Observational Study

Medicine®

OPEN

Establishment and verification of a nomogram prediction model of hypertension risk in Xinjiang Kazakhs

Yuezhen Xu, MM[a,b], Jinbao Liu, MM[a], Jiawei Wang, MM[b], Qiongling Fan, BS[b], Yuanyuan Luo, BS[b], Huaifeng Zhan, MB[c], Ning Tao, PhD[a], Shuping You, PhD[b,*]

Abstract

Hypertension is the main risk factor for cardiovascular and renal diseases. It is of great importance to develop effective risk prediction models to identify high-risk groups of hypertension. This study is to establish and verify a nomogram model for predicting the risk of hypertension among Kazakh herders in Xinjiang, China.

This is a prospective cohort study. Totally, 5327 Kazakh herders from the Nanshan pastoral area of Xinjiang were enrolled. They were randomly divided into the modeling set of 3729 cases (70%) and the validation set of 1598 cases (30%). In the modeling set, univariate analysis, least absolute shrinkage and selection operator regression and multivariate Logistic regression were used to analyze the influencing factors of hypertension, and a nomogram prediction model was constructed. We then validated the model in the validation set, and evaluated the accuracy of the model using receiver operating characteristic (ROC) and calibration curve.

Based on univariate analysis, least absolute shrinkage and selection operator regression and multivariate logistic regression analysis, we identified 14 independent predictors of hypertension in the modeling set, including age, smoking, alcohol consumption, baseline body mass index, baseline diastolic blood pressure, baseline systolic blood pressure, daily salt intake, yak-butter intake, daily oil intake, fruit and vegetable intake, low-density lipoprotein, cholesterol, abdominal circumference, and family history. The area under the receiver operating characteristic curve of the modeling set and the verification set was 0.803 and 0.809, respectively. Moreover, the calibration curve showed a higher agreement between the nomogram prediction and the actual observation of hypertension.

The risk prediction nomogram model has good predictive ability and could be used as an effective tool for the risk prediction of hypertension among Kazakh herders in Xinjiang.

Abbreviations: AUC = area under the ROC curve, BMI = body mass index, DBP = diastolic blood pressure, HDL = high-density lipoprotein, LASSO = least absolute shrinkage and selection operator, LDL = low-density lipoprotein, ROC = receiver operating characteristic, TC = total cholesterol, TG = triglyceride.

Keywords: hypertension, Kazakh, nomogram, risk prediction

Editor: Flavio Palmieri.

This work was supported by 2017 General Project of Humanities and Social Science Research of the Ministry of Education (No. 17YJCZH230).

Written informed consent was obtained from every patient and the study was approved by the Ethics Committee of the First Clinical Affiliated Hospital of Xinjiang Medical University (Approval No: 20140304-150).

The authors have no conflicts of interest to disclose.

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

[a] School of Public Health, Xinjiang Medical University, Urumqi, China, [b] Teaching and Research Department of Basic Nursing, School of Nursing, Xinjiang Medical University, Urumqi, China, [c] Shuixigou Health Center of Urumqi County, Urumqi, China.

[*] Correspondence: Shuping You, Teaching and Research Department of Basic Nursing, School of Nursing, Xinjiang Medical University, No. 393 Xinyi Road, Urumqi 830011 Xinjiang, China (e-mail: shupingyou@yeah.net).

How to cite this article: Xu Y, Liu J, Wang J, Fan Q, Luo Y, Zhan H, Tao N, You S. Establishment and verification of a nomogram prediction model of hypertension risk in Xinjiang Kazakhs. Medicine 2021;100:42(e27600).

Received: 4 March 2021 / Received in final form: 31 August 2021 / Accepted: 6 October 2021

http://dx.doi.org/10.1097/MD.0000000000027600

1. Introduction

Hypertension is the main risk factor for cardiovascular and renal diseases.[1–3] The common risk factors for hypertension include age, smoking, alcohol consumption, and family history. Risk prediction models can effectively identify high-risk groups of hypertension. Studies have shown that the more risk factors included in the intervention plan, the lower the average risk.[4–6] However, the predictive value of different models is variable, leading to certain limitations on their clinical application.[7,8] For example, the Framingham Heart Study in the United States developed a hypertension risk prediction model, which can be used for short-term follow-up and for assessment of the absolute risk of individual hypertension.[7] However, since 99% of Framingham Research Center participants are of European descent, the Framingham prediction function cannot be extrapolated to other populations without external verification.[7] Zheng et al[9] found that the Framingham model had low predictive value in the rural population of Northeast China. Therefore, prediction models in different regions need to be constructed in conjunction with the risk factors of hypertension in the local population.

At present, the commonly used analysis methods for risk prediction include Logistic regression analysis, Cox regression, artificial neural network, multilevel model, linear mixed model, Joint model analysis method, nomogram, and support vector machine model.[10,11] Among these models, the nomogram model allows for personalized and evidence-based risk assessment, and has gradually become a reliable and convenient tool for quantifying risks.[12] Kim et al[13] constructed a hypertension prediction model based on the Korean population, in which the included variables were age, body mass index (BMI), stroke, and family history. In addition, Shen et al[14] constructed a nomogram model of the risk of carotid atherosclerosis in patients with hypertension. However, currently, there is no report on the construction of a nomogram model of the incidence of hypertension in China.

Kazakhs in the pastoral areas of Xinjiang are among the 5 ethnic groups with the highest prevalence of hypertension.[15] They have special living environment and unique diet and lifestyle. It is of great importance to develop effective methods to assess the incidence of hypertension in Kazakh herders and to construct a predictive model suitable for Kazakh herders. Here, in this study, we constructed a nomogram model for hypertension risk prediction suitable for Kazakhs in Nanshan pastoral area, Xinjiang.

2. Materials and methods

2.1. Subjects

This is a prospective cohort study. We recruited Kazakh herders who received physical examination at 6 health centers/health service centers in 3 townships under the jurisdiction of Urumqi County, Xinjiang from January 2008 to December 2018, and established a dynamic cohort. We adopted stratified cluster sampling method and included the research objects according to the population of each township in the population of the entire Urumqi County. The inclusion criteria were: age ≥18 years; residence time in Xinjiang Nanshan Pastoral Area ≥3 years; complete follow-up data; Kazakh herdsmen. The exclusion criteria were: refusal to participate in the follow-up; patients with hypertension, stroke, coronary heart disease, and diabetes; patients with severe diseases such as liver cirrhosis, cancer, nephrotic syndrome, renal failure, respiratory failure, chronic obstructive pulmonary emphysema, idiopathic thrombocytopenic purpura, aplastic anemia, etc; patients with secondary hypertension. Exclusion of secondary hypertension was based on the guidelines for prevention and treatment of hypertension in China. In detail, patients with history of nephritis or anemia that suggests renal parenchymal hypertension; myasthenia and other hypokalemia signs that suggest primary aldosteronism; paroxysmal headaches, palpitations, and sweating that suggest pheochromocytoma; snoring accompanied by apnea; and, long-term use of drugs that increases blood pressure, were excluded. Written informed consent was obtained from every patient and the study was approved by the Ethics Committee of the First Clinical Affiliated Hospital of Xinjiang Medical University (Approval No: 20140304-150).

2.2. Baseline data

The baseline data was obtained from the medical records of physical examination and questionnaires. The physical examination included basic information such as gender, age, and education level. The self-designed "Questionnaire for Hypertension and Risk Factors of Kazakhs in Xinjiang Pastoral Areas" was used to conduct a baseline survey (Cronbach α 0.882). The items of questionnaire included smoking, alcohol consumption, intake of vegetable, salt, milk, and tea, and family history of hypertension, etc. The baseline survey was completed from January 2008 to December 2010. A total of 12,786 subjects participated in the baseline survey and the qualified questionnaires were obtained from 11,252 subjects (the response rate was 88%).

2.3. Follow-up

Among the subjects of the baseline survey, 4950 people with hypertension, coronary heart disease, diabetes, and stroke were excluded, and 6302 who met the inclusion criteria were followed up every 2 years. The follow-up outcome was hypertension. The follow-up lasted for 1.1 to 8.5 years, with a median follow-up time of 4.5 years. In the end, 5813 cases completed follow-up and 489 were lost to follow-up. During the follow-up period, there were 1985 new cases of hypertension. We excluded 213 subjects with incomplete blood pressure measurements and 273 subjects with incomplete outcome during follow-up. Finally, 5327 eligible subjects were included in this study.

2.4. Physical examination and laboratory examination

The physical examination included height, weight, blood pressure, and BMI. The blood pressure was measured by using a uniformly calibrated mercury sphygmomanometer. Each subject was measured at least twice, and the average value in mm Hg was used for data analysis. Besides, DP180 automatic biochemical analyzer (Dongtang Electronic Technology Co., Ltd., Guangdong, China) was used to detect the blood biochemical indicators of high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglyceride (TG), and total cholesterol (TC), etc.

2.5. Definitions

The diagnosis of hypertension was based on the "Guidelines for the Prevention and Treatment of Hypertension in China (2018)".[16] Hypertension was diagnosed when there was systolic blood pressure (SBP) ≥140 mm Hg and/or diastolic blood pressure (DBP) ≥90 mm Hg in the absence of antihypertensive drugs, or when patients had a history of high blood pressure and were currently taking antihypertensive drugs. The new-onset hypertension was diagnosed when the average SBP≥140 mm Hg and/or average DBP ≥90 mm Hg of 3 times, and/or when patients received antihypertensive medication within the past 2 weeks. Smoking history was defined as inhalation of smoke produced by cigarette burning for at least 1 day (>15 min/d) per week for more than 3 years or cumulative smoking of >100 cigarettes.[17] Alcohol consumption was defined as ≥50 ml/time, liquor (1 bottle of beer ≈100 ml liquor), on average ≥1 time/wk per week for 1 year.[18] Individual daily salt consumption (g) was calculated as total annual salt consumption of the family (g)/number of family members/365 (d). Dyslipidemia referred to TC, TG, and LDL levels higher than normal standards or HDL lower than normal standards under fasting state.[19] The normal values of blood lipids were: TC < 5.18 mmol/l; TG < 1.70 mmol/l; LDL < 3.37 mmol/l; HDL≥1.04 mmol/l.

2.6. Statistical analysis

All data were analyzed by using R software 3.6.0 (http://www.R-project.org, the R Foundation, Bell Laboratories, Holmdel, NJ, USA). The statistical analysis of this study was carried out strictly in accordance with the TRIPOD statement of the prediction model. Measurement data was expressed as M (P25, P75). Count data was presented as percentage or rate. The population was divided into a modeling set (3729 cases) and a verification set (1598 cases) using random split sampling (split ratio: 7:3). In the modeling set, a univariate analysis, and the least absolute shrinkage and selection operator (LASSO) regression method was used to screen the possible risk factors of hypertension. Furthermore, the variables screened by LASSO regression were subjected to multivariate Logistic regression analysis. The nomogram prediction model was established based on the results of the logistic regression analysis, and the nomogram model was verified in the verification set. The calibration curve was used to evaluate the calibration degree of the nomogram model. receiver operating characteristic (ROC) curve was plotted to analyze the predictive ability of the nomogram model for the risk of hypertension. The area under the ROC curve (AUC) was used to evaluate the discrimination degree of the model. A P value $<.05$ was considered statistically significant.

3. Results

3.1. Characteristics of included subjects

A total of 5327 subjects were finally included in the analysis, including 2499 (46.91%) males and 2828 (53.09%) females. The incidence of hypertension was 11.75/100 person-years (95% CI: 11.27-12.24). They were divided into the modeling set (3729 cases) and the validation set (1598 cases) using random split sampling (split ratio: 7:3). The 2 groups had no statistically significant differences in basic characteristics, such as gender, age, smoking, BMI, blood pressure, salt intake, and yak-butter intake (Table 1, $P>.05$).

3.2. Analysis of independent risk factors in the modeling set

In the univariate and multivariate logistic regression analysis, we included 22 variables, including gender, age, smoking, alcohol consumption, BMI, baseline SBP, baseline DBP, daily salt intake, yak-butter intake, daily oil intake, milk tea intake, exercise, fruit and vegetable intake, high-density lipoprotein, low-density lipoprotein, cholesterol, triglycerides, abdominal circumference, neck circumference, central obesity, prehypertension, and family history (Table 2). Using hypertension as the dependent variable and related factors as the independent variables, univariate logistic regression analysis showed that the included 22 variables were all related to hypertension. After that, we further included the significant variables of the univariate logistic regression analysis into the LASSO regression. After LASSO regression analysis (Fig. 1) and multivariate logistic regression analysis, we finally identified 14 independent predictors of hypertension in the modeling set, including age, smoking, alcohol consumption, baseline BMI classification, baseline DBP, baseline SBP, daily salt intake, yak-butter intake, daily oil intake, fruit and vegetable intake, low-density lipoprotein, cholesterol, abdominal circumference, and family history.

3.3. Construction of nomogram model

Based on the results of multivariate logistic regression analysis, we used hypertension as the outcome, and described the impact of each variable on the risk of hypertension in the form of a nomogram. Finally, we established a nomogram prediction model related to the occurrence of hypertension (Fig. 2).

3.4. Verification and calibration of the nomogram model for the risk prediction of hypertension among Kazakh herders

In order to verify the predictive ability of the nomogram model for the risk of hypertension, ROC analysis was performed. Figure 3 shows that AUC of the modeling set and validation set was 0.803 and 0.809, respectively. In addition, respectively, the optimal critical value in the ROC curve was 0.309 (0.628, 0.833) in the modeling set and 0.359 (0.676, 0.812) in the validation set.

The calibration curve was used to assess the calibration degree of the nomogram model in the modeling set and validation set. The results of the modeling set (Fig. 4A) and the validation set (Fig. 4B) showed that the actual prediction and the simulation prediction were basically the same, indicating a good agreement between the nomogram prediction and the actual observation result of hypertension. These results suggest that the model has good predictive performance.

4. Discussion

The results of the Framingham Research Center show that everyone will experience the risk of high blood pressure to varying degrees in their lifetime.[2] The conceal pathogenesis of hypertension determines the importance of hypertension prevention. The Framingham model is a very practical tool for predicting the risk of hypertension, and has achieved remarkable results in the prevention of cardiovascular and cerebrovascular diseases.[20] Usually, the risk factors of hypertension are clustered and also accompanied by synergistic effects. Therefore, appropriate strategies to stratify and predict the risk of hypertensive populations can maximize the benefits of hypertension prevention and control. The American Heart Association recommended the Framingham model as a first-level prediction tool for hypertension risk.[21–25] The prevalence of essential hypertension in the Kazakhs ranks among the top 3 of all ethnic groups in China.[26] Our previous results showed that the Kazakhs in pastoral areas in Xinjiang had high prevalence of blood pressure (as high as 48.88%).[27] However, it has been verified that the Framingham model has a poor fit among Kazakh herders in the Nanshan pastoral area of Xinjiang.[27] Thus, in order to better predict the risk of hypertension among Kazakh herders, we constructed a more suitable hypertension risk prediction model for Kazakhs in Xinjiang. The nomogram[28] is a convenient graph to show the mathematical model, which combines various important factors to predict a specific end event. The nomogram model can graphically visualize logistic regression results, and can be intuitively used to predict individual disease risks. By constructing a nomogram model for predicting the risk of hypertension among the Kazakh herders in the pastoral areas of Xinjiang, the screening of high-risk groups of hypertension among the Kazakh herders in the pastoral areas of Xinjiang can be improved, health resources can be optimized, work efficiency can be improved, and the incidence of diseases can be reduced.

Kazakhs in Xinjiang have unique regional characteristics and special diet and living habits, which cause differences in the risk factors of hypertension. Our previous results[11] found that among Kazakh herders in Xinjiang, the factors of age, smoking,

Table 1

Basic demographic characteristics of the study cohort.

Items	Total (n=5327)	Modeling set (n1=3729)	Validation set (n2=1598)	*P* value
Gender, no. (%)				.7575
Male	2499 (46.91)	1755 (47.06)	744 (46.56)	
Female	2828 (53.09)	1974 (52.94)	854 (53.44)	
Age, median (yr	50 (41, 60.5)	50 (41, 61)	50 (41, 60)	.6825
Smoking, no. (%)				.3755
No	4303 (80.78)	3000 (80.45)	1303 (81.54)	
Yes	1024 (19.22)	729 (19.55)	295 (18.46)	
Drinking, no. (%)				1
No	4307 (80.85)	3015 (80.85)	1292 (80.85)	
Yes	1020 (19.15)	714 (19.15)	306 (19.15)	
BMI, no. (%)				.952
<24.0 kg/m^2	2240 (42.05)	1563 (41.91)	677 (42.37)	
24-28 kg/m^2	2196 (41.22)	1540 (41.3)	656 (41.05)	
≥28 kg/m^2	891 (16.73)	626 (16.79)	265 (16.58)	
Baseline diastolic blood pressure, median (mm Hg)	79 (70, 83)	80 (70, 83)	78 (70, 83)	.2437
Baseline systolic blood pressure, median (mm Hg	128 (117, 134)	128 (117, 134)	128 (118, 134)	.6369
Daily salt intake exceeds 16.66 g/d, no. (%)				.1736
No	2414 (45.32)	1713 (45.94)	701 (43.87)	
Yes	2913 (54.68)	2016 (54.06)	897 (56.13)	
Yak-butter intake, no. (%)				.7064
Everyday	743 (13.95)	507 (13.6)	236 (14.77)	
4-6 times/wk	881 (16.54)	625 (16.76)	256 (16.02)	
1-3 times/wk	2142 (40.21)	1512 (40.55)	630 (39.42)	
Several times/mo	946 (17.76)	661 (17.73)	285 (17.83)	
None	615 (11.54)	424 (11.37)	191 (11.95)	
Daily edible oil intake, no. (%)				.547
20-25 g	910 (17.08)	636 (17.06)	274 (17.15)	
25-30 g	1250 (23.47)	867 (23.25)	383 (23.97)	
30-35 g	1290 (24.22)	924 (24.78)	366 (22.9)	
35-40 g	569 (10.68)	403 (10.81)	166 (10.39)	
Unknown	1308 (24.55)	899 (24.11)	409 (25.59)	
Exercise, no. (%)				.6751
No	4994 (93.75)	3492 (93.64)	1502 (93.99)	
Yes	333 (6.25)	237 (6.36)	96 (6.01)	
Fruits and vegetables intake, no. (%)				.5075
No	2855 (53.59)	1987 (53.29)	868 (54.32)	
Yes	2472 (46.41)	1742 (46.71)	730 (45.68)	
Excessive milk tea, no. (%)				.1609
No	2112 (39.65)	1455 (39.02)	657 (41.11)	
Yes	3215 (60.35)	2274 (60.98)	941 (58.89)	
Low-density lipoprotein cholesterol, median (IQR)	2.83 (2.09, 3.47)	2.72 (2.09, 3.47)	2.93 (2.09, 3.47)	.2844
High-density lipoprotein cholesterol, median (IQR)	1.32 (1.02, 1.5)	1.34 (1.02, 1.52)	1.31 (0.98, 1.48)	.0484
Triglycerides, median (IQR)	1.26 (0.94, 1.64)	1.26 (0.94, 1.64)	1.26 (0.94, 1.64)	.7483
Total cholesterol, median (IQR)	4.87 (4.25, 5.57)	4.87 (4.25, 5.56)	4.87 (4.27, 5.62)	.2433
Abdominal circumference, median (IQR)	90 (84, 99)	90 (84, 99)	90 (83, 99)	.1147
Neck circumference, median (IQR)	35 (25, 40)	35 (25, 40)	35 (25, 40)	.4826
Central obesity, no. (%)				.0422
No	2356 (44.23)	1615 (43.31)	741 (46.37)	
Yes	2971 (55.77)	2114 (56.69)	857 (53.63)	
Prehypertension, no. (%)				.58
No	1590 (29.85)	1122 (30.09)	468 (29.29)	
Yes	3737 (70.15)	2607 (69.91)	1130 (70.71)	
Family history, no. (%)	0.4581			
Father	3655 (68.61)	2544 (68.22)	1111 (69.52)	
Mother	1363 (25.59)	972 (26.07)	391 (24.47)	
Both father and mother	309 (5.8)	213 (5.71)	96 (6.01)	
Hypertension, no. (%)				.1949
0	3342 (62.74)	2318 (62.16)	1024 (64.08)	
1	1985 (37.26)	1411 (37.84)	574 (35.92)	

BMI=body mass index, IQR=inter-quartile range.

Table 2

Univariate and multivariate logistic regression analysis.

	Univariate analysis		Multivariate analysis	
Characteristics	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Gender, no. (%)				
Male	1		–	
Female	0.80 (0.70-0.92)	.001	–	–
Age, median (IQR)	1.04 (1.04, 1.09)	<.0001	1.03 (1.02-1.03)	<.0001
Smoking, no. (%)				
No	1		1	
Yes	1.59 (1.35, 1.87)	<.0001	1.52 (1.25-1.86)	<.0001
Drinking, no. (%)				
No	1		1	
Yes	1.40 (1.19, 1.65)	<.0001	1.23 (1.01-1.51)	.00399
BMI, no. (%)				
<24.0 kg/m^2	1		1	1
24-28 kg/m^2	1.66 (1.42, 1.92)	<.0001	1.39 (1.16-1.66)	.0003
≥28 kg/m^2	3.51 (2.90, 4.26)	<.0001	2.13 (1.68-2.71)	<.0001
Baseline diastolic blood pressure, median (IQR)	1.11 (1.10-1.12)	<.0001	1.07 (1.05-1.08)	<.0001
Baseline systolic blood pressure, median (IQR)	1.08 (1.07-1.09)	<.0001	1.04 (1.03-1.05)	<.0001
Daily salt intake exceeds 16.66 g/d, no. (%)				
No	1		1	
Yes	1.46 (1.28-1.67)	<.0001	1.46 (1.25-1.71)	<.0001
Yak-butter intake, no. (%)				
Everyday	1		1	
4-6 times/wk	0.71 (0.56-0.90)	.005	0.66 (0.5-0.88)	.0046
1-3 times/wk	0.55 (0.45-0.68)	<.0001	0.58 (0.46-0.74)	<.0001
Several times/mo	0.45 (0.35-0.57)	<.0001	0.49 (0.37-0.65)	<.0001
None	0.45 (0.34-0.59)	<.0001	0.54 (0.39-0.74)	.00012
Daily edible oil intake, no. (%)				
20-25 g	1		1	
25-30 g	1.59 (1.27-1.99)	<.0001	1.51 (1.17-1.97)	.0019
30-35 g	1.72 (1.38-2.15)	<.0001	1.79 (1.38-2.32)	<.0001
35-40 g	2.10 (1.61-2.74)	<.0001	2.25 (1.64-3.08)	<.0001
Unknown	2.56 (2.05-3.19)	<.0001	1.92 (1.49-2.49)	<.0001
Exercise, no. (%)				
No	1		–	
Yes	0.7 2 (0.54-0.95)	.0215	–	–
Fruits and vegetables intake, no. (%)				
No	1		1	
Yes	0.70 (0.61-0.80)	<.0001	0.81 (0.69-0.95)	.0077
Excessive milk tea, no. (%)				
No	1		–	–
Yes	1.30 (1.13-1.49)	.0002	–	–
Low-density lipoprotein cholesterol, median (IQR)	1.27 (1.19-1.36)	<.0001	1.16 (1.07-1.25)	.0002
High-density lipoprotein cholesterol, median (IQR)	0.76 (0.66-0.87)	.0002	–	–
Triglycerides, median (IQR)	1.48 (1.31-1.67)	<.0001	–	–
Total cholesterol, median (IQR)	1.24 (1.18-1.30)	<.0001	1.07 (1.02-1.13)	.0090
Abdominal circumference, median (IQR)	1.04 (1.04-1.05)	<.0001	1.01 (1.00-1.02)	.0036
Neck circumference, median (IQR)	1.01 (1.01-1.02)	.0003	–	–
Central obesity, no. (%)				
No	1		–	
Yes	2.17 (1.89-2.49)	<.0001	–	–
Prehypertension, no. (%)				
No	1		–	
Yes	3.97 (3.35-4.72)	<.0001	–	–
Family history, no. (%)				
Father	1		1	
Mother	1.99 (1.71-2.31)	<.0001	1.16 (0.97-1.39)	.097
Both father and mother	4.97 (3.69-6.77)	<.0001	2.96 (2.08-4.24)	<.0001

BMI = body mass index, CI = confidence interval, IQR = inter-quartile range, OR = odds ratio.

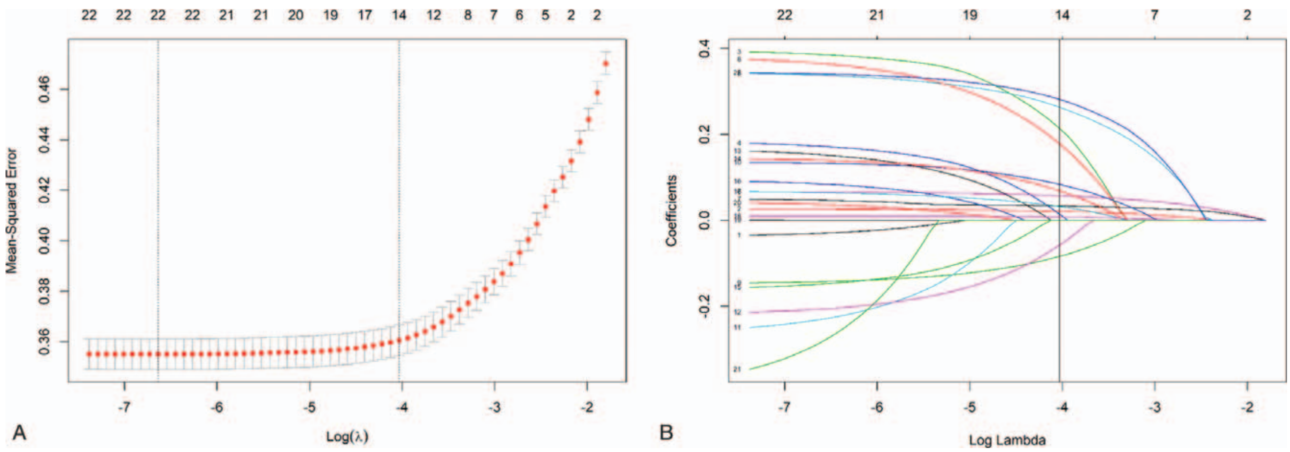

Figure 1. LASSO regression analysis. (A-B) The vertical line is marked as the number of hypertension-related risk factors included in the LASSO regression model when the optimal Lambda parameter is selected. LASSO = least absolute shrinkage and selection operator.

alcohol consumption, salt consumption, family history of hypertension, prehypertension, abdominal circumference, BMI, total cholesterol, triacylglycerol, and high-density lipoprotein were related to the onset of hypertension. Unhealthy lifestyles such as smoking and high salt intake could increase the risk of hypertension among Kazakh herders in pastoral areas in Xinjiang.[11] The dietary salt intake of most countries in the world is estimated to be 9 to 12 g/d.[29] However, the daily salt intake of the Kazakh population in the Nanshan pastoral area of Xinjiang was about 16.66 g, which is nearly 3 times the recommended salt intake of the WHO. The results of this study are consistent with the study of Rust et al,[30] suggesting that high salt intake is the main risk factor for hypertension.

In addition, we found that if the daily intake of fruits and vegetables was more than 500 g, the risk of hypertension was reduced. In a cohort study of dietary factors and chronic diseases conducted by Fang et al,[31] the incidence of hypertension was reduced by 1.9% for every day of fruit consumption. In addition, if there was every day of intake of fruits and vegetables, the incidence of high blood pressure decreased by 1.2%. BMI and waist circumference were independent risk factors for the risk of hypertension in female and male herders, respectively.[11] Herders have long used animal fats to make various pastries and stir-fries. Few intake of fresh fruits and vegetables and excessive intake of high-fat diets such as bacon, bacon, yak-butter, etc, causes fat to accumulate in the body, accompanied by

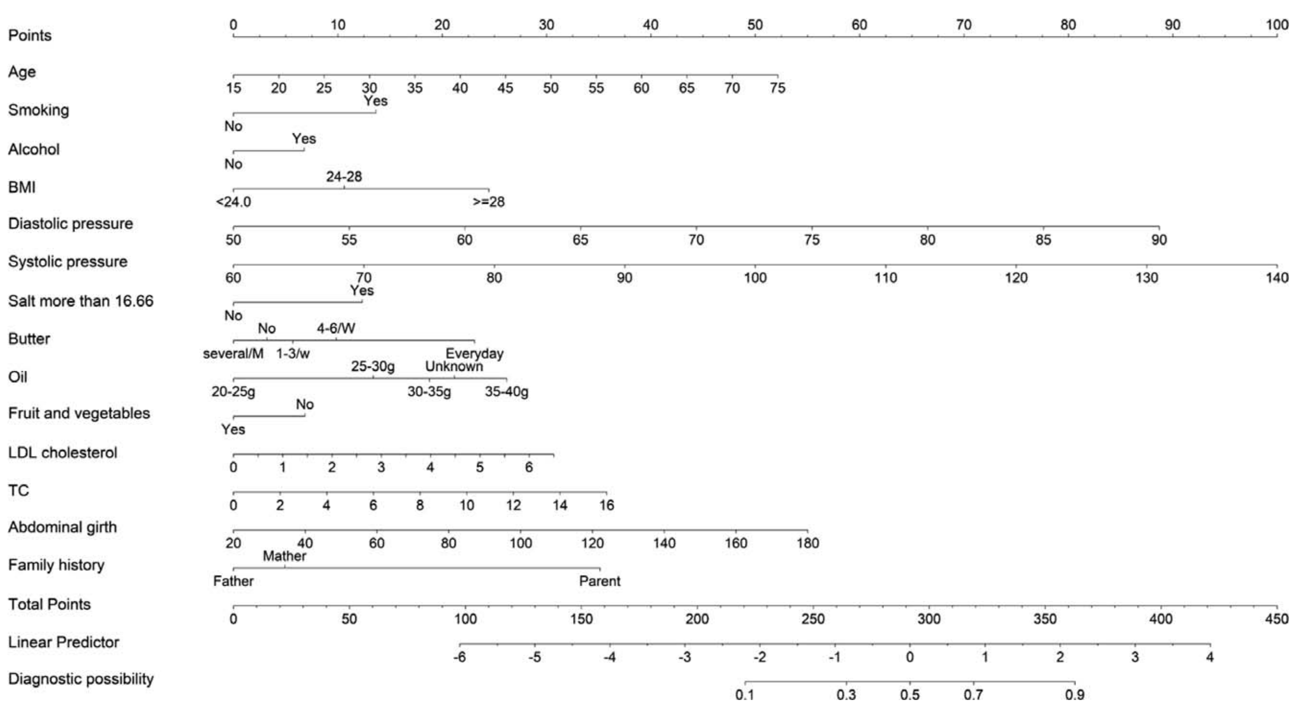

Figure 2. The nomogram prediction model. BMI = body mass index, LDL = low-density lipoprotein, TC = total cholesterol.

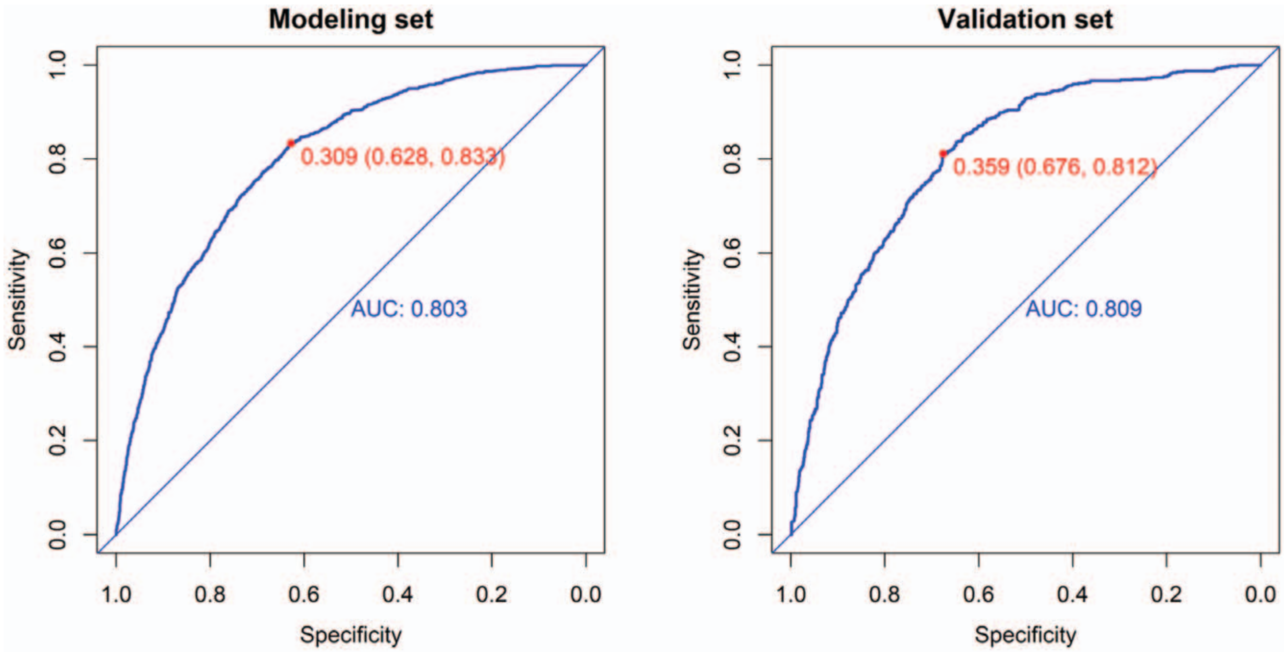

Figure 3. The ROC curve determined by the nomogram model of the modeling set and the validation set. AUC = area under the ROC curve, ROC = receiver operating characteristic.

changes in serological indicators, such as abnormal blood lipids, which are importantly related to vascular diseases such as high blood pressure.[31–35] Here, we found that abnormal total cholesterol in serum was also associated with high blood pressure. This may be related with their dietary lifestyles of high fat diets. However, these dietary lifestyles related to hypertension have not attracted attention among herders. Therefore, a systematic analysis of the factors affecting hypertension of Kazakh herders and screening of risk factors for hypertension to construct a risk prediction model can provide a basis for promoting Kazakh herders in Xinjiang pastoral areas to improve their unhealthy diet and lifestyle.

Here, in the Kazakh hypertension risk prediction nomogram model constructed by this study, we analyzed the risk factors based on univariate analysis, LASSO regression and multivariate logistic regression. In addition to the traditional 7 risk factors, including gender, age, SBP, DBP, smoking, family history, and BMI. We also included related risk factors such as salt intake, oil intake, yak-butter intake, and blood lipids. The results showed that in the nomogram model of hypertension prediction, the AUC of the modeling set and the validation set was 0.803 and 0.809, respectively. The best critical value in the ROC curve was 0.309 (0.628, 0.833) in the modeling set and 0.359 (0.676, 0.812) in the verification set. The AUC of 0.7 to 0.8 is considered acceptable

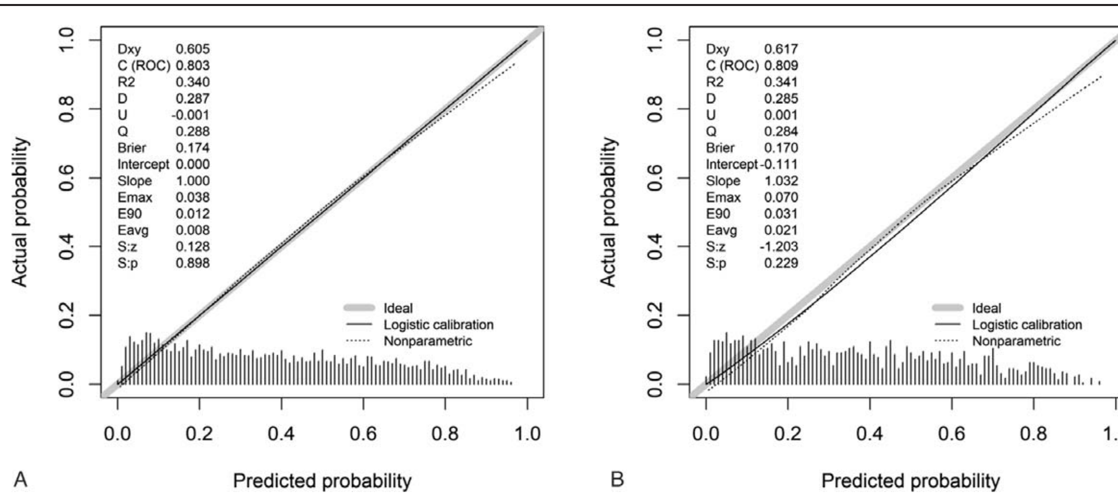

Figure 4. The calibration curve determined by the nomogram model Calibration of the modeling set (A) and the verification set (B). ROC = receiver operating characteristic.

for the model, and 0.8 to 0.9 is considered good.[36] Therefore, the nomogram model constructed in this study for hypertension prediction has good discriminative ability, and the calibration curve shows a good agreement between the nomogram prediction and the actual observation results of hypertension.

The nomogram model is a quantitative, reliable and easy-to-operate tool, and is now widely used in the prediction of a variety of diseases and the risk assessment of prognosis.[37] The nomogram model constructed in this study can be used as a simple tool for screening the high risk population of hypertension in the Kazakh population in the Nanshan pastoral area of Xinjiang. With this model, the high risk population of hypertension can be identified through assessing the information collected by physical examination. This not only facilitates the risk management of hypertension by primary health service workers, but also helps patients to understand their own risk of hypertension. Meanwhile, the nomogram model has low requirements on statistical knowledge and has the characteristics of visualization, which is convenient for medical personnel in the pastoral area of Nanshan, Xinjiang.

5. Conclusion

In conclusion, we constructed a nomogram model for hypertension risk prediction of Kazakhs based on Kazakh herdsmen aged 18 and above in the pastoral areas of Xinjiang, which has a good predictive ability for the risk of hypertension. However, this study is limited in that sample source was single. Additionally, due to different genetic background, this model was only applicable to Kazakh populations, but not Han Chinese populations. Therefore, further studies with larger sample sizes and multiple sample sources are needed.

Author contributions

Conceptualization: Shuping You.
Data curation: Yuezhen Xu, Jinbao Liu, Jiawei Wang, Qiongling Fan, Yuanyuan Luo, Huaifeng Zhan, Ning Tao.
Formal analysis: Yuezhen Xu, Jinbao Liu, Jiawei Wang, Qiongling Fan, Yuanyuan Luo, Huaifeng Zhan, Ning Tao.
Methodology: Yuezhen Xu, Jinbao Liu, Jiawei Wang, Qiongling Fan.
Software: Yuezhen Xu.
Supervision: Shuping You.
Writing – original draft: Yuezhen Xu.
Writing – review & editing: Shuping You.

References

[1] Zheng L, Sun Z, Zhang X, Li J, Hu D, Sun Y. The association between glomerular filtration rate and stroke in hypertensive patients in rural areas of China. J Hypertens 2012;30:901–7.
[2] He J, Whelton PK. Elevated systolic blood pressure and risk of cardiovascular and renal disease: overview of evidence from observational epidemiologic studies and randomized controlled trials. Am Heart J 1999;138:211–9.
[3] Klag MJ, Whelton PK, Randall BL, et al. Blood pressure and end-stage renal disease in men. N Engl J Med 1996;334:13–8.
[4] Whelton PK, He J, Appel LJ, et al. Primary prevention of hypertension: clinical and public health advisory from The National High Blood Pressure Education Program. JAMA 2002;288:1882–8.
[5] Whelton SP, Chin A, Xin X, He J. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. Ann Intern Med 2002;136:493–503.
[6] Xin X, He J, Frontini MG, Ogden LG, Motsamai OI, Whelton PK. Effects of alcohol reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 2001;38:1112–7.
[7] Kivimaki M, Batty GD, Singh-Manoux A, et al. Validating the Framingham Hypertension Risk Score: results from the Whitehall II study. Hypertension 2009;54:496–501.
[8] Goh LG, Welborn TA, Dhaliwal SS. Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study. BMC Womens Health 2014;14:118.
[9] Zheng L, Sun Z, Zhang X, et al. Predictive value for the rural Chinese population of the Framingham hypertension risk model: results from Liaoning Province. Am J Hypertens 2014;27:409–14.
[10] Ren Y, Fei H, Liang X, Ji D, Cheng M. A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records. BMC Med Inform Decis Mak 2019;19:51.
[11] Huang Y, Liang C, Liu Z. Nomogram for predicting pulmonary hypertension in patients without pulmonary embolism. Radiology 2016;280:327–8.
[12] Cheng H, Sun X, Ji X, et al. Risk factors and the potential of nomogram for predicting hospital-acquired pressure injuries. Int Wound J 2020;17:974–86.
[13] Kim M, Lee J. How to construct a nomogram for hypertension using complex sampling data from Korean adults. Commun Stat Theory Methods 2020; doi: 10.1080/03610926.2020.1774057.
[14] Shen QD. Establishment of a hospital-based carotid atherosclerosis risk nomogram for hypertensive patients. China Hosp Stat 2020;27:211–4.
[15] You S, Zhan H, Wang H. Analysis of the prevalence and risk factors of hypertension among the Kazakh population in the Nanshan pastoral area of Xinjiang. Mod Prev Med 2017;44:863–5.
[16] Chinese Guidelines for the Prevention and Treatment of Hyper-tensionGuidelines for the prevention and treatment of hypertension in China. Cardiovasc J 2018;19:1–45.
[17] Shen Z, Ding X, Mao D, et al. Cross-section survey on smoking behavior and cognition in Chongqing, 2014. Zhonghua Liu Xing Bing Xue Za Zhi 2015;36:1236–43.
[18] Wang J, Fan Q, Luo Y. Cox proportional hazard model analysis of factors affecting new-onset hypertension among Kazakh herders. Chin Gen Pract 2019;22:3672–8.
[19] Wei X, Sun G. Cholesterol intervention strategies based on the epidemic trend of cardiovascular disease in my country. Chin J Hypertens 2019;27:1107–10.
[20] Afshar M, Rong J, Zhan Y, et al. Risks of incident cardiovascular disease associated with concomitant elevations in lipoprotein(a) and low-density lipoprotein cholesterol-the Framingham Heart study. J Am Heart Assoc 2020;9:e014711.
[21] Du M, Yin S, Wang P, et al. Self-reported hypertension in Northern China: a cross-sectional study of a risk prediction model and age trends. BMC Health Serv Res 2018;18:475.
[22] Niiranen TJ, McCabe EL, Larson MG, et al. Risk for hypertension crosses generations in the community: a multi-generational cohort study. Eur Heart J 2017;38:2300–8.
[23] Zhang Y, Moran AE. Trends in the prevalence, awareness, treatment, and control of hypertension among young adults in the United States, 1999 to 2014. Hypertension 2017;70:736–42.
[24] Lu Y, Lu M, Dai H, et al. Lifestyle and risk of hypertension: follow-up of a young pre-hypertensive cohort. Int J Med Sci 2015;12:605–12.
[25] Ma GX, Lee M, Bhimla A, et al. Risk assessment and prevention of hypertension in Filipino Americans. J Community Health 2017;42:797–805.
[26] Wang YT, Adi D, Yu ZX, et al. The burden and correlates of hypertension among Chinese rural population in Han, Uygur, and Kazak: a cross-sectional study. J Am Soc Hypertens 2017;11:737–45 e3.
[27] Xu Y, Wang J, Liu J. Analysis of the prediction value of the Framingham hypertension risk prediction model in Xinjiang Kazakh herders. Chin Gen Pract 2019;23:1–7.
[28] Ma X, Cheng J, Zhao P, Li L, Tao K, Chen H. DNA methylation profiling to predict recurrence risk in stage I lung adenocarcinoma: development and validation of a nomogram to clinical management. J Cell Mol Med 2020;24:7576–89.
[29] Wu L, He Y, Jiang B, et al. Association between serum uric acid level and hypertension in a Chinese elderly rural population. Clin Exp Hypertens 2017;39:505–12.
[30] Rust P, Ekmekcioglu C. Impact of salt intake on the pathogenesis and treatment of hypertension. Adv Exp Med Biol 2017;956:61–84.
[31] Fang S, Zhang Q, Yang L, et al. Relationship between blood lipid level and blood pressure variability in patients with primary hypertension complicated with dyslipidemia. Guangxi Med J 2018;40:1925–8.

[32] Liu S, Zhang Y, Li S. Association between high-density lipoprotein subfractions and hypertensive status. Chin Circ J 2008;33:846–50.
[33] Chen Y, Zhu L, Yan W, Liu J, Lu J, Yan X. Relationship between dyslipidemia and hypertension among adult residents in Jiangxi Province. Chin J Health Educ 2008;34:503–8.
[34] Li Y, Li L. Analysis of the relationship between hypertension and dyslipidemia in physical examination population. World Latest Med Inform 2008;18:57.
[35] Lei X, Yang X, Yan X. Investigation on abnormal blood lipids of Han and Kazakh Nationalities in Zhaosu Reclamation Area. J Bingtuan Med 2018;3:31–4.
[36] Zhang R, Zheng L, Pan G. Application and establishment of disease risk prediction model. Chin Health Stat 2015;32:724–6.
[37] Deng QL, Dong S, Wang L, et al. Development and validation of a nomogram for predicting survival in patients with advanced pancreatic ductal adenocarcinoma. Sci Rep 2017;7:11524.