

Development and internal validation of a routine metabolic screening model for hyperuricemia among gas station workers in Guangxi, China

Jinbo Wang

Youjiang Medical College for Nationalities

Xinna He

Youjiang Medical College for Nationalities

Yinxia Lin

Youjiang Medical College for Nationalities

Zhuoheng Chen

Youjiang Medical College for Nationalities

Jianing Cong

Youjiang Medical College for Nationalities

Yaqin Pang

pangyaqin@126.com

Youjiang Medical College for Nationalities

Guangzi Qi

Youjiang Medical College for Nationalities

Research Article

Keywords: Hyperuricemia, Gas station workers, Routine metabolic indicators, LASSO regression, Logistic regression, Nomogram, Prediction model, Occupational health examination

Posted Date: September 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10456456/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background Hyperuricemia is a common metabolic abnormality in occupational health examinations and is closely associated with cardiovascular disease, chronic kidney disease, and various metabolic disorders. Gas station workers may face an elevated risk of uric acid abnormality due to occupational characteristics such as shift work, prolonged standing, and irregular lifestyles. Current identification of hyperuricemia relies mainly on serum uric acid measurements, and there remains a lack of simple screening tools based on routine metabolic indicators that are applicable to gas station workers.

Methods We conducted a cross-sectional study of 902 gas station workers in a region of Guangxi, China. Data on demographic characteristics, lifestyle factors, occupational factors, anthropometric measurements, and routine metabolic laboratory indicators were collected through questionnaire surveys, physical examinations, and laboratory tests. Hyperuricemia was defined as serum uric acid > 420 $\mu\text{mol/L}$ in men, > 360 $\mu\text{mol/L}$ in women, or a self-reported previous diagnosis by a physician. To develop a routine metabolic screening model suitable for occupational health settings, the main model excluded serum uric acid, urea, creatinine, estimated glomerular filtration rate, and urinary parameters. Participants were randomly divided into training and validation sets at a 7:3 ratio stratified by hyperuricemia status. Nineteen candidate predictors were included in the training set. Least absolute shrinkage and selection operator (LASSO) regression was used for variable selection, and the selected variables were incorporated into a multivariable logistic regression model to construct a prediction model and nomogram. Model discrimination was evaluated using the area under the receiver operating characteristic curve (AUC), calibration was assessed using Brier score and bootstrap calibration curves, clinical net benefit was evaluated using decision curve analysis, and risk stratification ability was assessed by grouping predicted probabilities into quartiles.

Results Among the 902 participants, 232 had hyperuricemia, yielding a prevalence of 25.7%. Under the λ_{1se} criterion, LASSO regression identified four predictors: sex, waist circumference, triglycerides, and high-density lipoprotein cholesterol (HDL-C). Multivariable logistic regression showed that waist circumference and triglycerides were positively associated with hyperuricemia, while HDL-C was negatively associated; the association for male sex showed an increasing trend but did not reach statistical significance after multivariable adjustment. The final logistic regression equation was: $\text{logit}(P) = -6.1767 + 0.4134 \times \text{Male} + 0.0742 \times \text{Waist circumference} + 0.1477 \times \text{Triglycerides} - 1.1388 \times \text{HDL-C}$. The AUC was 0.773 (95% CI: 0.733–0.814) in the training set and 0.815 (95% CI: 0.757–0.874) in the validation set, indicating good discriminative ability. The Brier scores were 0.155 and 0.143 for the training and validation sets, respectively. Bootstrap internal validation showed a corrected C-index of 0.768, a calibration slope of 0.972, and a mean absolute error of 0.011, suggesting stable calibration. Decision curve analysis demonstrated that the model provided net benefit over both "treat all" and "treat none" strategies within certain threshold probability ranges. In the validation set, hyperuricemia prevalence increased from 5.9% in the lowest quartile to 58.8% in the highest quartile (P for trend < 0.001), confirming good risk stratification ability.

Conclusions We developed a routine metabolic screening model for hyperuricemia using readily available occupational health examination indicators. The final model includes four easily obtainable variables—sex, waist circumference, triglycerides, and HDL-C—and demonstrates good discrimination, calibration, and risk stratification. The model does not rely on serum uric acid or renal function indicators and may serve as an auxiliary tool for initial screening and risk stratification of hyperuricemia in occupational health settings. Due to the cross-sectional design, the predicted probabilities reflect current hyperuricemia status rather than future risk; external validation in other regions and occupational populations, as well as prospective studies, are warranted.

1 Introduction

Hyperuricemia (HUA) is a common metabolic abnormality caused by purine metabolism disorder or reduced uric acid excretion. It serves as the biochemical basis for the development and progression of gout and is closely associated with obesity, dyslipidemia, hypertension, metabolic syndrome, chronic kidney disease, and cardiovascular diseases [1–4]. In recent years, due to changes in dietary patterns, the prevalence of overweight and obesity, and lifestyle transitions, the prevalence of hyperuricemia among Chinese adults has remained persistently high. Previous national surveys and health examination-based studies have shown that hyperuricemia prevalence is elevated in Chinese adults, with higher risk observed in men, and among individuals with obesity, dyslipidemia, and hypertension [5–7]. Hyperuricemia often lacks specific symptoms in its early stages, and reliance solely on clinical manifestations may delay early intervention. Therefore, identifying individuals at high likelihood of current hyperuricemia using routine health examination indicators in occupational health settings has practical significance for chronic disease prevention and occupational population health management.

Gas station workers represent an occupational group with distinct work-related characteristics. This population may be exposed to gasoline vapors and volatile organic compounds such as benzene, toluene, ethylbenzene, and xylene during their work, and their work schedules often involve shift work, prolonged standing, irregular rest patterns, and unstable dietary habits [8–11]. These occupational and lifestyle factors may interact with weight gain, dyslipidemia, blood pressure abnormalities, and metabolic syndrome, resulting in a dual influence of occupational and lifestyle factors on metabolic risk in gas station workers. Existing studies on hyperuricemia have largely focused on community residents, health examination populations, or clinical patients, with relatively limited research on hyperuricemia screening among petrochemical-related occupational groups such as gas station workers. Occupational health examinations have the advantages of wide coverage, standardized indicators, and regular repetition; if simple and stable predictors could be identified from routine health examination indicators, they could provide an operational tool for early identification of uric acid abnormalities in occupational populations.

Hyperuricemia is closely related to multiple metabolic indicators. Obesity, particularly central obesity, may influence serum uric acid levels through mechanisms including insulin resistance, hyperinsulinemia, and increased renal tubular uric acid reabsorption [4, 12–13]. Dyslipidemia is also an important accompanying feature of hyperuricemia, with elevated triglycerides and reduced high-density lipoprotein

cholesterol commonly co-occurring with insulin resistance and metabolic syndrome, and are closely associated with hyperuricemia [12–15]. Uric acid is primarily excreted by the kidneys. Although renal function indicators such as creatinine, urea, and estimated glomerular filtration rate are closely related to uric acid levels, these indicators are more proximal to the uric acid excretion pathway. Directly including them in a screening model may cause the model to rely excessively on renal function information, thereby weakening its independent explanatory value as a routine metabolic screening tool. Therefore, constructing a screening model that does not depend on serum uric acid or renal function indicators, but is primarily based on routine metabolic indicators such as anthropometric measurements and blood lipids, better aligns with the "early, simple, and scalable" application needs in occupational health examinations.

In recent years, the application of prediction models in early identification and risk stratification of chronic diseases has increased. Compared with single risk factor assessment, multivariable prediction models can integrate information from multiple dimensions and improve individualized risk assessment capacity [16–18]. LASSO regression can perform variable selection and coefficient shrinkage when there are many candidate predictors with potential correlations, helping to reduce model complexity and overfitting risk [19]. Nomograms can transform logistic regression models into intuitive scoring tools, presenting prediction results as individualized probabilities, facilitating application in health examination and management settings [20–21]. According to prediction model reporting guidelines, model evaluation should consider not only discrimination but also calibration, clinical net benefit, and risk stratification performance [17–18, 21–23]. Therefore, receiver operating characteristic (ROC) curves, bootstrap calibration curves, Brier score, decision curve analysis, and predicted probability stratification can be used collectively to evaluate model utility and stability.

Based on this, the present study focused on gas station workers in a region of Guangxi, China. Using data from questionnaire surveys, physical examinations, and laboratory tests, we aimed to screen routine metabolic predictors that are readily available in occupational health examinations without including serum uric acid or renal function indicators, and to develop a routine metabolic screening model for hyperuricemia using LASSO regression and multivariable logistic regression, with further construction of a nomogram. The model was internally validated in terms of discrimination, calibration, clinical net benefit, and risk stratification. This study aimed to develop a simple, interpretable, and applicable screening tool for hyperuricemia in occupational health settings, providing a reference for early identification and stratified management of current uric acid abnormalities among gas station workers.

2 Methods

2.1 Study population

Data for this study were derived from a cross-sectional survey conducted from July 1 to July 30, 2025, in Nanning, Guangxi Zhuang Autonomous Region, China. A cluster sampling method was employed, with

gas stations and oil depots under the Guangxi Nanning Petroleum Branch of Sinopec Sales Co., Ltd. as the sampling units. Active workers from both the urban districts of Nanning and its subordinate counties—Long'an, Mashan, Wuming, Shanglin, Binyang, Litang, and Hengzhou—were recruited to participate in questionnaire surveys, physical examinations, and laboratory tests. A total of 1,044 workers were initially surveyed, and after applying the inclusion and exclusion criteria, 902 participants were included in the final analysis. Inclusion criteria were: (1) age ≥ 18 years; (2) active employment for ≥ 1 year; (3) completion of the questionnaire, physical examination, and laboratory tests; and (4) completeness of data for the main study variables. Exclusion criteria were: (1) pregnant or lactating women; (2) individuals with severe heart, liver, kidney, or hematologic diseases, or malignancy; (3) use of urate-lowering medications within the past 3 months; and (4) missing data on key variables. The study was approved, and all participants provided written informed consent.

2.2 Outcome definition and candidate predictors

The outcome variable of this study was hyperuricemia (HUA). Based on serum uric acid measurements and previous diagnosis information, hyperuricemia was defined as serum uric acid $> 420 \mu\text{mol/L}$ in men, $> 360 \mu\text{mol/L}$ in women, or a self-reported previous diagnosis of hyperuricemia by a physician. This variable was included in the prediction model as a binary outcome.

A self-designed structured questionnaire was used to collect information on demographic characteristics, lifestyle factors, and occupational-related information (Additional file 1). Demographic characteristics included age, sex, educational level, marital status, monthly income, and household registration type. Lifestyle factors included smoking, drinking, and physical activity. Occupational-related information included work seniority, job category, daily working hours, work intensity, shift work schedule, and number of night shifts per month. Smoking was defined as currently smoking ≥ 1 cigarette per day for ≥ 6 consecutive months. Drinking was defined as currently consuming alcohol ≥ 3 times per week for ≥ 6 consecutive months. Physical activity was expressed as total days of physical activity per week.

Physical examinations included measurements of height, weight, waist circumference, and blood pressure. Body mass index (BMI) was calculated as weight (kg) divided by height squared (m^2). Waist circumference was measured at the midpoint between the lower margin of the rib cage and the iliac crest at the midaxillary line. Blood pressure was measured using a calibrated electronic sphygmomanometer after the participant had rested in a seated position for at least 5 minutes, with the measurement taken on the right upper arm; the average of repeated measurements was recorded.

Venous blood samples were collected from all participants after fasting for at least 8 hours. Laboratory tests included serum uric acid, urea, creatinine, fasting blood glucose, triglycerides, total cholesterol, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol, and complete blood count. Urine specimens were used to test for urine protein, urinary occult blood, urine specific gravity, and urine pH. Estimated glomerular filtration rate (eGFR) was calculated using the CKD-EPI equation.

2.3 Model development

The primary model of this study was positioned as a routine metabolic screening model for hyperuricemia. To avoid the model being dominated by serum uric acid itself or renal function indicators closely related to uric acid excretion, the candidate predictors for the primary model excluded serum uric acid, urea, creatinine, eGFR, and urinary parameters. Ultimately, 19 candidate predictors were included, covering demographic characteristics, lifestyle factors, occupational factors, anthropometric indicators, routine metabolic laboratory indicators, and disease history or examination-based diagnostic indicators. Specifically, these included age, sex, smoking, drinking, total days of physical activity, work seniority, job category, BMI, waist circumference, systolic blood pressure, diastolic blood pressure, fasting blood glucose, triglycerides, total cholesterol, HDL-C, low-density lipoprotein cholesterol, hypertension, diabetes mellitus, and hyperlipidemia.

To minimize the influence of extreme values on model estimation, participants were first divided into training and validation sets at a 7:3 ratio. The 1st and 99th percentiles of each continuous variable were calculated in the training set, and these thresholds were then applied to both the training and validation sets for winsorization (i.e., values below the 1st percentile were replaced with the 1st percentile, and values above the 99th percentile were replaced with the 99th percentile). All variable selection and model fitting were performed in the training set, while the validation set was used exclusively for model performance evaluation. Participants with missing data on the outcome variable, primary candidate predictors, or final model variables were excluded from the modeling analysis, resulting in a final analytical sample of 902 participants. Categorical variables were factorized according to their actual meanings, with sex coded as male = 1 and female = 0 as the reference; multicategory variables such as job category were converted into dummy variables when entering the model matrix. All candidate predictors were standardized prior to LASSO regression to reduce the impact of scale differences on variable selection; the final multivariable logistic regression model was fitted using variables on their original scales to facilitate interpretation of odds ratios and construction of the nomogram.

Participants were randomly divided into training and validation sets at a 7:3 ratio. The division was performed using stratified sampling based on hyperuricemia status to ensure that the proportions of hyperuricemia were approximately consistent between the two sets. The training set was used for variable selection, model fitting, and nomogram development; the validation set was used exclusively for model performance evaluation and risk stratification validation, and did not participate in LASSO variable selection or model parameter estimation to avoid data leakage.

Least absolute shrinkage and selection operator (LASSO) regression was used for variable selection in the training set. The penalty parameter λ was determined through 10-fold cross-validation, and the one-standard-error rule (λ_{1se}) was adopted as the final selection criterion to reduce overfitting risk. Variables with non-zero regression coefficients under the λ_{1se} criterion were incorporated into a multivariable logistic regression model to construct the routine metabolic screening model for hyperuricemia. To distinguish it from alternative models in sensitivity analyses, this primary model was designated as Model A1.

The final model included four predictors: sex, waist circumference, triglycerides, and HDL-C. Based on the multivariable logistic regression results, regression coefficients, odds ratios (OR), and 95% confidence intervals (CI) were estimated for each predictor, and a nomogram was further constructed to visually present the contribution of each predictor to the predicted probability of hyperuricemia. The final logistic regression equation was:

$$\text{logit}(P) = -6.1767 + 0.4134 \times \text{Male} + 0.0742 \times \text{Waist circumference} + 0.1477 \times \text{Triglycerides} - 1.1388 \times \text{HDL-C}$$

where P represents the predicted probability of hyperuricemia; Male is coded as 1 for men and 0 for women; Waist circumference is in cm; Triglycerides are in mmol/L; and HDL-C is in mmol/L. The predicted probability was calculated as: $P = \exp[\text{logit}(P)] / \{1 + \exp[\text{logit}(P)]\}$.

2.4 Model evaluation

Model performance was evaluated from four aspects: discrimination, calibration, clinical net benefit, and risk stratification ability. Discrimination was assessed using the area under the receiver operating characteristic curve (AUC), with AUC and its 95% CI calculated for both the training and validation sets. Calibration was evaluated using Brier score and bootstrap calibration curves with 1,000 resampling iterations, and the corrected Dxy, corrected C-index, calibration slope, and mean absolute error were reported.

Clinical net benefit was evaluated using decision curve analysis (DCA), which compared the net benefit of the final model with "treat all" and "treat none" strategies across different threshold probabilities. To further evaluate the risk stratification ability of the model, participants were divided into quartile groups based on the predicted probabilities from the final model. The prevalence of hyperuricemia and Wilson 95% CI were calculated for each group, and trend tests were used to evaluate the association between predicted probability strata and hyperuricemia prevalence.

2.4 Statistical Analysis

All statistical analyses were performed using R software. Continuous variables following an approximately normal distribution were presented as mean $\pm$ standard deviation, while those with skewed distributions were presented as median (interquartile range). Categorical variables were presented as frequencies and percentages. Depending on variable type and distribution, between-group comparisons were performed using t-tests, Mann-Whitney U tests, chi-square tests, or Fisher's exact tests.

LASSO regression was performed using the glmnet package; multivariable logistic regression and nomogram construction were performed using the rms package; ROC curves and 95% confidence intervals for AUC were calculated using the pROC package; calibration curves were plotted using bootstrap resampling; and decision curve analysis was performed using the dcurves package. All statistical tests were two-sided, and a P value < 0.05 was considered statistically significant.

To assess the robustness of the model, we further conducted sensitivity analyses and alternative model comparisons. Sensitivity analyses included: comparing LASSO variable selection results under the lambda.min and lambda.1se criteria; comparing baseline characteristics between the training and validation sets to evaluate the balance of sample partitioning; assessing multicollinearity of the final multivariable logistic regression model using variance inflation factor (VIF), with $VIF < 5$ indicating no substantial multicollinearity; repeating model performance evaluation after excluding individuals with $BMI \geq 40 \text{ kg/m}^2$; and evaluating model discrimination stratified by sex. Alternative model comparisons included a minimal metabolic model, an inflammation-extended model, and a VAI-based simplified model. Prediction performance across different models was compared using validation set AUC, Brier score, and ROC curves to evaluate the comprehensive advantages of the final model in terms of discrimination, parsimony, interpretability, and accessibility in occupational health examination settings.

3 Results

3.1 General Characteristics

A total of 902 gas station workers were included in this study, among whom 232 (25.7%) had hyperuricemia (HUA group) and 670 (74.3%) did not (Non-HUA group). Baseline characteristics of the two groups are presented in Table 1.

Regarding demographic and behavioral factors, the proportion of male participants was significantly higher in the HUA group than in the Non-HUA group (71.1% vs. 40.1%, $p < 0.001$). There was no statistically significant difference in age between the two groups [40.57 (8.53) years vs. 40.72 (7.60) years, $p = 0.803$]. The HUA group had higher rates of smoking (30.6% vs. 18.4%) and drinking (29.3% vs. 16.9%) compared with the Non-HUA group, with both differences being statistically significant (both $p < 0.001$). No statistically significant differences were observed between the two groups in educational level, monthly income, total physical activity level, work seniority, or job category distribution (all $p > 0.05$).

In terms of metabolic diseases and anthropometric indicators, the prevalence of hypertension and hyperlipidemia was higher in the HUA group than in the Non-HUA group (25.9% vs. 10.1% and 14.7% vs. 5.7%, respectively; both $p < 0.001$). The prevalence of diabetes mellitus did not differ significantly between the two groups ($p = 1.000$). BMI, waist circumference, systolic blood pressure, and diastolic blood pressure were all significantly higher in the HUA group compared with the Non-HUA group (all $p < 0.001$). Specifically, the HUA group had a BMI of 26.43 (3.58) kg/m^2 and waist circumference of 88.18 (9.74) cm, whereas the Non-HUA group had corresponding values of 23.36 (3.21) kg/m^2 and 78.52 (8.73) cm.

Regarding laboratory indicators, the HUA group had significantly higher levels of fasting blood glucose, triglycerides, total cholesterol, low-density lipoprotein cholesterol, urea, and serum creatinine, while high-density lipoprotein cholesterol and eGFR were significantly lower compared with the Non-HUA group (all

p < 0.01). Serum uric acid levels were markedly higher in the HUA group than in the Non-HUA group [449.50 (417.00, 493.75) $\mu\text{mol/L}$ vs. 304.00 (264.00, 348.00) $\mu\text{mol/L}$, p < 0.001].

In terms of derived indices, VAI was significantly higher in the HUA group than in the Non-HUA group [2.04 (1.24, 3.11) vs. 1.18 (0.79, 1.80), p < 0.001]. Among the four systemic inflammatory markers, only SIRI was slightly higher in the HUA group compared with the Non-HUA group [0.87 (0.63, 1.20) vs. 0.80 (0.58, 1.06), p = 0.024], whereas NLR, MLR, and SII showed no statistically significant differences between the two groups (all p > 0.05). The prevalence of proteinuria showed an increasing trend in the HUA group, but the difference did not reach statistical significance (5.6% vs. 2.7%, p = 0.058); no statistically significant difference was observed in the prevalence of hematuria between the two groups (p = 0.101).

Overall, the HUA group exhibited a more pronounced clustering of metabolic abnormalities, including higher levels of obesity, central obesity, blood pressure, dyslipidemia, and alterations in renal function-related indicators, whereas among the systemic inflammatory markers, only SIRI showed a mild elevation.

Table 1. Comparison of baseline characteristics between the hyperuricemia and non-hyperuricemia groups

Variables	Overall (n=902)	Non-HUA (n=670)	HUA (n=232)	p value
Age, years	40.68 (7.84)	40.72 (7.60)	40.57 (8.53)	0.803
Male sex, n (%)	434 (48.1)	269 (40.1)	165 (71.1)	<0.001
Education, n (%)				0.072
Junior high school and below	74 (8.2)	58 (8.7)	16 (6.9)	
Senior high school or technical secondary school	425 (47.1)	329 (49.1)	96 (41.4)	
Junior college or vocational college	293 (32.5)	209 (31.2)	84 (36.2)	
Bachelor's degree or above	110 (12.2)	74 (11.0)	36 (15.5)	
Marital status, n(%)				0.003
Married	725 (80.4)	556 (83.0)	169 (72.8)	
Unmarried	146 (16.2)	93 (13.9)	53 (22.8)	
Other	31 (3.4)	21 (3.1)	10 (4.3)	
Monthly income, n (%)				0.058
≤1999	46 (5.1)	37 (5.5)	9 (3.9)	
2000-2999	429 (47.6)	334 (49.9)	95 (40.9)	
3000-3999	278 (30.8)	197 (29.4)	81 (34.9)	
4000-4999	106 (11.8)	75 (11.2)	31 (13.4)	
≥5000	43 (4.8)	27 (4.0)	16 (6.9)	
Smoking, n (%)	194 (21.5)	123 (18.4)	71 (30.6)	<0.001
Drinking, n(%)	181 (20.1)	113 (16.9)	68 (29.3)	<0.001
Total physical activity, days/week	6.50 (3.00, 7.50)	6.50(2.50, 8.00)	6.50(3.00, 7.50)	0.772
Work seniority, years	14.63 (9.71)	14.49 (9.62)	15.03 (9.99)	0.464
Job category (%)				0.153
Fuel attendant	572 (63.4)	438 (65.4)	134 (57.8)	
Station manager	175 (19.4)	123 (18.4)	52 (22.4)	

Variables	Overall (n=902)	Non-HUA (n=670)	HUA (n=232)	p value
Administrative	54 (6.0)	38 (5.7)	16 (6.9)	
Technical	63 (7.0)	41 (6.1)	22 (9.5)	
Other	38 (4.2)	30 (4.5)	8 (3.4)	
Hypertension, n(%)	128 (14.2)	68 (10.1)	60 (25.9)	<0.001
Diabetes mellitus, n (%)	41 (4.5)	30 (4.5)	11 (4.7)	1
Hyperlipidemia, n(%)	72 (8.0)	38 (5.7)	34 (14.7)	<0.001
BMI, kg/m ²	24.15 (3.57)	23.36 (3.21)	26.43 (3.58)	<0.001
Waist circumference, cm	81.01 (9.94)	78.52 (8.73)	88.18 (9.74)	<0.001
SBP, mmHg	122.93 (16.92)	120.81 (16.10)	129.08 (17.76)	<0.001
DBP, mmHg	79.40 (11.28)	78.02 (10.63)	83.38 (12.14)	<0.001
FBG, mmol/L	5.02(4.71, 5.35)	4.99(4.69, 5.34)	5.14(4.85, 5.46)	<0.001
TG, mmol/L	1.23(0.84, 1.93)	1.08(0.77, 1.64)	1.81(1.15, 2.77)	<0.001
TC, mmol/L	5.08 (0.96)	5.02 (0.95)	5.28 (0.94)	<0.001
HDL, mmol/L	1.38(1.22, 1.57)	1.42(1.26, 1.63)	1.27(1.17, 1.44)	<0.001
LDL, mmol/L	2.80 (0.80)	2.71 (0.79)	3.06 (0.76)	<0.001
UA, umol/L	331.00(278.25, 397.75)	304.00(264.00, 348.00)	449.50(417.00, 493.75)	<0.001
UREA, mmol/L	4.70 (1.16)	4.64 (1.15)	4.88 (1.19)	0.007
CREA, mgdl	0.83 (0.17)	0.80 (0.16)	0.92 (0.17)	<0.001
eGFR, mL/min/1.73 m ²	100.56 (13.05)	101.93 (12.25)	96.61 (14.43)	<0.001
VAI	1.35(0.86, 2.19)	1.18(0.79, 1.80)	2.04(1.24, 3.11)	<0.001
SIRI	0.81(0.58, 1.09)	0.80(0.58, 1.06)	0.87(0.63, 1.20)	0.024
NLR	1.62(1.29, 2.06)	1.62(1.29, 2.07)	1.61(1.29, 2.04)	0.687
MLR	0.22(0.18, 0.26)	0.21(0.18, 0.26)	0.22(0.18, 0.25)	0.762
SII	459.49(353.34, 607.17)	457.07(352.49, 606.44)	462.24(354.93, 608.44)	0.727
Proteinuria, n (%)	31 (3.4)	18 (2.7)	13 (5.6)	0.058

Variables	Overall (n=902)	Non-HUA (n=670)	HUA (n=232)	p value
Hematuria, n (%)	137 (15.2)	110 (16.4)	27 (11.6)	0.101

Note: Normally distributed continuous variables are presented as mean (standard deviation), skewed continuous variables as median (interquartile range), and categorical variables as n (%). Between-group comparisons were performed using t-tests, Mann–Whitney U tests, chi-square tests, or Fisher's exact tests as appropriate based on variable type and distribution. HUA, hyperuricemia; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FBG, fasting blood glucose; TG, triglycerides; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; UA, serum uric acid; CREA, creatinine; eGFR, estimated glomerular filtration rate; VAI, visceral adiposity index; SIRI, systemic inflammatory response index; NLR, neutrophil-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; SII, systemic immune-inflammation index.

3.2 LASSO variable selection results

To reduce overfitting caused by an excessive number of candidate predictors, LASSO regression was applied to the 19 non-uric acid, non-renal function candidate variables in the training set. The LASSO coefficient path plot showed that as the penalty parameter λ gradually increased, the regression coefficients of the candidate variables shrank toward zero, indicating differential contributions of the variables to hyperuricemia identification (Figure 1A). The penalty parameter was determined using 10-fold cross-validation, and the one-standard-error rule (λ_{1se}) was adopted as the final selection criterion to minimize overfitting risk (Figure 1B).

Under the λ_{1se} criterion, four variables with non-zero coefficients were retained: sex, waist circumference, triglycerides, and HDL-C. These four variables were subsequently incorporated into the multivariable logistic regression model to construct the nomogram for hyperuricemia metabolic screening.

3.3 Logistic regression analysis of LASSO-selected variables associated with hyperuricemia

The variables selected by LASSO regression—sex, waist circumference, triglycerides, and HDL-C—were included in logistic regression analyses. Univariable logistic regression showed that male sex, higher waist circumference, and higher triglycerides were associated with a higher likelihood of hyperuricemia, while higher HDL-C was associated with a lower likelihood. Specifically, the odds of hyperuricemia were higher in men than in women (OR = 3.396, 95% CI: 2.322–5.030, $P < 0.001$); each 1-cm increase in waist circumference was associated with higher odds of hyperuricemia (OR = 1.106, 95% CI: 1.083–1.131, $P < 0.001$); each 1-mmol/L increase in triglycerides was associated with higher odds of hyperuricemia (OR = 1.476, 95% CI: 1.296–1.705, $P < 0.001$); and each 1-mmol/L increase in HDL-C was associated with lower odds of hyperuricemia (OR = 0.068, 95% CI: 0.028–0.156, $P < 0.001$).

After further multivariable logistic regression adjustment, waist circumference, triglycerides, and HDL-C remained independently associated with hyperuricemia. Each 1-cm increase in waist circumference was associated with a 7.7% increase in the odds of hyperuricemia (OR = 1.077, 95% CI: 1.051–1.105, $P < 0.001$); each 1-mmol/L increase in triglycerides was associated with a 15.9% increase in the odds of hyperuricemia (OR = 1.159, 95% CI: 1.015–1.329, $P = 0.030$); and each 1-mmol/L increase in HDL-C was associated with lower odds of hyperuricemia (OR = 0.320, 95% CI: 0.119–0.828, $P = 0.021$). The association between male sex and hyperuricemia was attenuated after multivariable adjustment and did not reach statistical significance (OR = 1.512, 95% CI: 0.971–2.361, $P = 0.068$). These findings suggest that central obesity and lipid metabolism abnormalities are the primary contributors to the routine metabolic screening model for hyperuricemia (Figure 2)

3.4 Nomogram development

A nomogram for predicting hyperuricemia risk was constructed based on the multivariable logistic regression model (Figure 3). The final nomogram incorporated four predictors: sex, waist circumference, triglycerides, and HDL-C. The nomogram assigns corresponding points to each predictor; the sum of these points yields a total score, which is then converted to the predicted probability of hyperuricemia.

In the nomogram, male sex, higher waist circumference, and higher triglycerides corresponded to higher points, while higher HDL-C levels corresponded to a lower predicted probability. Among these, waist circumference and HDL-C contributed relatively more to the total score, suggesting that central obesity and lipid metabolism abnormalities are the main contributors to the model's probability estimation. To illustrate the practical application of the nomogram, we provided an example prediction. For a male worker with a waist circumference of 90 cm, triglycerides of 2.0 mmol/L, and HDL-C of 1.0 mmol/L, the $\text{logit}(P)$ calculated from the final logistic regression equation was approximately 0.075, corresponding to a predicted probability of hyperuricemia of approximately 0.519. This example demonstrates that the nomogram can convert routine health examination indicators into individualized probability estimates, thereby facilitating preliminary identification of individuals with a high likelihood of hyperuricemia in occupational health settings.

3.5 Model performance evaluation

The performance of the final prediction model was evaluated using ROC curves, Brier score, bootstrap calibration curves, and decision curve analysis. The model achieved an AUC of 0.773 (95% CI: 0.733–0.814) in the training set and 0.815 (95% CI: 0.757–0.874) in the validation set, indicating good discriminative ability in both sets. The Brier scores were 0.155 and 0.143 for the training and validation sets, respectively, suggesting low overall prediction error between predicted probabilities and observed outcomes. Bootstrap internal validation showed a corrected Dxy of 0.536, a corrected C-index of 0.768, a calibration slope of 0.972, and a mean absolute error of 0.011, indicating stable calibration performance (Table 3).

The ROC curves further showed that the validation set ROC curve was located above the reference line, indicating that the model could effectively distinguish individuals with hyperuricemia from those without (Figure 4A). The bootstrap calibration curve showed that both the apparent curve and the bias-corrected curve were close to the ideal reference line, suggesting good agreement between predicted probabilities and observed outcomes (Figure 4B). Decision curve analysis demonstrated that the final model provided higher net benefit than the "treat all" and "treat none" strategies within certain threshold probability ranges, suggesting potential screening utility in occupational health settings (Figure 4C). As external validation samples were not included in this study, these results remain internal validation findings, and the generalizability of the model requires further validation in other regions and occupational populations.

Table 3. Discrimination, calibration, and internal validation metrics of the hyperuricemia prediction model

Metric	Training set	Validation set / Bootstrap correction
AUC95%CI	0.7730.733–0.814	0.8150.757–0.874
Brier score	0.155	0.143
Bootstrap-corrected Dxy	—	0.536
Bootstrap-corrected C-index	—	0.768
Bootstrap calibration slope	—	0.972
Mean absolute error	—	0.011

Note: AUC, area under the receiver operating characteristic curve; Brier score evaluates the overall prediction error between predicted probabilities and observed outcomes; bootstrap resampling was performed with 1,000 iterations; C-index was derived from Dxy using the formula $C-index = (Dxy + 1)/2$. AUC and Brier score for the training and validation sets were calculated based on the respective datasets; Dxy, corrected C-index, calibration slope, and mean absolute error are bootstrap internal calibration results.

3.6 Predicted probability stratification and hyperuricemia prevalence trends

To further evaluate the risk stratification ability of the final model, participants were divided into quartile groups based on predicted probabilities, and the trends in hyperuricemia prevalence across different predicted probability strata were compared. In the total cohort analysis, hyperuricemia prevalence increased progressively with higher predicted probability strata: 5.3% in Q1 (lowest risk), 13.3% in Q2

(low-intermediate risk), 28.9% in Q3 (high-intermediate risk), and 55.6% in Q4 (highest risk), with a statistically significant trend (P for trend < 0.001) (Figure 5A).

A consistent increasing trend was observed in the validation set. Hyperuricemia prevalence in Q1, Q2, Q3, and Q4 was 5.9% (95% CI: 2.3%–14.2%), 7.4% (95% CI: 3.2%–16.1%), 30.9% (95% CI: 21.2%–42.6%), and 58.8% (95% CI: 47.0%–69.7%), respectively, with a similarly significant trend (P for trend < 0.001) (Figure 5B). These findings indicate that the final model can stratify gas station workers into risk tiers with substantially different probabilities of hyperuricemia, demonstrating good risk stratification ability and potential utility in occupational health screening.

3.7 Sensitivity analysis results

Sensitivity analysis results are presented in Supplementary Tables S1–S9 and Supplementary Figure S1. Baseline characteristics were generally similar between the training and validation sets, indicating good comparability after sample partitioning (Supplementary Table S1). Comparison of different LASSO penalty parameters showed that the lambda.min criterion retained more variables, while the lambda.1se criterion further simplified the model to only four variables—sex, waist circumference, triglycerides, and HDL-C—consistent with the objective of constructing a parsimonious screening model (Supplementary Table S2). All VIF values for the final model were below 2, indicating no substantial multicollinearity (Supplementary Table S3). After excluding individuals with BMI ≥ 40 kg/m², the validation set AUC remained 0.815 (95% CI: 0.757–0.874), consistent with the primary analysis (Supplementary Table S4). Stratified by sex, the AUC in the validation set was 0.739 (95% CI: 0.607–0.871) for females and 0.772 (95% CI: 0.690–0.854) for males, suggesting that the model maintained acceptable discriminative ability across sexes (Supplementary Table S5). Comparison of alternative models showed that the final model had the highest validation set AUC (0.815), slightly outperforming the minimal metabolic model, the inflammation-extended model, and the VAI-based simplified model (Supplementary Table S6, Supplementary Figure S1). To assess whether model performance depended on a single data split, we repeated 100 random 7:3 splits; the mean validation AUC was 0.785 (SD = 0.028), and the primary analysis AUC (0.816) was positioned at the higher end of the repeated distribution, suggesting that model performance was not driven by chance in a single split (Supplementary Table S7). The Hosmer-Lemeshow test in the validation set yielded $\chi^2 = 8.767$ (df = 8, $P = 0.362$), further supporting good calibration (Supplementary Table S8). Bootstrap optimism correction showed that the optimism of the training AUC was 0.004, with a corrected AUC of 0.770, close to the observed value of 0.773, indicating minimal overfitting (Supplementary Table S9). Considering predictive performance, parsimony, interpretability, and accessibility in occupational health examination settings, the final model incorporating sex, waist circumference, triglycerides, and HDL-C was selected as the routine metabolic screening model for hyperuricemia.

4 Discussion

In this study, we developed and internally validated a routine metabolic screening model for hyperuricemia based on occupational health examination and questionnaire data from gas station workers in a region of Guangxi, China. Among 902 participants, 232 had hyperuricemia, yielding a prevalence of 25.7%, which is higher than the prevalence reported in recent general adult populations in China, suggesting a potentially elevated burden of uric acid metabolic abnormalities in this occupational group. Using LASSO regression and multivariable logistic regression, the final model incorporated four routine health examination indicators: sex, waist circumference, triglycerides, and HDL-C. The model demonstrated good discriminative ability in both the training set (AUC = 0.773) and the validation set (AUC = 0.815). Bootstrap calibration results indicated stable calibration performance, and decision curve analysis suggested potential net benefit within certain threshold probability ranges. Furthermore, after stratifying by predicted probability quartiles, hyperuricemia prevalence in the validation set increased from 5.9% in the lowest quartile to 58.8% in the highest quartile, showing a clear upward trend. These findings indicate that the model not only discriminates current hyperuricemia status but also possesses good risk stratification ability.

The most important feature of this study is that the final model excluded serum uric acid, creatinine, urea, eGFR, and urinary parameters, and instead relied on routine metabolic indicators including sex, waist circumference, triglycerides, and HDL-C for screening. Serum uric acid itself is the basis for outcome definition, and renal function indicators are closely related to uric acid excretion. Although including these indicators might have improved apparent predictive performance, it could also have caused the model to rely excessively on proximal physiological pathways, thereby weakening its independent explanatory value as an early metabolic screening tool. Therefore, the model developed in this study is better positioned as a "preliminary screening and risk stratification tool" in occupational health settings, rather than as a replacement for serum uric acid testing or diagnostic criteria.

Previous studies have shown that hyperuricemia is closely associated with multiple chronic conditions, including obesity, central obesity, dyslipidemia, hypertension, metabolic syndrome, and chronic kidney disease [1–4, 12–15]. National surveys in China have indicated that the prevalence of hyperuricemia has been increasing in recent years, with higher risk observed in men, and among individuals with overweight, obesity, and metabolic abnormalities [5–6]. The hyperuricemia prevalence of 25.7% observed in our gas station workers is higher than that reported in general adult populations in China, which may be attributable to the occupational nature of the study population, the higher proportion of male participants, clustering of metabolic abnormalities, and occupational lifestyle characteristics.

Gas station workers have distinct occupational characteristics. This population may have long-term exposure to gasoline vapors and volatile organic compounds such as benzene, toluene, ethylbenzene, and xylene, while also experiencing shift work, prolonged standing, and irregular dietary and rest patterns [8–11]. Although we did not directly measure benzene series exposure levels and cannot establish a causal relationship between occupational exposure and hyperuricemia, occupational environmental and lifestyle factors may jointly influence body weight, blood lipids, blood pressure, and metabolic status, thereby indirectly contributing to uric acid metabolic abnormalities. Therefore, conducting hyperuricemia

screening based on occupational health examination indicators in gas station workers has practical significance.

From the perspective of model development, previous prediction studies on hyperuricemia have often included indicators such as BMI, blood lipids, creatinine, eGFR, liver function, or lifestyle variables, among which renal function-related indicators tend to have high predictive contributions. However, for early screening in occupational health examinations, over-reliance on creatinine or eGFR may render the model closer to "renal function-uric acid excretion status identification" rather than "metabolic abnormality screening." In this study, we restricted the primary model to exclude uric acid and renal function-related indicators, yet still achieved good discriminative performance and risk stratification in the validation set, indicating that a screening tool with practical value can be developed using routine metabolic indicators alone.

The variables included in the final model—waist circumference, triglycerides, and HDL-C—all reflect metabolic abnormalities and have strong biological plausibility. Waist circumference is a simple indicator of central obesity and visceral fat accumulation. Central obesity may influence serum uric acid levels through mechanisms including insulin resistance, hyperinsulinemia, abnormal adipokine secretion, and increased renal tubular uric acid reabsorption [4, 12–13]. In this study, waist circumference remained positively associated with hyperuricemia in the multivariable model, suggesting that waist circumference may be an important anthropometric screening indicator for uric acid metabolic abnormalities in occupational health examinations of gas station workers.

Elevated triglycerides and reduced HDL-C are key manifestations of dyslipidemia and important components of metabolic syndrome. Previous studies have shown that metabolic syndrome and its components, including central obesity, hypertriglyceridemia, and low HDL-C, are closely associated with hyperuricemia [12–15]. The association between dyslipidemia and hyperuricemia may be mediated through insulin resistance, enhanced hepatic lipid synthesis, abnormal free fatty acid metabolism, and reduced renal uric acid excretion. In this study, triglycerides were positively associated with hyperuricemia, while HDL-C showed a protective association, consistent with previous findings. These results suggest that simultaneous attention to waist circumference, triglycerides, and HDL-C in occupational health examinations may help identify individuals with a high likelihood of hyperuricemia who have clustering of central obesity and dyslipidemia.

The inclusion of sex in the final model is also biologically reasonable. Hyperuricemia is more common in men, which may be related to differences in sex hormone levels, body fat distribution, drinking behavior, dietary patterns, and renal uric acid handling [5–6]. In this study, male sex was significantly associated with hyperuricemia in univariable analysis, but the association was attenuated after adjusting for waist circumference, triglycerides, and HDL-C, suggesting that the higher odds of hyperuricemia in men may be partially explained by central obesity and dyslipidemia. Although male sex did not reach conventional statistical significance in the multivariable model, it was selected by LASSO into the final model and has clear epidemiological relevance; therefore, its retention as a predictor is justified.

Notably, the final model did not include serum uric acid, renal function indicators, or blood count-derived inflammatory markers. Uric acid is primarily excreted by the kidneys, and indicators such as creatinine and eGFR are naturally correlated with hyperuricemia, but this correlation may also cause the prediction model to rely on proximal renal function information. Inflammatory markers such as SII, NLR, MLR, and SII can reflect systemic inflammatory status, but they have low specificity and are easily influenced by infection, stress, sleep, medication, and short-term physiological fluctuations. If these markers do not provide substantial incremental information when competing with metabolic indicators such as waist circumference, triglycerides, and HDL-C, their exclusion from the final model helps maintain parsimony and interpretability.

In addition to AUC, calibration curves, and DCA, we further evaluated the risk stratification ability of the model using predicted probability quartiles. The results showed that hyperuricemia prevalence in the validation set increased significantly across quartiles, reaching 58.8% in the highest-risk quartile (Q4) compared with the lowest-risk quartile (Q1). This finding indicates that the model can stratify the screening population into risk tiers with distinct probabilities of hyperuricemia, demonstrating good public health utility.

In occupational health management, risk stratification is more practical than AUC alone. For individuals at low predicted risk, health education and regular monitoring may be prioritized; for those at moderate-to-high risk, serum uric acid re-testing, dietary and weight management, drinking control, and lipid interventions could be strengthened; and for those at high risk, further uric acid metabolic evaluation and comprehensive chronic disease management may be warranted. This stratification strategy based on routine health examination indicators can help allocate limited health management resources to those most in need, thereby enhancing the translational value of occupational health examination results.

Moreover, the variables in our model are derived from routine health examination items with low cost, high accessibility, and intuitive interpretability, making the model suitable for dissemination in primary occupational health management and workplace health promotion. Compared with complex machine learning models, logistic regression-based models and nomograms offer greater transparency, facilitating clinicians' understanding of variable contributions and enabling clearer explanation of risk sources to workers. Prediction model reporting guidelines emphasize that model evaluation should comprehensively consider discrimination, calibration, and the practical application context. In this study, we evaluated the model from multiple aspects—including ROC curves, bootstrap calibration, DCA, and predicted probability stratification—which collectively enhanced the interpretability and utility of the model results [17–23].

Several limitations should be acknowledged. First, this was a cross-sectional study, and the model predicts current hyperuricemia status rather than future risk of developing hyperuricemia. Therefore, the "predicted probability" should be interpreted as an estimate of current hyperuricemia status and cannot be construed as incident risk or used to infer causality. Second, serum uric acid, blood lipid, and anthropometric measurements were based on single health examination assessments and may be

influenced by short-term dietary intake, fluid consumption, physical activity, sleep, medication use, and blood sampling timing, introducing measurement error. Third, although we included multiple routine health examination indicators and questionnaire variables, we did not fully capture important factors affecting uric acid metabolism, such as high-purine diet, sugar-sweetened beverage consumption, specific alcohol intake, diuretic or antihypertensive medication use, menopausal status, family history, genetic susceptibility, and uric acid excretion-related parameters. Residual confounding or omitted predictors may remain. Fourth, the study population was drawn from gas station workers of a petroleum sales enterprise in a specific region of Guangxi, with a relatively concentrated sample, and the healthy worker effect may be present; caution is needed when generalizing the model to other regions, industries, or the general population. Fifth, although we performed training/validation set partitioning and bootstrap internal validation, these remain internal validation methods; the lack of independent external validation samples means that some optimistic bias may persist. Sixth, we did not conduct quantitative assessments of occupational exposure among gas station workers, such as environmental benzene series concentrations, urinary metabolites, shift intensity, site-specific exposure differences, or cumulative exposure doses; therefore, we could not further evaluate the incremental value of occupational exposure indicators for the hyperuricemia screening model. Seventh, LASSO selection results are influenced by sample structure, candidate variable sets, variable coding, and data preprocessing methods; the final selected variables should be understood as predictor combinations optimal under the conditions of this dataset rather than as stable causal factors.

Future studies should conduct external validation in larger, multi-region, and diverse occupational populations, and incorporate prospective follow-up data to further evaluate the model's ability to predict new-onset hyperuricemia. Additionally, integrating dietary, medication, genetic factors, uric acid excretion indicators, and quantitative occupational exposure data would enable comparison of different models in terms of discrimination, calibration, net benefit, and generalizability, thereby further optimizing screening strategies for hyperuricemia in occupational populations.

Conclusions

In this study, we developed and internally validated a routine metabolic screening model for hyperuricemia that does not rely on serum uric acid or renal function-related indicators, based on occupational health examination data from gas station workers in a region of Guangxi, China. The final model incorporated four routine health examination indicators—sex, waist circumference, triglycerides, and HDL-C—and demonstrated good discrimination, calibration, and clinical net benefit in both the training and validation sets. Predicted probability stratification showed that hyperuricemia prevalence increased markedly with higher predicted risk, indicating good risk stratification value. The model may serve as an auxiliary tool for preliminary screening and risk stratification of hyperuricemia among gas station workers in occupational health settings. Future multi-center external validation and prospective cohort studies are needed to further evaluate its stability, generalizability, and practical utility.

Declarations

Ethics approval and consent to participate

Ethics approval was obtained from the Ethics Committee of Youjiang Medical University for Nationalities, Baise, Guangxi, China. The ethics approval number will be provided upon request. All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments. All participants provided written informed consent before taking part in the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Authors' contributions

JW: Conceptualization, Methodology, Writing – Original Draft; XH: Conceptualization, Methodology, Writing – Original Draft; YL: Investigation, Data Curation, Formal Analysis; ZC: Methodology, Data Curation; JC: Investigation, Data Curation; YP: Conceptualization, Supervision, Project administration, Funding Acquisition, Writing – Review & Editing; GQ: Conceptualization, Supervision, Project administration, Writing – Review & Editing. All authors read and approved the final manuscript.

Acknowledgements

We thank all study participants and the staff of the Guangxi Nanning Petroleum Branch for their support and cooperation.

Authors' information

Not applicable.

References

1. Chinese Society of Endocrinology, Chinese Medical Association. Chinese guideline for the diagnosis and treatment of hyperuricemia and gout (2019)[J]. *Chin J Endocrinol Metabolism*. 2020;36(1):1–13.
2. Multidisciplinary Expert Consensus Task Force on Hyperuricemia and Related Diseases, Severe Metabolic Diseases Branch, China National Health Association. Chinese multidisciplinary expert consensus on the diagnosis and treatment of hyperuricemia and related diseases (2023 edition)[J]. *Chin J Practical Intern Med*. 2023;43(6):461–80.
3. Du L, Zong Y, Li H, et al. Hyperuricemia and its related diseases: mechanisms and advances in therapy[J]. *Signal Transduct Target Therapy*. 2024;9:212.
4. Gong M, Wen S, Ngan I, et al. Converging relationships of obesity and hyperuricemia with special reference to metabolic disorders and plausible therapeutic implications[J]. *Diabetes Metabolic Syndrome Obes*. 2020;13:943–62.
5. Zhang M, Zhu X, Wu J, et al. Prevalence of hyperuricemia among Chinese adults: findings from two nationally representative cross-sectional surveys in 2015-16 and 2018-19[J]. *Front Immunol*. 2022;12:791983.
6. Song J, Jin C, Shan Z, et al. Prevalence and risk factors of hyperuricemia and gout in China: a national cross-sectional survey[J]. *Clin Rheumatol*. 2022;41(4):1209–18.
7. She D, Wang Y, Liu J, et al. Changes in the prevalence of hyperuricemia in clients of health examination in Eastern China, 2009 to 2019[J]. *BMC Endocr Disorders*. 2022;22:202.
8. Chen Y, Yang Y, Zheng Z, et al. Influence of occupational exposure on hyperuricemia in steelworkers: a nested case-control study[J]. *BMC Public Health*. 2022;22:1508.
9. Li Q, Wang J, Zhang Y, et al. Association between occupational heat exposure and hyperuricemia among petrochemical workers[J]. *Front Public Health*. 2025;13:1648619.
10. Kuranchie FA, Angove MJ, Rahman MM. Occupational exposure of benzene, toluene, ethylbenzene and xylene among gasoline station workers: a review[J]. *Environ Reviews*. 2019;27(3):309–20.
11. Tunsaringkarn T, Siriwong W, Ritthidej G, et al. Occupational exposure of gasoline station workers to BTEX compounds in Bangkok, Thailand[J]. *Int J Occup Environ Med*. 2012;3(3):117–25.
12. Tu YC, Chen KW, Chen TY, et al. Metabolic syndrome and its components are associated with new-onset hyperuricemia: a longitudinal study[J]. *Nutrients*. 2023;15(6):1337.
13. Cui W, Liu C, Chen C, et al. Triglycerides and HDL cholesterol mediate the association between waist circumference and hyperuricemia in normal-weight men[J]. *Front Nutr*. 2024;11:1481517.
14. Peng TC, Wang CC, Kao TW et al. Relationship between hyperuricemia and lipid profiles in US adults[J]. *BioMed Research International*, 2015, 2015: 127596.
15. Chen L, Zhu WH, Chen ZW, et al. Relationship between hyperuricemia and metabolic syndrome[J]. *J Zhejiang Univ Sci B*. 2007;8(8):593–8.
16. Zheng Z, Yang Y, Zhang L, et al. Risk prediction for the development of hyperuricemia: model development using an occupational health examination dataset[J]. *Int J Environ Res Public Health*.

2023;20(4):3411.

17. Collins GS, Dhiman P, Ma J, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J]. *BMJ*. 2024;385:e078378.
18. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement[J]. *Ann Intern Med*. 2015;162(1):55–63.
19. Tibshirani R. Regression shrinkage and selection via the lasso[J]. *J Roy Stat Soc B*. 1996;58(1):267–88.
20. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*[M]. 2nd ed. Cham, Switzerland: Springer; 2015.
21. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*[M]. 2nd ed. Cham, Switzerland: Springer; 2019.
22. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis[J]. *Diagn Prognostic Res*. 2019;3:18.
23. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models[J]. *Med Decis Making*. 2006;26(6):565–74.

Figures

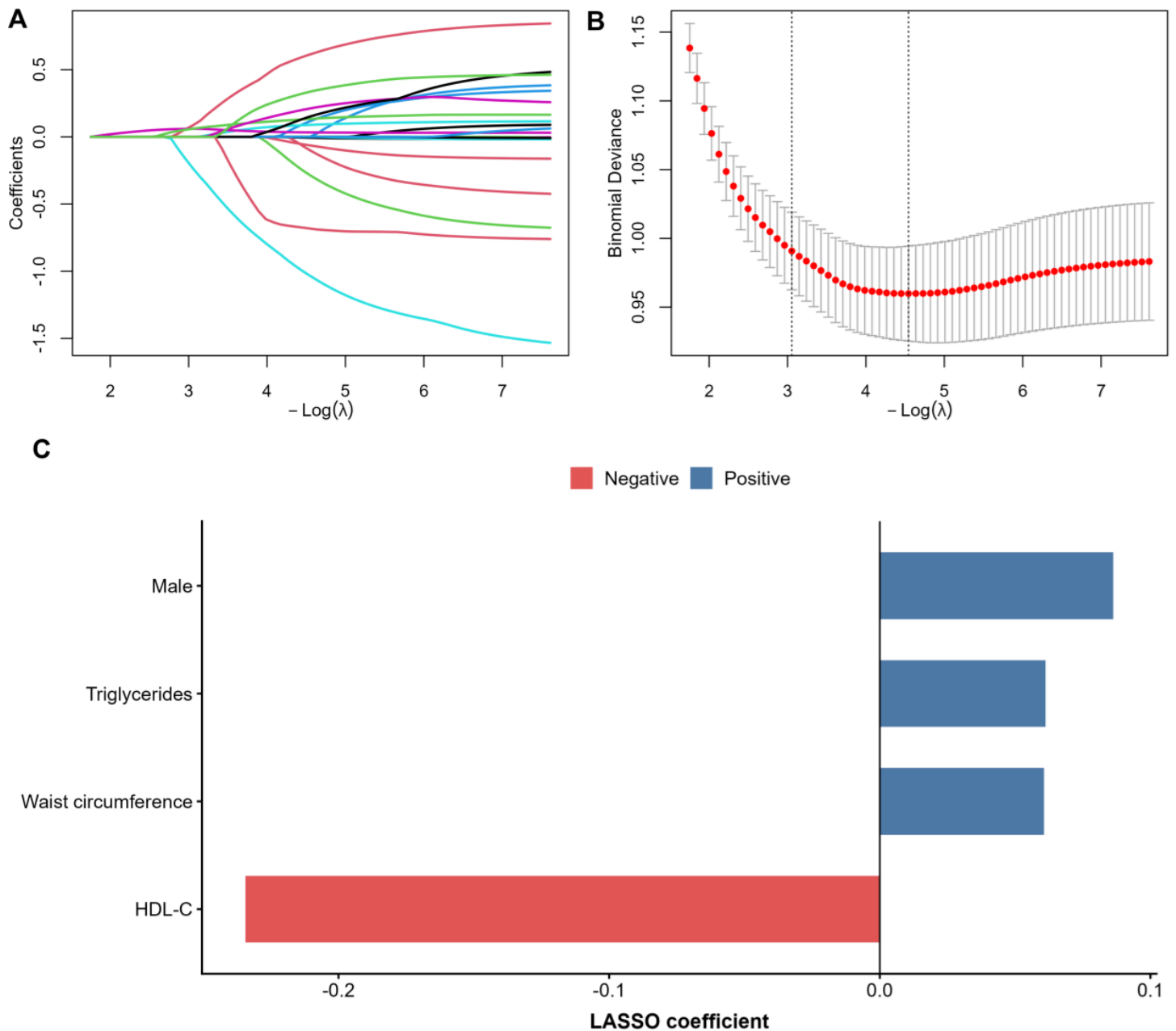


Figure 1

LASSO regression for selection of hyperuricemia predictors

Note: (A) LASSO coefficient paths; (B) 10-fold cross-validation for determining the optimal penalty parameter λ ; (C) non-zero coefficient variables selected under the λ_{1se} criterion. Dashed vertical lines indicate $\lambda_{\min}$ and λ_{1se} , respectively. BMI, body mass index; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol.

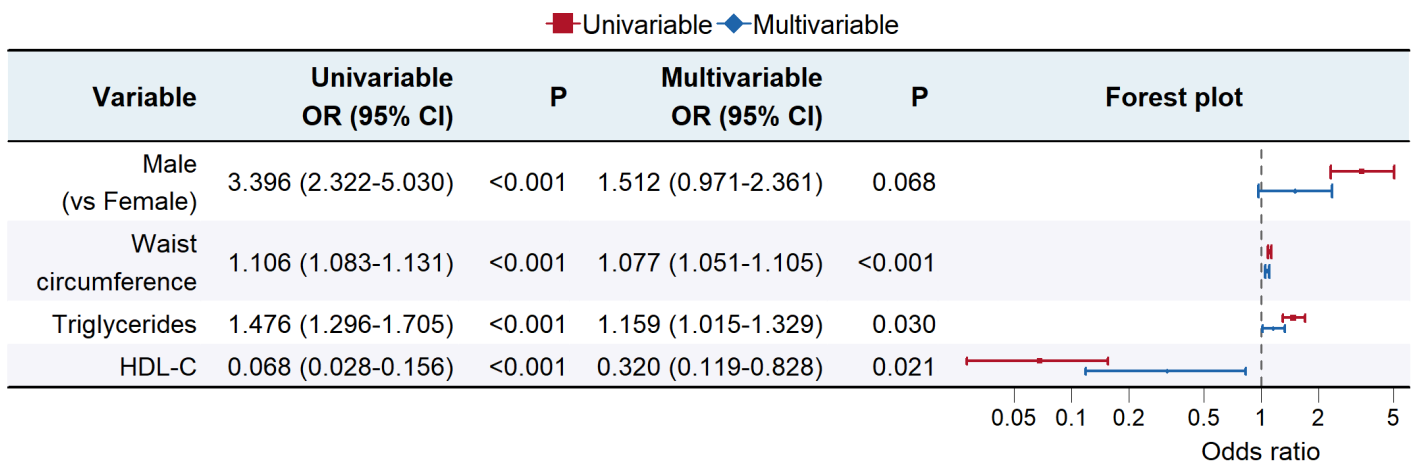


Figure 2

Forest plot of univariable and multivariable logistic regression for LASSO-selected variables associated with hyperuricemia

Note: Points represent odds ratios, horizontal lines represent 95% confidence intervals, and the vertical dashed line indicates OR = 1. Continuous variables were analyzed per 1-unit increase: waist circumference in cm, triglycerides in mmol/L, and HDL-C in mmol/L. OR, odds ratio; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol.

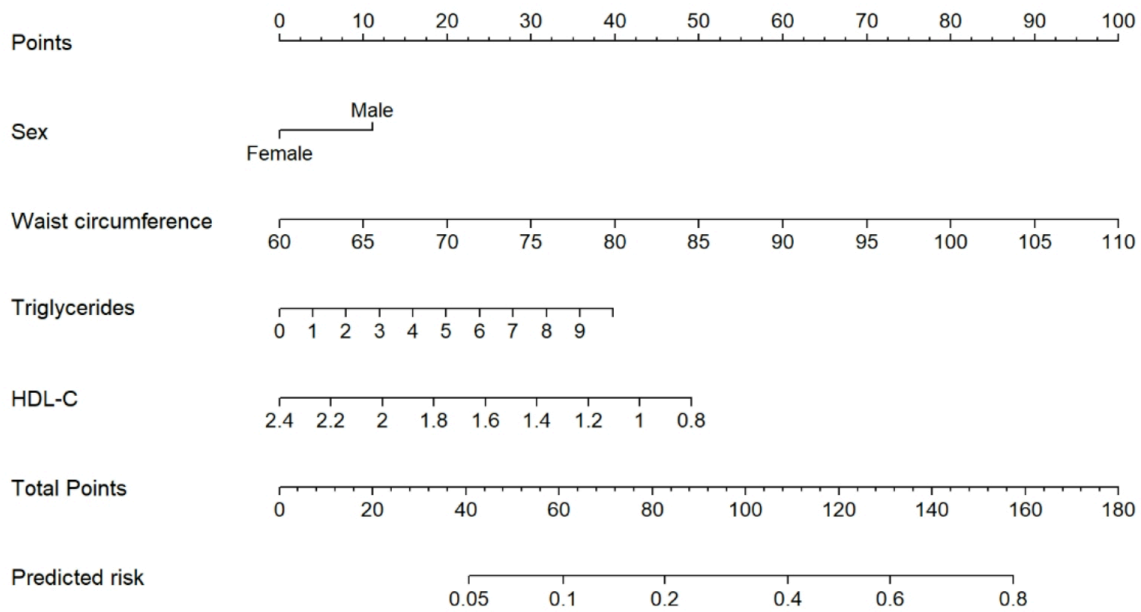


Figure 3

Nomogram for predicting hyperuricemia risk based on routine health examination indicators

Note: The nomogram was constructed based on the multivariable logistic regression model and includes four predictors: sex, waist circumference, triglycerides, and HDL-C. Points for each variable are summed to obtain a total score, which is then converted to the predicted probability of hyperuricemia. HDL-C, high-density lipoprotein cholesterol.

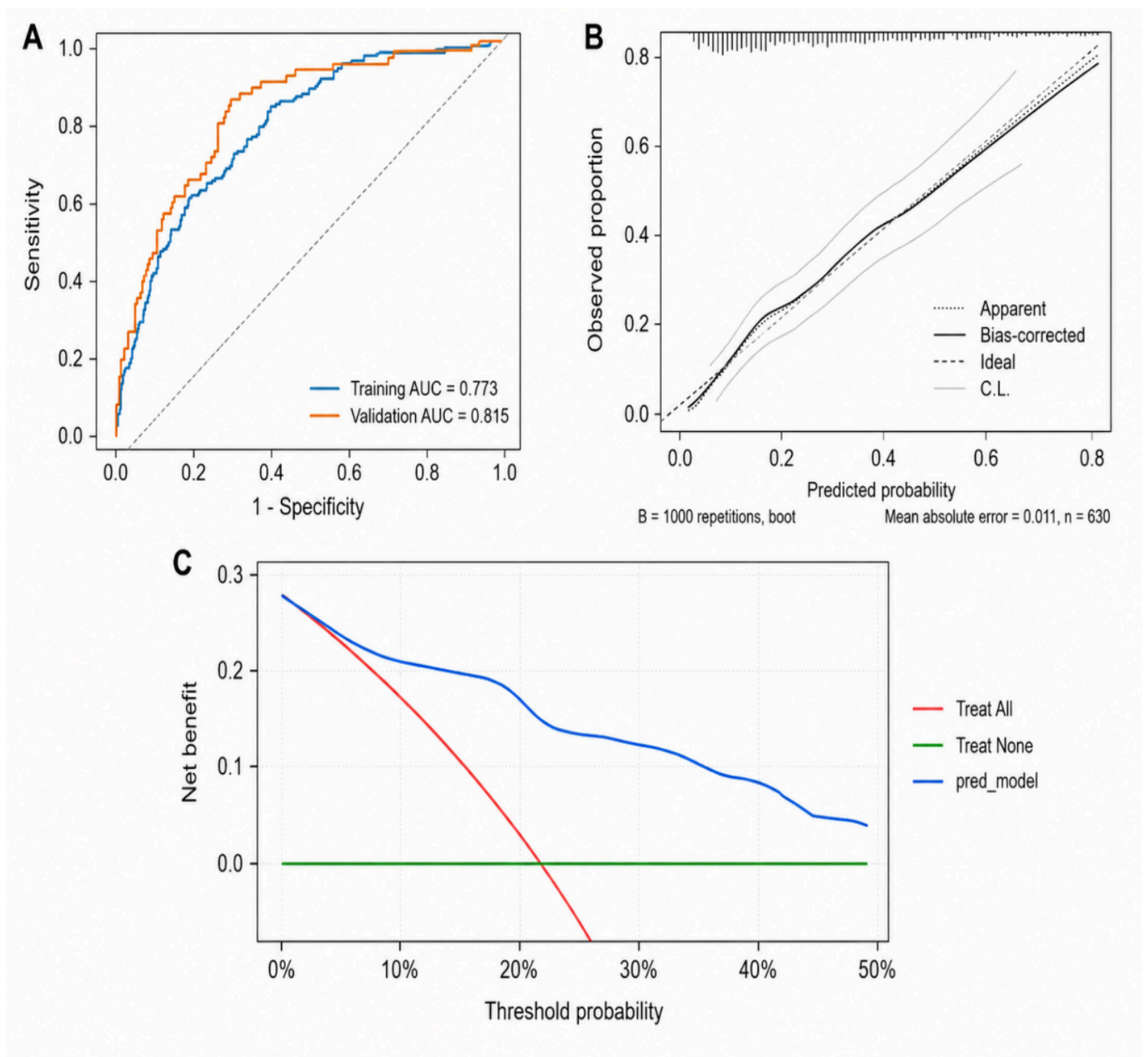


Figure 4

Discrimination, calibration, and clinical net benefit evaluation of the hyperuricemia prediction model

Note: (A) ROC curves for the training and validation sets; (B) Bootstrap calibration curves with 1,000 resampling iterations; (C) Decision curve analysis. ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis.

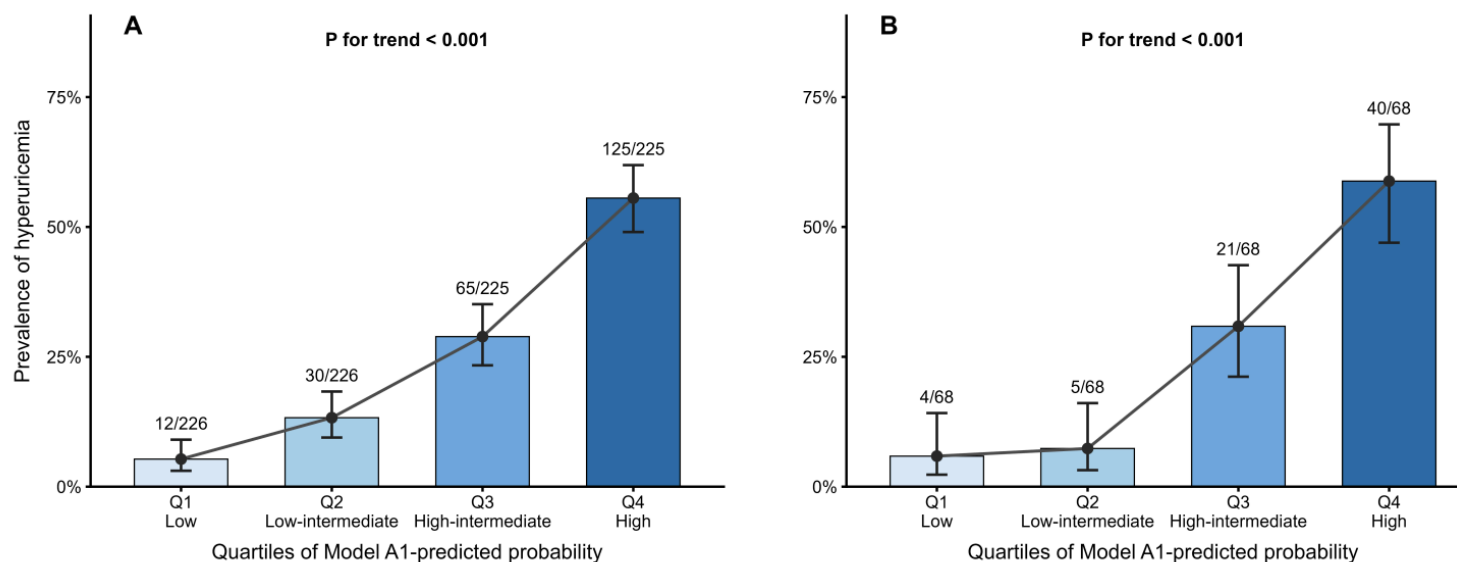


Figure 5

Predicted probability stratification and hyperuricemia prevalence trends

Note: (A) Total cohort; (B) Validation set. Participants were divided into quartile groups based on predicted probabilities from the final model: Q1, lowest predicted risk; Q4, highest predicted risk. Bars represent hyperuricemia prevalence in each group, error bars indicate Wilson 95% confidence intervals, and numbers above bars indicate hyperuricemia cases/total participants in each group. HUA, hyperuricemia.

Supplementary Files

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

- [SupplementaryMaterials.docx](#)
- [Additionalfile1QuestionnaireEnglish.pdf.pdf](#)