucose metabolism and type 2 diabetes in normal weight Chinese adults. J Diabetes Complicat. 2014;28:470–6.

32. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2018;15:11–20.

33. Malone JIHansen BC. Does obesity cause type 2 diabetes mellitus (T2DM)? Or is it the opposite? Pediatr Diabetes. 2019;20:5–9.

34. Cryer MJ, Horani T, DiPette DJ. Diabetes and hypertension: a comparative review of current guidelines. J Clin Hypertens (Greenwich). 2016;18: 95–100.

35. [No Authors Listed]. Hypertension in Diabetes Study (HDS): I. Prevalence of hypertension in newly presenting type 2 diabetic patients and the association with risk factors for cardiovascular and diabetic complications. J Hypertens. 1993;11:309–17.

36. Pechere-Bertschi A, Greminger P, Hess L, Philippe J, Ferrari P. Swiss Hypertension and Risk Factor Program (SHARP): cardiovascular risk factors management in patients with type 2 diabetes in Switzerland. Blood Press. 2005;14:337–44.

37. Emdin CA, Rahimi K, Neal B, et al. Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. JAMA. 2015;313:603–15.

38. Lombardi R, Pisano G, Fargion S. Role of serum uric acid and ferritin in the development and progression of NAFLD. Int J Mol Sci. 2016;17:548.

39. Targher G, Byrne CD. Clinical review: nonalcoholic fatty liver disease: a novel cardiometabolic risk factor for type 2 diabetes and its complications. J Clin Endocrinol Metab. 2013;98:483–95.

40. Shih MH, Lazo M, Liu SH, et al. Association between serum uric acid and nonalcoholic fatty liver disease in the US population. J Formos Med Assoc. 2015;114:314–20.

41. Fan N, Zhang L, Xia Z, et al. Sex-specific association between serum uric acid and nonalcoholic fatty liver disease in type 2 diabetic patients. J Diabetes Res. 2016;2016:3805372.

42. Johnson RJ, Perez-Pozo SE, Sautin YY, et al. Hypothesis: could excessive fructose intake and uric acid cause type 2 diabetes? Endocr Rev. 2009;30: 96–116.

43. Han K, Yun JS, Park YM, et al. Development and validation of a risk prediction model for severe hypoglycemia in adult patients with type 2 diabetes: a nationwide population-based cohort study. Clin Epidemiol. 2018;10:1545–59.

44. Abraldes JG, Bureau C, Stefanescu H, et al. Noninvasive tools and risk of clinically significant portal hypertension and varices in compensated cirrhosis: the "Anticipate" study. Hepatology. 2016;64: 2173–84.

45. Cappellari M, Turcato G, Forlivesi S, et al. START-ING-SICH nomogram to predict symptomatic intracerebral hemorrhage after intravenous thrombolysis for stroke. Stroke. 2018;49:397–404.

46. Chung SM, Park JC, Moon JS, Lee JY. Novel nomogram for screening the risk of developing diabetes in a Korean population. Diabetes Res Clin Pract. 2018;142:286–93.

47. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther. 2008;88:1254–64.

48. Radaelli MG, Martucci F, Perra S, et al. NAFLD/NASH in patients with type 2 diabetes and related treatment options. J Endocrinol Invest. 2018;41: 509–21.

49. Ahmed MH, Husain NEO, Almobarak AO. Nonalcoholic Fatty liver disease and risk of diabetes and cardiovascular disease: what is important for primary care physicians? J Family Med Primary Care. 2015;4:45–52.