Observational Study

Medicine®

OPEN

Development and validation of a 2-year new-onset stroke risk prediction model for people over age 45 in China

Qiang Yao, MS[a], Jing Zhang, MS[a], Ke Yan, MS[a], Qianwen Zheng, MS[a], Yawen Li, MS[a], Lu Zhang, MS[a], Chenyao Wu, MS[a], Yanling Yang, MS[a], Muke Zhou, MD[b], Cairong Zhu, PhD[a,*]

Abstract

Multiple factors, including increasing incidence, poor knowledge of stroke and lack of effective, noninvasive and convenient stroke risk prediction tools, make it more difficult for precautions against stroke in China. Effective prediction models for stroke may assist to establish better risk awareness and management, healthier lifestyle, and lower stroke incidence for people.

The China Health and Retirement Longitudinal Survey was the development cohort. Logistic regression was applied to model's development, in which the candidate variables with statistically significant coefficient were included in the prediction model. The area under receiver operating characteristic curve (AUC) and 10-times cross-validation were used for internal validation. Cutoff point of high-risk group was measured by Youden index. The China Health and Nutrition Survey was the validation cohort.

The development cohort and the validation cohort included 16557 and 5065 participants, and the incidence density was 358.207/100,000 person-year and 350.701/100,000 person-year, respectively. The model for 2-year new-onset stroke risk prediction included age, hypertension, diabetes, heart disease, and smoking. The AUC and cross-validation AUC were 0.707 (95% confidence interval[CI]: 0.664, 0.750) and the 0.710 (95% CI: 0.650, 0.736). The sensitivity, specificity and accuracy of the cutoff point were 0.774, 0.545, and 0.319. The AUC and cross-validation AUC were 0.800 (95% CI: 0.744, 0.856) and 0.811(95% CI:0.714, 0.847), and the sensitivity, specificity and accuracy of cutoff point being 0.857,0.569, and 0.426 in external validation.

A simple prediction tool using 5 noninvasive and easily accessible factors can assist in 2-year new-onset stroke risk prediction in Chinese people over 45 years old, which is believed to be applicable in identifying high-risk individuals and health management in China.

Abbreviations: AUC = area under receiver operating characteristic curve, BMI = body mass index, Charls = the China Health and Retirement Longitudinal Survey, CHNS = the China Health and Nutrition Survey, CI = confidence interval, cvAUC = cross-validation AUC, FSRS = the Framingham Stroke Risk Score, ICVD = the Ischemic Cardiovascular Disease model, ROC = receiver operating characteristic curve, WHO = World Health Organization, WHtR = waist-to-height ratio.

Keywords: China, risk Assessment, stroke

1. Introduction

The incidence of stroke has been declining worldwide, but it is still increasing by 5.4% per year in China.[1] By 2016, China ranked the first in terms of age-specific stroke incidence and lifetime stroke risk[1,2] around the world and stroke became the first cause of death[3,4] and disability[5] in China. The incidence of stroke increases significantly with age, and has a much higher morbidity among people over age 45.[1,6] However, 45% of people at this age are unaware of any risk factor.[7] About 50% of the residents in China have limited knowledge about the risk factors of stroke.[8] Thus, multiple tasks make stroke prevention undoubtedly more intensified.[9,10] Both the annually increasing

Editor: Lenan Zhuang.

This research was supported by grants from National Natural Science Foundation of China (grant no. 30600511 and no. 81673273, http://www.nsfc.gov.cn/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funding agreement ensured the authors' independence in designing the study.

The authors have no conflicts of interest to disclose.

The data that support the findings of this study are available from a third party, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Data are available from the authors upon reasonable request and with permission of the third party.

[a] Department of Epidemiology and Health Statistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, [b] Department of Neurology, West China Hospital of Sichuan University, Chengdu, Sichuan, China.

[*] Correspondence: Cairong Zhu, No. 17 Section 3, Renmin South Road, Chengdu 610041, Sichuan, China (e-mail: cairong.zhu@hotmail.com).

How to cite this article: Yao Q, Zhang J, Yan K, Zheng Q, Li Y, Zhang L, Wu C, Yang Y, Zhou M, Zhu C. Development and validation of a 2-year new-onset stroke risk prediction model for people over age 45 in China. Medicine 2020;99:41(e22680).

Received: 10 March 2020 / Received in final form: 16 August 2020 / Accepted: 8 September 2020

http://dx.doi.org/10.1097/MD.0000000000022680

incidence of stroke and the lack of relevant knowledge among high-risk individuals cause the dilemma of stroke prevention in China.

Prediction model for individuals can be used for risk alarming and lifestyle planning.[11] As a result, stroke risk assessment, and management are of great significance for middle-aged and elderly people, who would be more aware of risk factors, and make behavioral adjustments to finally reduce the incidence.

Recently, stroke risk prediction model development has attracted much attention. The Framingham Stroke Risk Score (FSRS) involves gender, age, blood pressure, heart diseases, and diabetes mellitus to assess the risk in the next 10 years.[12] With AUC of 0.588,[13] FSRS poorly predicts the risk of stroke in Chinese population, and may overestimate the risk in Asian populations.[14] Subsequently, FSRS was modified particularly for Chinese population, but AUCs of the modified model for the male and female were only 0.726 and 0.656 respectively.[15] The Ischemic Cardiovascular Disease model (ICVD) is specially developed for the Chinese population and widely used, involving age, systolic blood pressure, body mass index (BMI), smoking, diabetes, and total cholesterol for risk prediction.[16] However, not only stroke, but Ischemic heart disease could be predicted by ICVD, so its ability to predict the single outcome of stroke is unclear. Additionally, the 2 models above are outdated (FSRS in 1994, ICVD in 2006) to evaluate the changing risk factors and effects on stroke.[17] Meanwhile, the target population in these 2 models fails to cover all middle-aged and elderly people (FSRS and ICVD cover age groups 55–85 and 35–59 respectively). Other stroke risk prediction tools for the Chinese population are mostly developed for special patient groups (e.g., diabetes mellitus, atrial fibrillation) rather than for the general population.[18–20]

Therefore, a new stroke risk prediction tool for Chinese people over age 45 was developed with larger sample size and relatively newer data in this study.

2. Methods

2.1. Study population

2.1.1. Development cohort. The development cohort was the China Health and Retirement Longitudinal Survey (Charls), which is conducted among families and individuals aged 45 and above in China every 2 years since 2011 and covers 28 provinces. The whole survey consists of 3 parts:

1. face-to-face electronic questionnaire: used to collect demographic information, health status and lifestyle from the respondents by professionals;
2. physical examination: including height, weight, waist circumference, blood pressure and other information;
3. laboratory examination (blood sample): conducted by the Chinese Center for Disease Control and Prevention. More information such as sampling methods, quality control, and laboratory inspection methods of Charls were published.[21,22]

2.1.2. Validation cohort. The validation cohort was the China Health and Nutrition Survey (CHNS), which is conducted by the Carolina Population Center at the University of North Carolina, the National Institute for Nutrition and Health and Chinese Center for Disease Control and Prevention since 1989, covering 9 provinces in China. Two waves of data were included in this study in 2009 and 2011. Face-to-face questionnaires and physical examinations were conducted by professionals, and blood samples were added in 2009. The samples were tested by the local community health service center and China-Japan Friendship Hospital. More information on CHNS were published.[23,24]

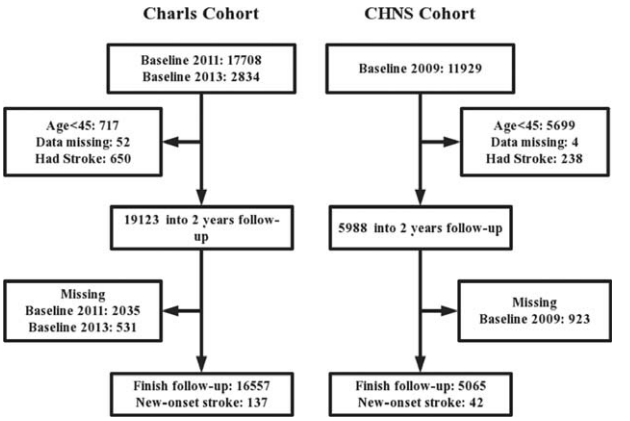

Figure 1. Derivation of the study population.

Figure 1 shows the derivation of 2 cohorts. Ethical approval is not necessary in this research.

2.2. Candidate predictors and outcome indicators

Based on previous studies, common understanding of epidemiology and the characteristics of development and validation cohorts, a total of 12 candidate predictive factors were included[19,25–31]: age, education, marital status, hypertension, diabetes, dyslipidemia, heart disease, sleep duration, smoking, alcohol consumption, BMI, and waist-to-height ratio (WHtR).

Hypertension status was determined based on the Hypertension Diagnostic Criteria of World Health Organization (WHO) and the Primary Prevention Guideline for Cerebrovascular Diseases of Chinese Population.[32,33] It was determined according to both self-reported result and physical examination because middle-aged and elderly people in China have limited awareness of hypertension.[34] Hypertension status was divided into 4 categories, including no hypertension, unknown hypertension, well-controlled hypertension, and poorly-controlled hypertension. If the measured blood pressure of a self-reported patient with no hypertension reached the WHO diagnostic criteria, this patient was classified as unknown hypertension. According to the guideline for primary prevention of stroke in China,[35] patients with self-reported hypertension, under age 65, with systolic blood pressure below 140 mm Hg and diastolic blood pressure below 90 mm Hg was considered as well-controlled hypertension, and otherwise, as poorly-controlled hypertension. For any patient aged 65 and above and reporting hypertension, the cutoff point of good systolic blood pressure control is 150 mmHg.[35] The mean blood pressure averaged from 3 times of physical examination was taken as the blood pressure of the individual.

The laboratory diagnosis of diabetes was based on the glucose metabolism status issued by WHO in 1999[35] and the Glycosylated Hemoglobin Standard by WHO in 2011.[36] The individual who met any of the diagnostic criteria for diabetes (fasting blood glucose ≥7.0 mmol/L, postprandial blood glucose ≥11.1 mmol/L or hemoglobin a1c ≥6.5%) or with reported diabetes was considered as diabetic.

Dyslipidemia was defined as the condition of self-report or meeting any of the following criteria: total cholesterol ≥240 mg/dL, triglyceride ≥200 mg/dL, high-density lipoprotein < 40 mg/dL, and low-density lipoprotein > 160 mg/dL.[37]

Drinking was defined as drinking any kind of alcohol more than once a month. BMI was calculated by kg (weight)/m^2(height). In China, BMI ≥24 kg/m^2 and BMI ≥28 kg/m^2 are considered as overweight and obesity respectively.[38,39] WHtR was calculated as the waist circumference (cm) divided by the height (cm). WHtR ≥ 0.5 is regarded to be abdominal obesity.[40]

The definition of heart disease in the development and validation cohorts is different. Heart disease includes heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems in the development cohort, but only includes myocardial infarction in the validation cohort.

The outcome indicator is the 2-year new-set intracranial hemorrhage or ischemic stroke during the 2-year follow-up period. In both cohorts, the outcome was clarified by face-to-face questionnaire, by asking "Have you been diagnosed with stroke by a doctor in the last 2 years?".

2.3. Statistics methods

T test and Chi-squared test was applied in the comparison of baseline characteristics and outcome between Charls and CHNS. Logistic regression was used to select the model variables. Heteroscedasticity is often found in generalized linear regression, so the heteroscedasticity robust standard error method was used for model correction.[41] In consideration of the confounding effect and convenience, a total of 12 candidate factors were included to develop the model by logistic regression, while the variables with statistically significant coefficient were included in the prediction system. The total risk score of each individual was calculated first by multiplying the logistic regression coefficient by 10 as the risk score for each predictor and then by summing the risk scores.

The reliability of the scores was examined by the Hosmer-Lemeshow goodness of fit test. Receiver operating characteristic curve (ROC) was plotted and the area under the curve (AUC) was calculated to evaluate the validity of scores. The Youden index was adopted to calculate the cutoff point of high risk, which was used to divide the high-risk group and the low-risk group. The sensitivity, specificity and accuracy of the cutoff point were calculated.

Internal validation of the model was performed with 10-fold cross-validation to calculate the cross-validation AUC (cvAUC), with 1995 as the random seed. The 95% confidence interval (CI) of cvAUC was computed by the Bootstrap method with 1000 times. The mean cvAUC of 10 times of 10-fold cross-validation was also calculated (random seed: 1986–1995). External validation of the model was conducted in the validation cohort. In addition to cross-validation, the sensitivity, specificity and accuracy of the cutoff point were also verified. Fisher exact test was applied to compare the difference in incidence density between low-risk group and high-risk group in both cohorts.

Sensitivity analysis was applied to compare the AUC and cvAUC when all 12 risk factors were taken as the predictors (Situation A) and when the excluded covariates were taken as the predictors (Situation B).

Stata (Stata13 Corp, College Station, TX) was used for statistics analysis and the level of 2-sided significance was < 0.05.

3. Results

3.1. Baseline characteristics and incidence

Table 1 showed that the proportions of males, education level at high school or above, heart diseases, smokers, and WHtR ≥0.5 were all higher in Charls than those in CHNS, while the proportions of dyslipidemia, drinkers, and overweight or obesity were all less than those in CHNS. A total of 137 and 42 patients suffered stroke during the 2-year follow-up in Charls and CHNS, with incidence density of 358.207/ 100,000 person-year and 350.701/100000 person-year respectively.

3.2. Risk factor selection and model construction

The development cohort indicated that heart disease, hypertension status, age, diabetes, and smoking were risk factors for stroke (Table 2). Compared with those without hypertension, hypertension patient of any state were more likely to suffer stroke. Compared with those aged 45 to 55 years old, the older ones suffered higher risk of stroke. Regression coefficients and risk score for each factor were showed in Table 2.

3.3. Model prediction effect and internal validation

The risk of stroke got higher when there was a higher total risk score, with the regression coefficient being 0.104 (Odds Ratio: 1.109, 95%CI: 1.084, 1.134). In the Hosmer-Lemeshow goodness-of-fit test for the prediction model, *P* value was .178. AUC was 0.707 (95%CI: 0.664, 0.750) (Fig. 2).

The cvAUC was 0.710(95%CI: 0.650, 0.736) (Fig. 2). The cvAUC of 10 times 10-fold cross-validation ranged from 0.703 to 0.714, with the average of 0.707. The ROC line and cross-validation ROC line in Charls were close.

When the Youden index was the largest, the total risk score was 9.485, which was taken as the cutoff point of high-risk. The high-risk group involved 7581 individuals, 106 of whom suffered stroke in 2 years; the low-risk group involved 8976 individuals, 31 of whom suffered stroke. The incidence was shown in Figure 3. The sensitivity, specificity and accuracy of this cutoff point were 0.774 (95%CI: 0.704, 0.844), 0.545 (95%CI: 0.537, 0.552), and 0.547 (95%CI: 0.539, 0.554), respectively.

For sensitivity analysis, no risk factor was excluded, so all 12 risk factors were taken as the predictors. In the development cohort, the Hosmer-Lemeshow goodness-of-fit test showed *P* value was .337. AUC for the new total risk score was 0.720, and cvAUC was 0.724. The sensitivity, specificity and accuracy of new cutoff point were 0.745, 0.595, and 0.596, respectively.

3.4. External validation

In the validation cohort, the regression coefficient of total risk score was 0.151, and OR was 1.163 (95%CI: 1.121, 1.206). The Hosmer-Lemeshow goodness-of-fit test indicated the model fitted well ($P = .260$), and AUC was 0.800 (95%CI: 0.744, 0.856) (Fig. 2).

The cvAUC was 0.811 (95%CI: 0.714, 0.847) (Fig. 2). The cvAUC of 10-fold cross-validation for 10 times was between 0.789 and 0.822, with an average of 0.802. The cross-validation ROC line in CHNS was higher than the ROC line, and these 2 lines in CHNS were not as close as in Charls.

In the validation cohort, the high-risk group involved 2,199 individuals, 36 of whom suffered stroke; the low-risk group

Table 1

Baseline characteristics and outcome of the study populations.

Variable	Charls N = 16557	CHNS N = 5065	*P* value
Outcome			
Number of new-onset stroke	137	42	.990
Incidence density (/100,000 py)*	358.207	350.701	
Gender			**.006**
Male n(%)	8048 (48.6%)	2351 (46.4%)	
Age Mean±SD	58.99 ± 9.95	58.94 ± 9.94	.711
45– n(%)	6115 (36.9%)	1958 (38.7%)	.210
55– n(%)	6006 (36.3%)	1737 (34.3%)	
65– n(%)	3064 (18.5%)	946 (18.6%)	
75– n(%)	1372 (8.3%)	424 (8.4%)	
Marital status			**<.001**
Married and living together n(%)	13481 (81.4%)	4737 (93.5%)	
Married but separated n(%)	1031 (6.2%)	15 (0.3%)	
Not married n(%)	2045 (12.4%)	313 (6.2%)	
Education level			**<.001**
Senior high-school or above n(%)	5426 (32.8%)	969 (19.1%)	
Hypertension			.903
No hypertension n(%)	10413 (62.9%)	3105 (61.3%)	
Unknown hypertension n(%)	2262 (13.7%)	1073 (21.2%)	
Hypertension with good control n(%)	2280 (13.8%)	287 (5.7%)	
Hypertension with bad control n(%)	1602 (9.7%)	600 (11.8%)	
Dyslipidemia n(%)†	4881 (29.5%)	1727 (34.1%)	**<.001**
Total cholesterol ≥240 mg/dL n(%)	1135 (6.9%)	573 (11.3%)	**<.001**
Triglyceride ≥200 mg/dL n(%)	1465 (8.8%)	962 (19.0%)	**<.001**
HDL < 40 mg/dL n(%)	2459 (14.9%)	544 (10.7%)	**<.001**
LDL > 160 mg/dL n(%)	1043 (6.3%)	648 (12.8%)	**<.001**
Diabetes mellitus n(%)	1906 (11.5%)	632 (12.5%)	.062
Heart disease n(%)‡	1904 (11.7%)	65 (1.3%)	**<.001**
Sleep duration<7h n(%)§	7652 (46.2%)	151 (14.2%)	
Smoking n(%)			**<.001**
No smoke or quit smoke	11266 (68%)	3632 (71.7%)	
Yes	5291(32%)	1433 (28.3%)	
Drinking n (%)			**<.001**
No drinking or quit drinking	12238(73.9%)	3600 (71.1%)	
Yes	4319 (26.1%)	1465 (28.9%)	
BMI Mean ± SD	23.55 ± 3.57	23.60 ± 3.39	.480
<24 n(%)	11175 (75.7%)	2956 (58.4%)	**<.001**
24–28 n(%)	3857 (23.3%)	1616 (31.9%)	
≥28 n(%)	1525 (9.2%)	493 (9.7%)	
WHtR≥0.5 n(%)	12818 (77.4%)	3395 (67.0%)	**<.001**

* Including missing people after 2-year follow-up.

† Dyslipidemia was defined as the condition of self-report or meeting any of the following criteria: total cholesterol ≥240 mg/dL, triglyceride ≥200 mg/dL, high-density lipoprotein < 40 mg/dL, and low-density lipoprotein > 160 mg/dL.

‡ The definition of heart disease is different in 2 cohort.

§ Too many missing in CHNS cohort and it is 151 in 1063 people with sleep duration data.

Py stands for person-years.

BMI = body mass index, Charls = the China Health and Retirement Longitudinal Survey, CHNS = the China Health and Nutrition Survey.

involved 2866 patients, 6 of whom suffered stroke in 2 years. Figure 3 shows the incidence of high-risk and low-risk groups. The sensitivity, specificity and accuracy of this cutoff point were 0.857 (95%CI:0.751, 0.963), 0.569 (95%CI:0.556, 0.583), and 0.572 (95%CI:0.558, 0.585) respectively.

3.5. Sensitivity analysis

In the development cohort, in Situation A, when all 12 risk factors were taken as the predictors, the AUC for the new total risk score was 0.720, and cvAUC was 0.724. In Situation B, when the excluded covariates, as education, marital status, dyslipidemia, sleep duration, alcohol consumption, BMI, and WHtR were taken as the predictors, the AUC for covariates was 0.610, and cvAUC for covariates was 0.614.

We did not include sleep duration into sensitivity analysis because there were too many missing in sleep duration data in CHNS cohort. The AUC was 0.819 and cvAUC was 0.815 in Situation A and the AUC for covariates and cvAUC for covariates were 0.614 and 0.599 in Situation B.

4. Discussion

In this study, a 2-year risk prediction model for new-onset stroke in Chinese people over 45 years old has been developed. The model showed high predictive value and discrimination ability in

Table 2

Multivariable logistic model and risk score associated with each risk factor for 2-year risk of new-onset stroke for participants in the China Health And Retirement Longitudinal Survey study.

Variable	β	OR	95%CI	*P* value	Risk Score
Heart disease					
No (Ref)	0.000	1.000			0
Yes	0.536	1.709	1.119–2.611	**.013**	5.36
Hypertension status					
No hypertension (Ref)	0.000	1.000			0
Unknown hypertension	0.789	2.202	1.370–3.541	**.001**	7.89
Hypertension with good control	0.827	2.286	1.405–3.720	**.001**	8.27
Hypertension with bad control	0.971	2.642	1.555–4.487	**<.001**	9.71
Age					
45– (Ref)	0.000	1.000			0
55–	0.539	1.714	1.039–2.826	**.035**	5,39
65–	0.738	2.091	1.209–3.616	**.008**	7.38
75–	1.349	3.853	2.090–7.105	**<.001**	13.49
Diabetes mellitus					
No (Ref)	0.000	1.000			0
Yes	0.436	1.546	1.005–2.379	**.047**	4.36
Smoking					
No (Ref)	0.000	1.000			0
Yes	0.387	1.473	1.007–2.156	**.046**	3.87
Sleep duration<7h	−0.006	0.994	0.706–1.401	.975	
Drinking	0.053	1.054	0.704–1.579	.798	
Marital status					
Married and living together(Ref)	0.000	1.000			
Married but separated	−0.478	0.620	0.232–1.657	.341	
Not married	0.203	1.225	0.771–1.945	.391	
Education level					
Senior high-school or above(Ref)	0.000	1.000			
Less than senior high-school	−0.073	0.930	0.632–1.369	.713	
BMI					
<24 (Ref)	0.000	1.000			
24–28	0.049	1.051	0.696–1.586	.814	
≥28	−0.600	0.549	0.260–1.158	.115	
Dyslipidemia	0.070	1.072	0.728–1.578	.724	
WHtR≥0.5	0.341	1.406	0.865–2.286	.169	
Constant	−6.303	0.002	0.001–0.004	**<.001**	

BMI = body mass index, CI = confidence interval, OR = odds ratio, Ref = reference, WHtR = waist-to-height ratio.

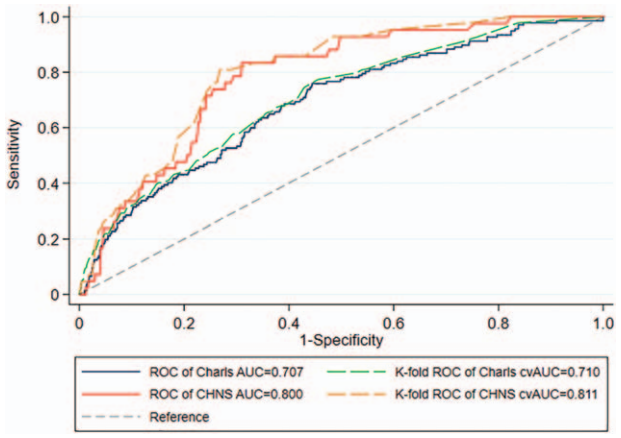

Figure 2. Cross-validation ROC curves of 2-year new-onset stroke risk prediction model in the China Health and Retirement Longitudinal Survey study and the China Health and Nutrition Survey study: cross-validation area under receiver operating characteristic curve stands for cross-validation AUC.

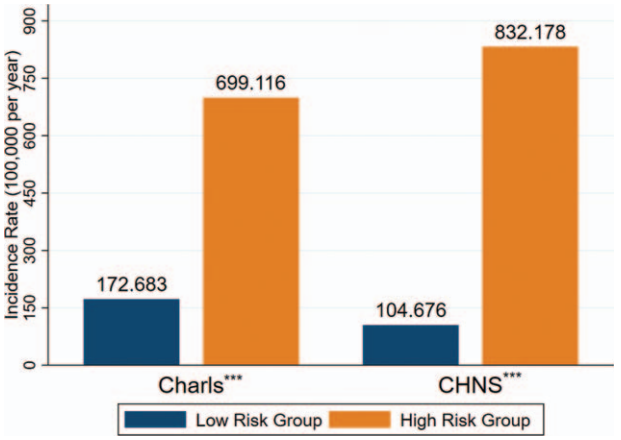

Figure 3. Different incidence rate between low risk group and high risk group in the China Health and Retirement Longitudinal Survey and he China Health and Nutrition Survey: *** indicates $P < .001$ (Fisher exact test between low risk group and high risk group).

external validation and is believed to be applicable in health management to identify individuals with higher risk of 2-year new-onset stroke in Chinese people over 45 years old.

The prediction tool fitted well, and a higher risk score indicated a higher risk of the disease. In the development cohort, stroke risk had a tendency of increasing 10.9% when the total risk score increased by 1 point, and the trend was also observed in the validation cohort. The model showed good prediction performance in both development cohort and validation cohort. The cvAUC was 0.710 in development cohort, and 0.811in validation cohort. The average AUC in this study was higher than that in FSRS among Chinese and in modified FSRS, indicating that this model might perform better in predicting risk among Chinese. The sensitivity of high-risk cutoff point exceeded 0.75 in both development and validation cohort, while the specificity did not reach 0.6 in either cohort, and only reached 0.612 among the female in the CHNS cohort. Despite the relatively low specificity, the incidence among the screened high-risk groups was far higher than that in the low-risk group, indicating the model could be well applied in screening high-risk groups. Though there were some differences in prediction performance in 2 cohorts, this model was believed to distinguish the high-risk group of new-onset stroke in Chinese population.

The classification of hypertension status in this study is quite different from other stroke prediction tools. In consideration of the difference in the risk of stroke among people with different awareness of hypertension and blood pressure control,[34,44,45] we divided hypertension status into 4 types of no, unknown, well-controlled, and poorly-controlled hypertension, so the model could be used to predict the risk of stroke in different states of hypertensive patients. The risks of stroke in patients with unknown, well-controlled and poorly-controlled hypertension were 2.202, 2.286, and 2.642 times higher than those without hypertension, respectively. A meta-analysis has argued that antihypertensive therapy could effectively reduce the risk of stroke in the elderly.[42,43] Those patients with undiagnosed hypertension, unreasonable treatment and poor blood pressure control were at much higher risk of stroke than those with normal blood pressure.[44] Every 10 mm Hg reduction in systolic blood pressure significantly reduced the risk of stroke, with relative risk being 0.73.[45] Other factors including diabetes, heart diseases, age, and smoking are all clearly correlated with stroke.[25,26,29,31]

The sensitivity and AUC of the model in CHNS were higher, but the ROC line and cross-validation ROC line were closer in Charls, and the gap between AUC and cvAUC was larger in CHNS, which might be correlated with some differences between the 2 cohorts. The heart diseases in Charls involved a number of different types of heart problems, while only heart attack/myocardial infarction were included in CHNS. This suggested myocardial infarction had greater impact on the new-onset stroke than other heart diseases. As reported, the risk of Ischemic stroke increases due to residual ischemic risk within 4 years after the onset of myocardial infarction.[46] In addition, the samples in Charls were more representative due to the coverage of 28 provinces of China, while the samples in CHNS only coverd 9 provinces, and the sample size of Charls was larger than CHNS's, which might lead to the different prediction performance in 2 cohorts. China is vast and the morbidity of stroke is higher in north China and relatively lower in south China.[47,48] As a consequence, the multilevel model may be applied to develop a disease risk prediction model in the future.

Logistic regression rather than Cox proportional hazard regression was used here because the follow-up interval was 2 years and the minimum time gap was year (in the questionnaire "When (in which year) was the stroke first diagnosed or known by yourself?"). Some individuals did not know the new-onset age or the disease-free time.

Only factors with statistically significant coefficient were selected to construct the prediction model, which contributed to the simplified and convenient application of the model. Sensitivity analysis shows the AUC and cvAUC were only enlarged by 0.013 and 0.014 in Charls when taking all candidate factors in the model construction. In CHNS, the increase of AUC and cvAUC were only enlarged by and 0.019 and 0.004, respectively. Although the prediction performance improved, the improvement is fairly limited in the consideration of the inconvenience in application with 7 additional factors. The AUC and cvAUC for covariates were lower than the AUC and cvAUC of this prediction tool in both Charls and CHNS, which meant these covariates poorly predicted stroke.

Although the model has good prediction and discrimination performances, there are still some limitations. For example, indispensable risk factors such as family history of stroke, precise diagnosis of heart disease like atrial fibrillation and antiplatelet drugs were not included owing to the limitations of Charls and CHNS data. In addition, only 50% of the respondents in the Charls survey answered the items of physical exercise, so this factor was ignored. The diagnosis of stroke was based on self-report in both Charls and CHNS cohorts and both cohorts failed to distinguish intracranial hemorrhage or Ischemic stroke, which might lead to potential bias. Due to lack of onset time, Cox proportional hazard regression cannot be used for model development.

The novel model is available to all people over 45 years of age in China and divides the hypertensive population into 4 categories and considers the impact of blood pressure control of hypertensive patients on the stroke risk, which is rarely involved in previous risk prediction models. This model can predict new-onset 2-year stroke risk, so it is expected to be beneficial for people to be aware of their risk, improving living habits or seek medical and health services so as to enhance the primary prevention of stroke.

Acknowledgments

We thank Dr Fengdi Cao (DDs & PhD) for his sincerely help to this study and invaluable guidance to author Qiang Yao. We also thank the investigators, staff, and participants of the Charls study and CHNS study for their valuable contributions. A full list of Charls and CHNS study investigators/institutions can be found at http://charls.pku.edu.cn/ and https://www.cpc.unc.edu/projects/china/.

Author contributions

Conceptualization: Qiang Yao.
Data curation: Qiang Yao.
Formal analysis: Qiang Yao, Yanling Yang.
Funding acquisition: Cairong Zhu.
Methodology: Yawen Li, Lu Zhang, Chenyao Wu.
Project administration: Cairong Zhu.
Resources: Jing Zhang, Ke Yan, Qianwen Zheng.
Software: Cairong Zhu.
Supervision: Muke Zhou, Cairong Zhu.
Validation: Qiang Yao.

Visualization: Qiang Yao, Jing Zhang, Ke Yan.
Writing – original draft: Qiang Yao.
Writing – review & editing: Jing Zhang, Ke Yan, Qianwen Zheng.

References

[1] GBD 2016 Stroke CollaboratorsGlobal, regional, and national burden of stroke, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol 2019;18:439–58.
[2] Feigin VL, Nguyen G, Cercy K, et al. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. N Engl J Med 2018;379:2429–37.
[3] Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the global burden of disease study 2013. Lancet 2016;387:251–72.
[4] Chen Z. The Third Nationwide Survey on Causes of Death. Beijing: The Peking Union Medical College Press; 2008.
[5] Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet 2019;394:1145–58.
[6] Wang W, Jiang B, Sun H, et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. Circulation 2017;135:759–71.
[7] Slark J, Bentley P, Majeed A, et al. Awareness of stroke symptomatology and cardiovascular risk factors amongst stroke survivors. J Stroke Cerebrovasc Dis 2012;21:358–62.
[8] Sun H, Chen S, Jiang B, et al. Public knowledge of stroke in Chinese urban residents: a community questionnaire study. Neurol Res 2011;33: 536–40.
[9] Samsa GP, Cohen SJ, Goldstein LB, et al. Knowledge of risk among patients at increased risk for stroke. Stroke 1997;28:916–21.
[10] Soomann M, Vibo R, Korv J. Do stroke patients know their risk factors? J Stroke Cerebrovasc Dis 2016;25:523–6.
[11] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Eur J Clin Invest 2015;45:204–14.
[12] D'Agostino RB, Wolf PA, Belanger AJ, et al. Stroke risk profile: adjustment for antihypertensive medication. the Framingham study. Stroke 1994;25:40–3.
[13] Huang XY, Fu WJ, Chen ZC, et al. Association between FSP, CVHI, inflammatory cytokines and the incidence of primary stroke. J Clin Neurosci 2017;45:265–9.
[14] Grundy SM, D'Agostino RS, Mosca L, et al. Cardiovascular risk assessment based on US cohort studies: findings from a national heart, lung, and blood institute workshop. Circulation 2001;104:491–6.
[15] Huang J, Cao Y, Guo J, et al. Modified Framingham stroke profile in the prediction of the risk of stroke among Chinese. Chin J Cerebrovasc Dis 2013;10:228–32.
[16] Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation 2006;114:2217–25.
[17] Dufouil C, Beiser A, McLure LA, et al. Revised framingham stroke risk profile to reflect temporal trends. Circulation 2017;135:1145–59.
[18] Zhang H, Wang C, Ren Y, et al. A risk-score model for predicting risk of type 2 diabetes mellitus in a rural Chinese adult population: a cohort study with a 6-year follow-up. Diabetes Metab Res Rev 2017;33:e2911.
[19] Li TC, Wang HC, Li CI, et al. Establishment and validation of a prediction model for ischemic stroke risks in patients with type 2 diabetes. Diabetes Res Clin Pract 2018;138:220–8.
[20] Menon BK, Saver JL, Prabhakaran S, et al. Risk score for intracranial hemorrhage in patients with acute ischemic stroke treated with intravenous tissue-type plasminogen activator. Stroke 2012;43:2293–9.
[21] Zhao Y, Strauss J, Yang G, et al. China health and Retirement Longitudinal Study—2011-2012 National Baseline Users' Guide. Beijing: Peking University; 2013.
[22] Zhao Y, Hu Y, Smith JP, et al. Cohort profile: the China health and retirement longitudinal study (CHARLS). Int J Epidemiol 2014;43:61–8.
[23] Popkin BM, Du S, Zhai F, et al. Cohort profile: the China health and nutrition survey–monitoring and understanding socio-economic and health change in China, 1989-2011. Int J Epidemiol 2010;39:1435–40.
[24] Zhang B, Zhai FY, Du SF, et al. The China health and nutrition survey, 1989–2011. Obes Rev 2014;15(Suppl 1):2–7.
[25] O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet 2016;388:761–75.
[26] Wang Y, Liu J, Wang W, et al. Lifetime risk of stroke in young-aged and middle-aged Chinese population. J Hypertens 2016;34:2434–40.
[27] Yu P, Pan Y, Zheng H, et al. Association of high waist-to-height ratio with functional outcomes in patients with acute ischemic stroke: a report from the ACROSS-China study. Medicine (Baltimore) 2017;96:e6520.
[28] Li W, Wang D, Cao S, et al. Sleep duration and risk of stroke events and stroke mortality: a systematic review and meta-analysis of prospective cohort studies. Int J Cardiol 2016;223:870–6.
[29] Feigin VL, Roth GA, Naghavi M, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the global burden of disease study 2013. Lancet Neurol 2016;15:913–24.
[30] Jia Q, Liu L, Wang Y. Risk factors and prevention of stroke in the Chinese population. J Stroke Cerebrovasc Dis 2011;20:395–400.
[31] Wang J, Wen X, Li W, et al. Risk factors for stroke in the Chinese population: a systematic review and meta-analysis. J Stroke Cerebrovasc Dis 2017;26:509–17.
[32] World Health OrganizationGlobal Status Report on Noncommunicable Diseases 2010. Geneva: World Health Organization; 2011.
[33] Neurology CSO, Society CS. Guidelines for the primary prevention of cerebrovascular diseases in China 2019. Chin J Neurol 2019;52:684–709.
[34] Ning M, Zhang Q, Yang M. Comparison of self-reported and biomedical data on hypertension and diabetes: findings from the China health and retirement longitudinal study (CHARLS). BMJ Open 2016;6:e9836.
[35] World Health OrganizationDefinition and Classification of Diabetes Mellitus and its Complications-Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: World Health Organization; 1999.
[36] World Health Organization. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus:abbreviated report of a WHO consultation; 2011.
[37] Joint Committee on Revision of Guideline for Prevention and Treatment of Dyslipidemia in Chinese AdultsGuideline for prevention and treatment of dyslipidemia in Chinese adults (2016 edition). Chin Circ J 2016;31:937–53.
[38] Department of Disease Control, Ministry of Health of the People's Republic of ChinaGuidelines for Prevention and Control of Overweight and Obesity in Chinese adults. Beijing: People's Health Publishing House; 2006.
[39] Zhou BF. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults–study on optimal cut-off points of body mass index and waist circumference in Chinese adults. Biomed Environ Sci 2002;15:83–96.
[40] Yang H, Xin Z, Feng JP, et al. Waist-to-height ratio is better than body mass index and waist circumference as a screening criterion for metabolic syndrome in Han Chinese adults. Medicine (Baltimore) 2017;96:e8192.
[41] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817–38.
[42] Gaciong Z, Sinski M, Lewandowski J. Blood pressure control and primary prevention of stroke: summary of the recent clinical trial data and meta-analyses. Curr Hypertens Rep 2013;15:559–74.
[43] Parsons C, Murad MH, Andersen S, et al. The effect of antihypertensive treatment on the incidence of stroke and cognitive decline in the elderly: a meta-analysis. Future Cardiol 2016;12:237–48.
[44] Han TS, Wang HH, Wei L, et al. Impacts of undetected and inadequately treated hypertension on incident stroke in China. BMJ Open 2017;7: e16581.
[45] Ettehad D, Emdin CA, Kiran A, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016;387:957–67.
[46] Abtan J, Bhatt DL, Elbez Y, et al. Residual ischemic risk and its determinants in patients with previous myocardial infarction and without prior stroke or TIA: insights from the REACH registry. Clin Cardiol 2016;39:670–7.
[47] Wu Z, Yao C, Zhao D, et al. Sino-MONICA project: a collaborative study on trends and determinants in cardiovascular diseases in China, part i: morbidity and mortality monitoring. Circulation 2001;103:462–8.
[48] Xu G, Ma M, Liu X, et al. Is there a stroke belt in China and why? Stroke 2013;44:1775–83.