BMJ Open

Risk Factors and Nomogram for Predicting Stroke in Middle-Aged and Elderly Chinese Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-048734
Article Type:	Original research
Date Submitted by the Author:	08-Jan-2021
Complete List of Authors:	Yu, Qi; the First Hospital of Putian, Department of Scientific Research and Education Wu, Yuanzhe; the First Hospital of Putian, Department of Scientific Research and Education Jin, Qingdong; the First Hospital of Putian, Department of Neurosurgery Chen, Yanqing; the First Hospital of Putian, Department of Scientific Research and Education Lin, Qingying; the First Hospital of Putian, Department of Scientific Research and Education Liu, Xinru; the First Hospital of Putian, Department of Scientific Research and Education
Keywords:	Epidemiology < TROPICAL MEDICINE, Stroke < NEUROLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Risk Factors and Nomogram for Predicting Stroke in Middle-Aged and Elderly Chinese Population

Qi Yu[1*]

Yuanzhe Wu[1]

Qingdong Jin[2]

Yanqing Chen[1]

Qingying Lin[1]

Xinru Liu[1]

1. Department of Scientific Research and Education,the First Hospital of Putian, Putain city, Fujian province,China

2. Department of Neurosurgery,the First Hospital of Putian, Putain city, Fujian province,China

Correspondence to:

Qi Yu, Department of Scientific Research and Education,the First Hospital of Putian, Putain city, Fujian province,China. Tel: +86 594 6923557, E-mail: madyuqi@126.com

Key words:Stroke,Risk factor,Epidemiology

Word count：3673

1. Abstract

Objective:This study was aimed to assess the association of known and emerging risk factors with stroke and its primary subtypes, and develop a nomogram for predicting stroke risk in middle-aged and elderly Chinese population.

Method:This is a retrospective cohort study from a prospectively collected database. We included a total 3,214 adults aged 45-80 years free of stroke or myocardial infarction at baseline in the 2009-2015 cohort of China Health and Nutrition Survey (CHNS). Multivariate Cox regression analyses were applied to identify independent predictors. A nomogram was constructed to predict 6-year risk of stroke based on the multivariate analysis results. Bootstraps with 1000 resamples were applied to both C-index and calibration curve .

Result:The overall prevalence of overall stroke was 2.98%. After multivariate analysis , age,gender,hypertension and low-density lipoprotein cholesterol(LDL-C) were found as significant risk predictors for overall stroke and ischemic stroke;age,gender,hypertension and body mass index (BMI)were found as significant risk predictors for hemorrhagic stroke.The nomogram was constructed using significant variables included in the model,with a bootstrap-corrected C-index of 0.75,0.77 and 0.81 for overall stroke,ischemic stroke and hemorrhagic stroke model,respectively.The calibration curves demonstrating the good agreements between predicted and observed 6-year risk probability.

Conclusion:Our nomogram could be convenient, easy to use, and effective prognoses for predicting 6-year risk of developing stroke in middle-aged and elderly Chinese population. Nomogram could be used to assist the level 1 prevention of stroke and provide key information to reduce the incidence and burden of diseases.

Key words:Stroke,Risk factor,Epidemiology

Strengths and limitations of this study

Strengths

1. Previous studies that develop the stroke risk prediction models were initially developed from the western populations[6]-[8],thus application of these model to general Chinese population is still questionable. In addition,for young adults,using 5-year or 10-year risk model could affect early identification and health education of persons with low short-term risk but high long-term risk[9] .Therefore,we developed a simple, convenient, and efficient model to predicting the 6-year risk of stroke among middle-aged and elderly Chinese adults.

2. Furthermore,many stroke risk prediction model only built for overall stroke[10][11]. However, there are some notable differences of risk factors between ischemic and hemorrhagic stroke[12]. Therefore,we developed a

simple, convenient, and efficient model to predicting the 6-year risk of overall ,ischemic and hemorrhagic stroke among middle-aged and elderly Chinese adults.

limitations

1. Our study has several limitations. First, we only screen residents older older then 45 and younger then 80 years old from nine provinces of China. Therefore, our current results cannot be generalized to all populations in China.

2. Second,the long-term,>6 years,incidence of stroke for these risk factors are unknown .In future studies, we project to prospective follow-up participants, gather data about mortality and incidence of stroke in participants.

3. Finally, the nomogram has not been externally validate.However, we used a 1000 bootstrap resampling strategy for internal validation and the nomogram showed good performance in terms of calibration and discrimination for predicting risk of stroke in middle-aged and elderly Chinese person .

2. Introduction

Stroke is the leading cause of death and is a major cause of permanent disability worldwide[1][2].In 1990-2010,the incidence of stroke has decreased by 12% in high-income countries, because of effective strategies for preventing

cerebrovascular risk factor and good health services in developed countries. On the contrary , the age-adjusted incidence of stroke significantly increased by 12% in low-income and middle-income countries[3]. In China,the world's most populous country,the incidence of stroke increase by 8.3% among adults aged 40 years and older during 2002 to 2013[4][5].

Therefore, stroke prevention is essential for enhancing public health and reducing social burden in countries with heavy stroke burden such as China. Prediction model for actively assessing stroke risk are required to ensure targeted strategies for stroke prevention and management in high-risk groups .

Previous studies that develop the stroke risk prediction models were initially developed from the western populations[6]-[8],thus application of these model to general Chinese population is still questionable. In addition,for young adults,using 5-year or 10-year risk model could affect early identification and health education of persons with low short-term risk but high long-term risk[9] .Furthermore,many stroke risk prediction model only built for overall stroke[10][11]. However, there are some notable differences of risk factors between ischemic and hemorrhagic stroke[12]. Therefore,we developed a simple, convenient, and efficient model to predicting the 6-year risk of overall ,ischemic and hemorrhagic stroke among middle-aged and elderly Chinese adults.

3. Method

3.1. Study design

The China Health and Nutrition Survey(CHNS) was conducted by University of North Carolina at Chapel Hill, and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention in nine provinces (Heilongjiang, Liaoning, Shandong, Jiangsu, Henan, Hubei, Hunan, Guangxi, and Guizhou) to examine the status of economic , public resources, health and nutrition . A multistage, random cluster process was used to draw the samples in the Chinese population. The survey is ongoing nationwide study that started in 1989 and subsequently conducted in1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011and 2015.All participants provided written informed consent. Details about the study design are available elsewhere [13][14].Baseline data collection included demographic information, a medical history, standardized medical examination, laboratory tests, and anthropometric measurements.

3.2. Study population

For this study, data were drawn from the 2009-2015 CHNS cycles(n=12,178). After excluding participants who were younger than 45 years or older than 80 years at baseline (n = 6,148), persons who had history of stroke or myocardial infarction(n = 291), persons who lost to follow up(n = 1,770), and those without complete physical survey data or blood measure data at baseline(n = 845).As a result, 1,434

men and 1,690 women were available for analysis.

3.3. Data collection

History of diseases ,individual activities, lifestyle, health status, marriage and birth history were acquired through a individual questionnaire .Adults and children receive detailed physical examinations that included weight, height, arm and head circumference, mid-arm skinfold measurements , and blood pressure .Blood pressure was measured thrice by experienced physicians with the participant in the sitting position.The biomarker data collected in CHNS 2009 involves the release of 26 fasting blood measures on individuals aged 7 and older.Frozen serum samples were sent to a national central lab in Beijing (medical laboratory accreditation certificate ISO 15189:2007) for measurement of serum lipid levels.

3.4. Definitions

Age variable was divided into three groups as follows: 45-55 years old, 55–65 years old and 65–80 years old.Low-density lipoprotein cholesterol(LDL-C)was classified into three categories as follows:0-3mmol/L,3-4mmol/L and ≥4mmol/L; high-density lipoprotein cholesterol(HDL-C)was classified into two categories as follows:0-1mmol/L and ≥1mmol/L;total cholesterol(TC) was classified into three categories as follows:0-5mmol/L ,5-6mmol/L and ≥6mmol/L .Body mass index (BMI) was was classified into three categories as follows: 0-24kg/m^2, 24-28kg/m^2 and

≥28kg/m^2.Systolic blood pressure(SBP) and diastolic blood pressure(DBP) was defined as mean SBP and DBP of there test result. Hypertension was defined as blood pressure >160/90mmHg or taking antihypertensive drugs. Smoking status was classified into two categories as follows: never smoke and ever or current smoke;alcohol drinking status was divide into two groups:never drink and ever or current drink. Diabetes was identified by self reports of a history of diabetes diagnosis.

Incident stroke was was defined by a doctor's diagnosis or treatment history for stroke during the follow-up period (2009-2015).Cases were censored at the date of diagnosis of stroke or the final visit, whichever came first.

3.5. Statistical analysis

Results are presented number (percentage) for categorical variables. Cox proportional hazards models were used to explore the relation between baseline risk factors and overall stroke incidence.In addition,we also constructed the models with the outcomes of ischemic stroke and hemorrhagic stroke(unknown stroke participants were excluded). In multivariate analysis , we included confirmed or conceivable confounding factors for stroke as covariates, including age,gender (male or female) , BMI(kg/m^2), hypertension(no or yes) , LDL-C(mmol/L) , HDL-C(mmol/L),TC(mmol/L),drinking status(no or current/ever),smoking status (no or current/ever) , diabetes(no or yes) and site(rural or urban).

Risk factors found to be statistically significant(P<0.1) in multivariate analysis

were further evaluated using nomogram .A score was assigned to each risk factor in the nomogram in order that total points could be easily calculated to estimate the probability of stroke. To assess the accuracy of the nomogram,bootstraps with 1000 resamples were applied to both C-index and calibration curve .

All statistical analyses were carried out using R software version 3.6.1 (http://www.R-project.org).

4. Result

4.1. Baseline clinical characteristics of participants

The baseline clinical characteristics of the patients are listed in Table 1. During a median follow-up of 6 years ,2.98%(93) participants developed firstever overall stroke events, including 37 ischemic strokes, 25 hemorrhagic strokes, and 31 unspecified stroke events.There were 1001(32.04%) currently or ever smoking, 1066(34.12%) currently or ever drinking, 411(13.16%) had history of diabetes,and 2218(71.00%) live in rural areas.

4.2. multivariable analysis predicting stroke

Table 2 displays the HR [95% confidence interval (CI)] for stroke in different subgroups. Multivariate Cox regression analysis showed that age, gender(for female vs male,HR(95%CI)=1.58[0.92-2.72]), Hypertension(for no vs yes,HR(95%CI)=2.77[1.81-4.24]) and LDL-C (for 0-3 vs 3-4mmol/L,HR(95%CI)=1.75[0.99-3.09];for 0-3 vs ≥4mmol/L,HR(95

%CI)=1.86[0.79-4.37]) were independent risk predictors for developing overall stroke (p<0.1), which were further used to build a nomogram.

The significant risk factors (p<0.1) for ischemic stroke were age, gender (for female vs male,HR(95%CI)=2.00[0.86-4.66]), Hypertension(for no vs yes,HR(95%CI)=2.46[1.25-4.84]) and LDL-C(for 0-3 vs 3-4mmol/L,HR(95%CI)=4.38[1.61-11.90];for 0-3 vs ≥4mmol/L,HR(95%CI)=6.55[1.65-25.98]).The significant risk factors (p<0.1) for hemorrhagic stroke were age, gender (for female vs male,HR(95%CI)=3.48[1.20-10.10]), Hypertension(for no vs yes,HR(95%CI)=2.86[1.24-6.57]) and BMI(for 0-24 vs ≥28 kg/㎡, HR(95%CI)=3.28[1.01-10.61]).

4.3. Nomogram construction and validation

These independently associated risk factors were used to construct an overall stroke, ischemic stroke and hemorrhagic stroke risk estimation nomogram (Fig. 1,Fig. 2). The resulting model was internally validated using the bootstrap validation method. The nomogram demonstrated good accuracy in estimating the risk of overall stroke, ischemic stroke and hemorrhagic stroke ,with a bootstrap-corrected C-index of 0.75,0.77 and 0.81 respectively.The calibration plots for 6-year overall stroke-free, ischemic stroke-free and hemorrhagic stroke-free probability showed good agreement between the predicted possibility and the actual observation.These results indicated that the nomogram could accurately predict 6-year risk of overall stroke, ischemic stroke and hemorrhagic stroke among middle-aged and elderly Chinese population(Fig. 3,Fig. 4).

5. Discussion

We used data from the CHNS study, which is a nationally representative prospective cohort in nine provinces around China, enabling us to provide a general method to estimate the risk of stroke for middle-aged and elderly Chinese population.we created a practical nomogram based on age,gender,hypertension and LDL-C, to predict 6 year risk of overall stroke and ischemic stroke among middle-aged and elderly Chinese adults .Discrimination was supported by the C index values of 0.75 and 0.77, respectively.In addition,among three different models we established, hemorrhagic stroke model incorporating 4 variables (age, gender, hypertension and BMI)showed the best discrimination predictive ability (C -index =0.81). Calibration curves for 6-year overall ,ischemic and hemorrhagic stroke free probability demonstrating the good agreements between predicted and observed probability.

Hypertension is important and modifiable predictive risk factors for stroke[15][16].Results from INTERSTROKE study show that hypertension,using a definition : history of hypertension or blood pressure >160/90mm Hg, is highly important risk factor for stroke in developing countries and account for 35% of all stroke[17].Our result confirm the association between hypertension and risk of ischemic and hemorrhagic stroke and a more potent association for haemorrhagic stroke than for ischemic stroke.In China, the prevalence of hypertension rapidly increased in the past 30 years. However, awareness, treatment and control of hypertension declined or remained unchanged in China during 2000 to 2010 [4][18],

while they have increased significantly in the developed countries[19], these may affect the incidence of stroke.

The association between cholesterol and stroke is complex,with an increased risk for ischemic stroke with elevated TC[20]-[22],and a decreased risk for hemorrhagic stroke with elevated TC[23][24]. However,our study provides different information about the relationship of cholesterol and stroke risk. For risk of overall stroke and ischaemic stroke , we recorded association with elevated LDL-C , but we did not record significant associations with TC and HDL-C. Further,our analysis indicated that BMI exerted the impact on the onset of hemorrhagic stroke. A notable correlation between BMI and hemorrhagic stroke has also been reported [25][26], although conflicting study exists [27][28]. In addition,most studies consistently show that diabetes is a significant predict factor for ischemic stroke[29][30], and the risk of hemorrhagic stroke in diabetic patients does not seem to increase, and may even decrease[31].In accordance with these result, our result indicated that diabetes seem to be an important risk factor for ischemic stroke ,whereas no statistical significiant association was observed.

A nomogram is a excellent visual tool that is convenient, easy to use, and effective prognoses,which can easily estimate the risk of stroke by adding the score of each predicting variable to predicted probability .In addition,early treatment can be given to patients at high risk of early stroke.In contrast,these at low risk of stroke patients can be managed expectantly to prevent the potential treatment . Therefore, patients and doctors can use the nomogram to assist the level 1 prevention of stroke

and provide key information to reduce the incidence and burden of diseases. Furthermore,Non-professionals can learn about the nomogram through the Internet , TV or newspaper in a short period of time, thereby improving their understanding of the risk factors of diseases.

Our study has several limitations. First, we only screen residents older older then 45 and younger then 80 years old from nine provinces of China. Therefore, our current results cannot be generalized to all populations in China.Second,the long-term,>6 years,incidence of stroke for these risk factors are unknown .In future studies, we project to prospective follow-up participants, gather data about mortality and incidence of stroke in participants. Finally, the nomogram has not been externally validate.However, we used a 1000 bootstrap resampling strategy for internal validation and the nomogram showed good performance in terms of calibration and discrimination for predicting risk of stroke in middle-aged and elderly Chinese person .

6. Conclusion

In conclusion, we find that age,gender,hypertension and LDL-C were significant risk factors for overall and ischemic stroke,while age, gender, hypertension and BMI were stronger risk factors for hemorrhagic stroke in middle-aged and elderly Chinese adults .A nomogram that incorporates risk factors can be conveniently used to facilitate the individualized prediction of stroke.

Conflict of Interest

The authors declare that there is no conflict of interest.

Statement of Ethics

Data of this study was from the China Health and Nutrition Survey (CHNS). The Institutional Review Committees of the North Carolina at Chapel Hill, and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention approved the survey protocols and instruments and the process for obtaining informed consent for the survey. All participants (or their parents or guardians) agreed to participate in the survey and provided written informed consent.

Authors' contributions

Qi Yu conceptualized and designed the study, carried out the initial analyses, drafted the initial manuscript, and reviewed and revised the manuscript; Yuanzhe Wu, Qingdong Jin, Yanqing Chen,Qingying Lin and Xinru Liu critically reviewed and revised the manuscript; and all authors approved the final manuscript for submission.

Acknowledgement

This research uses data from CHNS . We thank the National Institute of

Nutrition and Food Safety, China Centre for Disease Control and Prevention, Carolina Population Centre, the University of North Carolina at Chapel Hill, and the Fogarty International Centre, National Institute of Health for financial support for the CHNS data collection and analysis files.

Funding Sources

None.

Patient and Public Involvement

No patient involved.

Reference

[1]Feigin VL, Norrving B, Mensah GA. Global Burden of Stroke. Circ Res. 2017;120(3):439-448. doi:10.1161/CIRCRESAHA.116.308413

[2]Writing Group Members, Mozaffarian D, Benjamin EJ, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association [published correction appears in Circulation. 2016 Apr 12;133(15):e599]. Circulation. 2016;133(4):e38-e360. doi:10.1161/CIR.0000000000000350

[3]Feigin VL, Forouzanfar MH, Krishnamurthi R, et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010 [published correction appears in Lancet. 2014 Jan 18;383(9913):218]. Lancet. 2014;383(9913):245-254. doi:10.1016/s0140-6736(13)61953-4

[4]Guan T, Ma J, Li M, et al. Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013. Neurology. 2017;89(1):53-61. doi:10.1212/WNL.0000000000004056

[5]Wang Z, Hu S, Sang S, Luo L, Yu C. Age-Period-Cohort Analysis of Stroke Mortality in China: Data From the Global Burden of Disease Study 2013. Stroke. 2017;48(2):271-275. doi:10.1161/STROKEAHA.116.015031

[6]Dufouil C, Beiser A, McLure LA, et al. Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. Circulation. 2017;135(12):1145-1159. doi:10.1161/CIRCULATIONAHA.115.021275

[7]Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study [published correction appears in Am J Epidemiol. 2004 Nov 1;160(9):927]. Am J Epidemiol. 2004;160(3):259-269. doi:10.1093/aje/kwh189

[8]Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ. 2013;346:f2573. Published 2013 May 2. doi:10.1136/bmj.f2573

[9]Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [published correction appears in Circulation. 2014 Jun 24;129(25 Suppl 2):S74-5]. Circulation. 2014;129(25 Suppl 2):S49-S73. doi:10.1161/01.cir.0000437741.48606.98

[10]Wang Y, Liu J, Wang W, et al. Lifetime risk of stroke in young-aged and middle-aged Chinese population: the Chinese Multi-Provincial Cohort Study. J Hypertens. 2016;34(12):2434-2440. doi:10.1097/HJH.0000000000001084

[11]Xing X, Yang X, Liu F, et al. Predicting 10-Year and Lifetime Stroke Risk in Chinese Population. Stroke. 2019;50(9):2371-2378. doi:10.1161/STROKEAHA.119.025553

[12]Boehme AK, Esenwa C, Elkind MS. Stroke Risk Factors, Genetics, and Prevention. Circ Res. 2017;120(3):472-495. doi:10.1161/CIRCRESAHA.116.308398

[13]Popkin BM, Du S, Zhai F, Zhang B. 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(6):1435-1440. doi:10.1093/ije/dyp322

[14]Yan S, Li J, Li S, et al. The expanding burden of cardiometabolic risk in China: the China Health and Nutrition Survey. Obes Rev. 2012;13(9):810-821. doi:10.1111/j.1467-789X.2012.01016.x

[15]Chobanian AV, Bakris GL, Black HR, et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003;42(6):1206-1252. doi:10.1161/01.HYP.0000107251.49515.c2

[16]Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies [published correction appears in Lancet. 2003 Mar 22;361(9362):1060]. Lancet. 2002;360(9349):1903-1913. doi:10.1016/s0140-6736(02)11911-8

[17]O'Donnell MJ, Xavier D, Liu L, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet. 2010;376(9735):112-123. doi:10.1016/S0140-6736(10)60834-3

[18]Wu S, Wu B, Liu M, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. Lancet Neurol. 2019;18(4):394-405. doi:10.1016/S1474-4422(18)30500-3

[19]Mills KT, Bundy JD, Kelly TN, et al. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. Circulation. 2016;134(6):441-450. doi:10.1161/CIRCULATIONAHA.115.018912

[20]Kurth T, Everett BM, Buring JE, Kase CS, Ridker PM, Gaziano JM. Lipid levels and the risk of ischemic stroke in women. Neurology. 2007;68(8):556-562. doi:10.1212/01.wnl.0000254472.41810.0d

[21]Soyama Y, Miura K, Morikawa Y, et al. High-density lipoprotein cholesterol and risk of stroke in Japanese men and women: the Oyabe Study. Stroke. 2003;34(4):863-868. doi:10.1161/01.STR.0000060869.34009.38

[22]Wannamethee SG, Shaper AG, Ebrahim S. HDL-Cholesterol, total cholesterol, and the risk of stroke in middle-aged British men. Stroke. 2000;31(8):1882-1888. doi:10.1161/01.str.31.8.1882

[23]Iso H, Jacobs DR Jr, Wentworth D, Neaton JD, Cohen JD. Serum cholesterol levels and six-year mortality from stroke in 350,977 men screened for the multiple risk factor intervention trial. N Engl J Med. 1989;320(14):904-910. doi:10.1056/NEJM198904063201405

[24]Zhang X, Patel A, Horibe H, Wu Z, Barzi F, Rodgers A, MacMahon S, Woodward M; Asia Pacific Cohort Studies Collaboration. Cholesterol, coronary heart disease, and stroke in the Asia Pacific region. Int J Epidemiol. 2003 Aug;32(4):563-72. doi: 10.1093/ije/dyg106. PMID: 12913030.

[25]Bazzano LA, Gu D, Whelton MR, et al. Body mass index and risk of stroke among Chinese men and women. Ann Neurol. 2010;67(1):11-20. doi:10.1002/ana.21950

[26]Song YM, Sung J, Davey Smith G, Ebrahim S. Body mass index and ischemic and hemorrhagic stroke: a prospective study in Korean men. Stroke. 2004;35(4):831-836. doi:10.1161/01.STR.0000119386.22691.1C

[27]Yonemoto K, Doi Y, Hata J, et al. Body mass index and stroke incidence in a Japanese community: the Hisayama study. Hypertens Res. 2011;34(2):274-279. doi:10.1038/hr.2010.220

[28]Strazzullo P, D'Elia L, Cairella G, Garbagnati F, Cappuccio FP, Scalfi L. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2 million participants. Stroke. 2010;41(5):e418-e426. doi:10.1161/STROKEAHA.109.576967

[29]Banerjee C, Moon YP, Paik MC, et al. Duration of diabetes and risk of ischemic stroke: the Northern Manhattan Study. Stroke. 2012;43(5):1212-1217. doi:10.1161/STROKEAHA.111.641381

[30]Sui X, Lavie CJ, Hooker SP, et al. A prospective study of fasting plasma glucose and risk of stroke in asymptomatic men. Mayo Clin Proc. 2011;86(11):1042-1049. doi:10.4065/mcp.2011.0267

[31]Roehmholdt ME, Palumbo PJ, Whisnant JP, Elveback LR. Transient ischemic attack and stroke in a community-based diabetic cohort. Mayo Clin Proc. 1983;58(1):56-58.

legend

Fig. 1 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population.Measurement: Age(year) and LDL-C(mmol/L).The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. 2 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population(A.Ischemic stroke, B.Hemorrhagic stroke).Measurement: Age(year), BMI(kg/m2)and LDL-C(mmol/L).The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. 3 Calibration curves for the nomogram(overall stroke).

Fig. 4 Calibration curves for the nomogram(A.Ischemic stroke, B.

```
Hemorrhagic stroke).
```

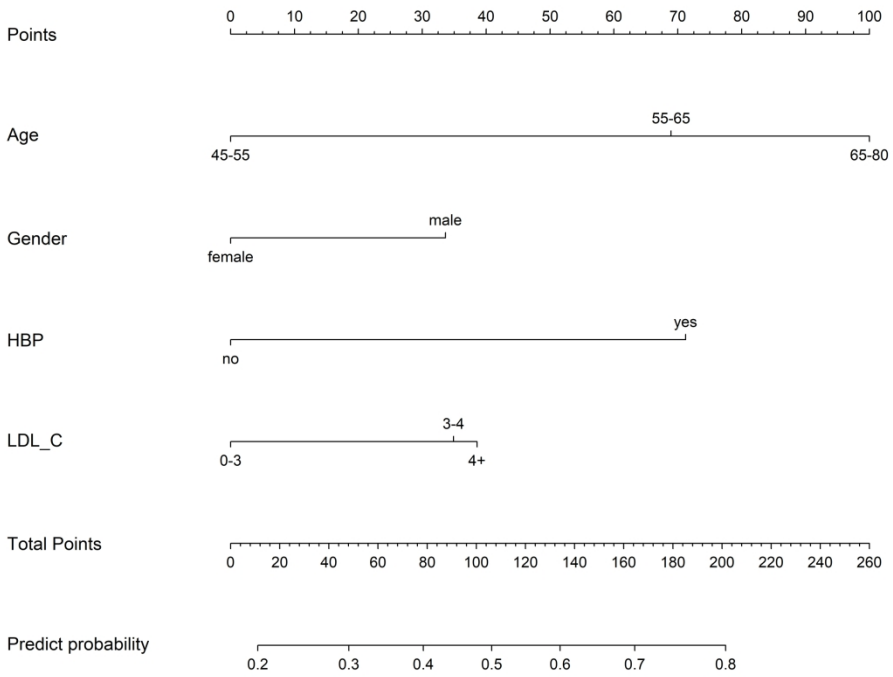
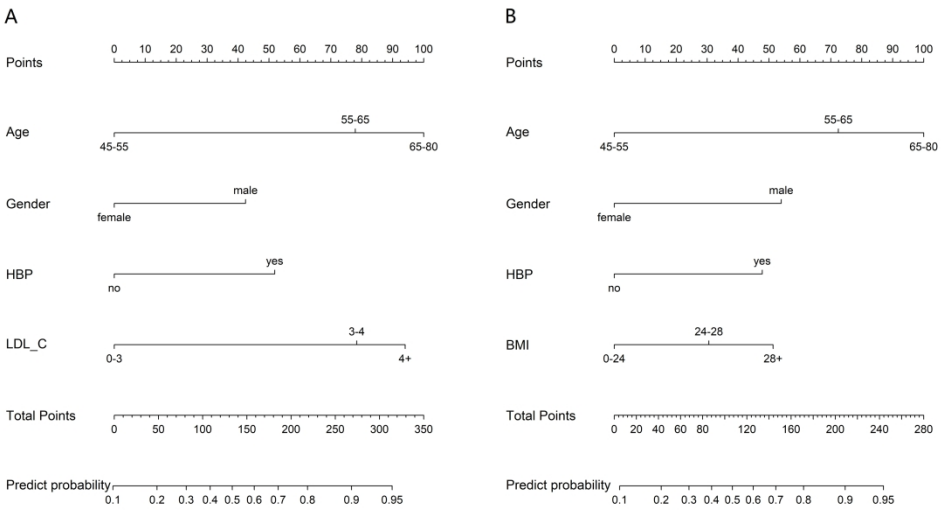


Fig. 1 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population.Measurement: Age(year) and LDL-C(mmol/L).The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

237x203mm (300 x 300 DPI)



Caption : Fig. 2 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population(A.Ischemic stroke, B.Hemorrhagic stroke).Measurement: Age(year), BMI(kg/m2)and LDL-C(mmol/L).The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

321x203mm (300 x 300 DPI)

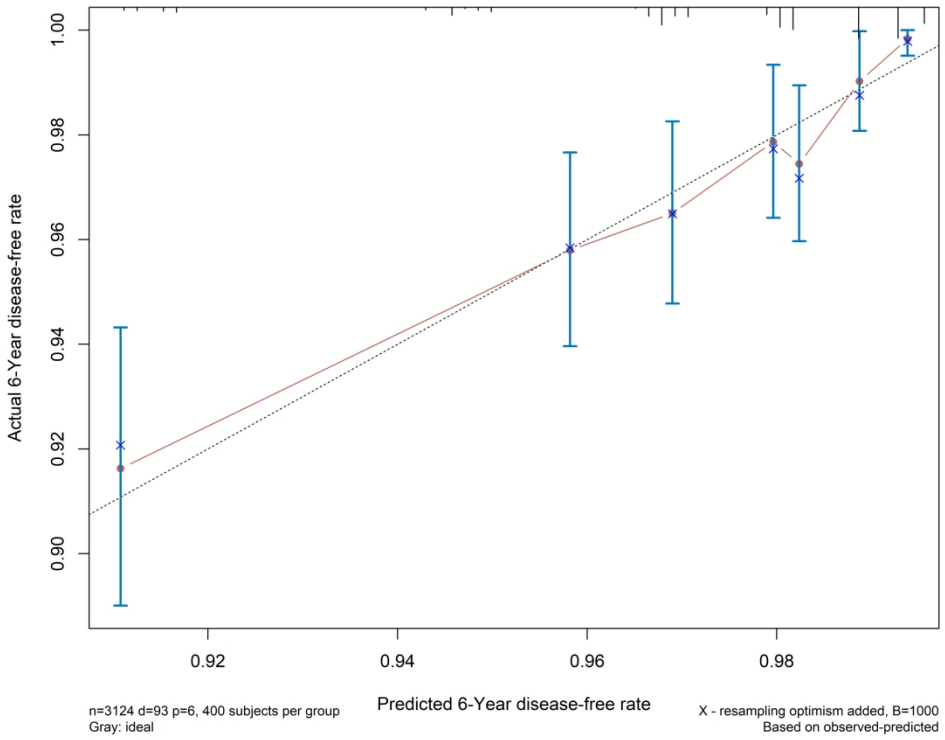


Fig. 3 Calibration curves for the nomogram(overall stroke).

237x203mm (300 x 300 DPI)

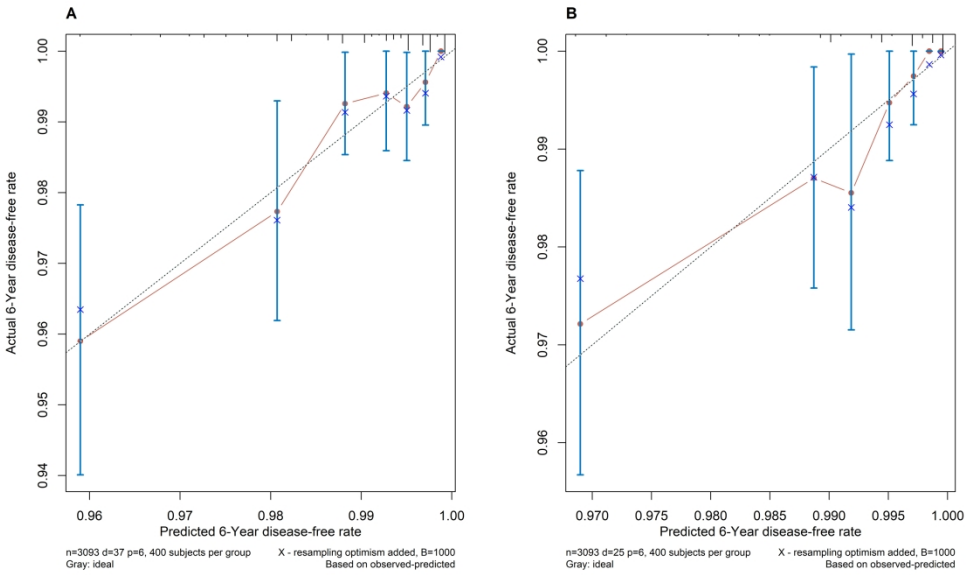


Fig. 4 Calibration curves for the nomogram(A.Ischemic stroke, B. Hemorrhagic stroke).

321x203mm (300 x 300 DPI)

BMJ Open

Development and internal validation of a multivariable prediction model for 6-year risk of stroke: a cohort study in middle-aged and elderly Chinese population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-048734.R1
Article Type:	Original research
Date Submitted by the Author:	11-May-2021
Complete List of Authors:	Yu, Qi; the First Hospital of Putian, Department of Scientific Research and Education Wu, Yuanzhe; the First Hospital of Putian, Department of Scientific Research and Education Jin, Qingdong; the First Hospital of Putian, Department of Neurosurgery Chen, Yanqing; the First Hospital of Putian, Department of Scientific Research and Education Lin, Qingying; the First Hospital of Putian, Department of Scientific Research and Education Liu, Xinru; the First Hospital of Putian, Department of Scientific Research and Education
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	Epidemiology < TROPICAL MEDICINE, Stroke < NEUROLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and internal validation of a multivariable prediction model for 6-year risk of stroke: a cohort study in middle-aged and elderly Chinese population

Qi Yu[1*]

Yuanzhe Wu[1]

Qingdong Jin[2]

Yanqing Chen[1]

Qingying Lin[1]

Xinru Liu[1]

1. Department of Scientific Research and Education, the First Hospital of Putian, Putain city, Fujian province, China

2. Department of Neurosurgery, the First Hospital of Putian, Putain city, Fujian province, China

Correspondence to:

Qi Yu, Department of Scientific Research and Education, the First Hospital of Putian, Putain city, Fujian province, China. Tel: +86 594 6923557, E-mail: madyuqi@126.com

Key words Stroke, Risk factor, Epidemiology

Word count：4990

Abstract

Objective To develop and internally validate a prediction model for 6-year risk of stroke and its primary subtypes in middle-aged and elderly Chinese population.

Design This is a retrospective cohort study from a prospectively collected database.

Participants We included a total 3,124 adults aged 45-80 years free of stroke or myocardial infarction at baseline in the 2009-2015 cohort of China Health and Nutrition Survey (CHNS).

Primary and secondary outcome measures The outcome of the prediction model was stroke. Investigated predictors were: age, gender, body mass index(BMI), low-density lipoprotein cholesterol(LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol(TC), hypertension(HBP), drinking status, smoking status, diabetes, and site. Stepwise multiple Cox regression was applied to identify independent predictors. A nomogram was constructed to predict 6-year risk of stroke based on the multiple analysis results. Bootstraps with 1000 resamples were applied to both C-index and calibration curve.

Result The overall incidence of overall stroke was 2.98%. Age, gender, HBP, and TC were found as significant risk predictors for overall stroke; age, gender, HBP, and LDL-C were found as significant risk predictors for ischemic stroke; age, gender, HBP, BMI, and HDL-C were found as significant risk predictors for hemorrhagic stroke. The nomogram was constructed using significant variables included in the

Model, with a C-index of 0.74(95% CI: 0.72–0.76), 0.74(95% CI: 0.71–0.77), and 0.81(95% CI: 0.78–0.84) for overall stroke, ischemic stroke, and hemorrhagic stroke model, respectively. The calibration curves demonstrated the good agreements between predicted and observed 6-year risk probability.

Conclusion Our nomogram could be convenient, easy to use, and effective prognoses for predicting 6-year risk of stroke in middle-aged and elderly Chinese population.

Key words Stroke, Risk factor, Epidemiology

Strengths and limitations of this study

This is the first study to develop a nomogram for predicting 6-year incidence rate of total stroke and its subtypes in a Chinese population.

Stepwise multiple regression and bootstrap internal validation were used to construct the models.

The model performance was also assessed by imputing missing values using an imputational regression model.

Because of lacking long-term follow-up data, the long-term result was uncertain.

This study has a relatively high sample size when compared with previous studies on stroke. However, for the development of a prediction model, the sample size is relatively low, and therefore we had to conduct a strong variable selection and model validation procedure.

1. Introduction

Stroke is the leading cause of death and is a major cause of permanent disability worldwide[1][2]. In 1990-2010, the incidence of stroke had decreased by 12% in high-income countries, because of effective strategies for preventing cerebrovascular risk factors and good health services in developed countries. On the contrary, the age-adjusted incidence of stroke had significantly increased by 12% in low-income and middle-income countries[3]. In China, the world's most populous country, the incidence of stroke increased by 8.3% among adults aged 40 years and older from 2002 to 2013[4][5].

Therefore, stroke prevention is essential for enhancing public health and reducing social burden in countries with heavy stroke burdens such as China. A prediction model for actively assessing stroke risk is required to ensure targeted strategies for stroke prevention and management in high-risk groups.

Previous studies that develop the stroke risk prediction models were initially developed from the western populations[6][7][8], thus application of these models to the general Chinese population is still questionable. In addition, many stroke risk prediction models were only built for overall stroke[9][10][11]. However, there are some notable differences in risk factors between ischemic and hemorrhagic stroke[12]. Therefore, we developed a simple, convenient, and efficient model to predict the 6-year risk of overall, ischemic and hemorrhagic stroke among middle-aged and elderly Chinese adults.

2. Method

2.1. Study design

The China Health and Nutrition Survey(CHNS) was conducted by the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention in nine provinces (Heilongjiang, Liaoning, Shandong, Jiangsu, Henan, Hubei, Hunan, Guangxi, and Guizhou) to examine the status of economic, public resources, health and nutrition. A multistage, random cluster process was used to draw the samples in the Chinese population. The survey was an ongoing nationwide study that started in 1989 and subsequently conducted in 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011 and 2015. All participants provided written informed consent. Details about the study design are available elsewhere [13][14]. Baseline data collection included demographic information, medical history, standardized medical examination, laboratory tests, and anthropometric measurements.

2.2. Study population

For this study, data was drawn from the 2009-2015 CHNS cycles(n=12,178). After excluding participants who were younger than 45 years or older than 80 years at baseline (n = 6,148), persons who had a history of stroke or myocardial infarction(n = 291), persons who lost to follow up(n = 1,770), and those without complete physical survey data or blood measure data at baseline(n = 845). As a result, 1,434 men and

1,690 women were available for analysis.

2.3. Data collection

History of diseases, individual activities, lifestyle, health status, marriage and birth history were acquired through an individual questionnaire. Adults and children received detailed physical examinations that include weight, height, arm and head circumference, mid-arm skinfold measurements, and blood pressure. Blood pressure was measured thrice by experienced physicians with the participant in the sitting position. The biomarker data collected in CHNS 2009 involves the release of 26 fasting blood measures on individuals aged 7 and older. Frozen serum samples were sent to a national central lab in Beijing for measurement of serum lipid levels.

2.4. Definitions

Systolic blood pressure(SBP) and diastolic blood pressure(DBP) were defined as mean SBP and DBP of three test results. Hypertension(HBP) was defined as blood pressure > 160/90mmHg or taking antihypertensive drugs. Smoking status was classified into three categories as follows: never, smoker, ever smoker and current smoker; alcohol drinking status was divided into two groups: never drinker and ever or current drinker. Diabetes was identified by self reports of a history of diabetes diagnosis.

Incident stroke was defined by a doctor's diagnosis or treatment history for stroke during the follow-up period (2009-2015). Cases were censored at the date of diagnosis of stroke or the final visit, whichever came first.

2.5. Statistical analysis

In this analysis, we included the following clinical candidate predictors: age(year), gender (male or female), body mass index(BMI, kg/m^2), low-density lipoprotein cholesterol(LDL-C, mmol/L), high-density lipoprotein cholesterol(HDL-C, mmol/L), total cholesterol(TC, mmol/L), HBP(no or yes), drinking status(no or ever/current), smoking status (no, ever or current), diabetes(no or yes), and site(urban or rural). Results were presented as mean±standard deviation for continuous variables, and number (percentage) for categorical variables. To determine statistical significance, analysis of variance (ANOVA) and chi-square tests were used for continuous variables and categorical variables according to sex-specific groups.

Cox proportional hazards models were used to explore the relationship between baseline risk factors and stroke incidence. In addition, we also constructed the models with the outcomes of ischemic stroke and hemorrhagic stroke(unknown stroke participants were excluded). The proportional hazards assumption of the Cox models was examined using Schoenfeld residuals. The significance of each risk factor in the cohort was assessed by univariate cox regression analysis for investigating the independent variable of incident stroke. Risk factors associated with stroke at a significant level were further evaluated using multivariate Cox proportional hazards regression analysis by the stepwise selection method($P<0.1$). The independent predictors related to incident stroke were used to build a nomogram. A score was assigned to each risk factor in the nomogram in order that total points could be easily calculated to estimate the probability of stroke. To assess the accuracy of the

nomogram, bootstraps with 1000 resamples were applied to both C-index and calibration curve. Moreover, we did stratified analyses to explore whether the association of risk factors with overall stroke varied across gender. We also assessed the models by imputing missing values using an imputational regression model (with the transcan function in R).

Internal validation is the process of determining internal validity. Bootstrap is an important internal verification method. Bootstrap samples are drawn with replacement from the original sample, reflecting the drawing of samples from an underlying population. Bootstrap samples are of the same size as the original sample. For bootstrap validation, a prediction model is developed in each bootstrap sample. This model is evaluated both in the bootstrap sample and in the original sample. The bootstrap is used to estimate the optimism: the decrease between performance in the bootstrap sample and performance in the original sample. This optimism is subsequently subtracted from the original estimate to obtain an optimism-corrected performance estimate[15][16].

We can use the bootstrap approach to any performance measure, including the C-index and calibration measures. The C-index is the most commonly used performance measure to depict the discriminative ability of generalized linear regression models. For a binary outcome, C-index is identical to the area under the receiver operating characteristic (ROC) curve. The ROC curve is a plot of the sensitivity (true positive rate) versus 1-specificity (false-positive rate) for continuous cutoffs for the probability of an outcome. The area under the curve can be interpreted

as the probability that a patient with the outcome is given a higher probability of the outcome by the model than a randomly chosen patient without the outcome[17]. Another important performance of a model is calibration: the consistency between observed results and predictions. For example, if we predict 70% probability of stroke event for a person, the observed frequency of stroke event should be 70 out of 100 such persons. Calibration curves describe the calibration of each model in terms of the agreement between the predicted risks of stroke and observed frequency of stroke. The y-axis represents the actual stroke rate. The x-axis represents the predicted stroke risk. The gray line represents a perfect prediction by an ideal model. The red solid line represents the performance of the nomogram, of which a closer fit to the gray line indicates a better prediction.

All statistical analyses were carried out using R software version 3.6.1 (http://www.R-project.org).

3. Result

3.1. Baseline clinical characteristics of participants

The baseline clinical characteristics of the patients are listed in Table 1. During a mean follow-up of 5.89 years, 2.98%(93) participants developed firstever stroke events, including 37 ischemic strokes, 25 hemorrhagic strokes, and 31 unspecified stroke events.

Table 1 Baseline clinical characteristics of participants

Characteristics	Total	Gender		
		Female	Male	P
Age(year)	58.80±8.64	58.83±8.72	58.75±8.54	0.880
Gender				
female	1690(54.10%)	NA	NA	
male	1434(45.90%)	NA	NA	
BMI(kg/m^2)	23.72±3.43	23.94±3.53	23.47±3.30	<.001
HBP				0.075
no	2283(73.08%)	1257(74.38%)	1026(71.55%)	
yes	841(26.92%)	433(25.62%)	408(28.45%)	
LDL-C(mmol/L)	3.14±1.00	3.24±0.98	3.02±1.01	<.001
HDL-C(mmol/L)	1.46±0.45	1.48±0.44	1.43±0.45	<.001
TC(mmol/L)	5.05±0.98	5.15±0.99	4.94±0.96	<.001
Smoking status				<.001
no	2123(67.96%)	1628(96.33%)	495(34.52%)	
ever	106(3.39%)	6(0.36%)	100(6.97%)	
current	895(28.65%)	56(3.31%)	839(58.51%)	
Drinking status				<.001
no	2058(65.88%)	1535(90.83%)	523(36.47%)	
ever/current	1066(34.12%)	155(9.17%)	911(63.53%)	
Diabetes				0.009
no	2713(86.84%)	1475(87.28%)	1238(86.33%)	
yes	411(13.16%)	215(12.72%)	196(13.67%)	
Site				0.992
urban	906(29.00%)	490(28.99%)	416(29.01%)	
rural	2218(71.00%)	1200(71.01%)	1018(70.99%)	

Stroke type				0.072
Ischemic stroke	37(39.78%)	15(37.50%)	22(41.51%)	
Hemorrhagic stroke	25(26.88%)	7(17.50%)	18(33.96%)	
Unknown	31(33.33%)	18(45.00%)	13(24.53%)	

3.2. Multivariable analysis predicting stroke

Variables identified as predictors of incident stroke are listed in Table 2. Univariate analysis indicated that age, gender, BMI, HBP, LDL-C, HDL-C, TC, drinking status, smoking status, and diabetes were significant risk factors for all types of stroke. These predictor variables were included in the stepwise multiple Cox regression models. The proportional hazards assumption of the Cox models were examined for three prediction models (overall stroke, ischemic stroke and hemorrhagic stroke) using Schoenfeld residuals, which showed no significant departure from proportionality over time ($p > 0.05$).

Table 3 displays the HR [95% confidence interval (CI)] for stroke in different subgroups. Stepwise multiple Cox regression analysis showed that age(for each 1 year increase,HR(95%CI)=1.07[1.04-1.09]), gender(for female vs male, HR(95%CI)=1.66[1.09-2.52]), HBP(for no vs yes, HR(95%CI)=2.63[1.73-3.98]) and TC(for each 1 mmol/L increase, HR(95%CI)=1.26[1.03-1.53]) were independent risk predictors for developing overall stroke, which were further used to build a nomogram. The significant risk factors for ischemic stroke were age(for each 1 year increase, HR(95%CI)=1.07[1.03-1.11]), gender (for female vs male, HR(95%CI)=1.85[0.96-3.58]), HBP(for no vs yes, HR(95%CI)=2.29[1.19-4.40]) and

LDL-C(for each 1 mmol/L increase, HR(95%CI)=1.27[1.07-1.51]). The significant risk factors for hemorrhagic stroke were age(for each 1 year increase, HR(95%CI)=1.08[1.03-1.13]), gender (for female vs male,HR(95%CI)=3.49[1.43-8.55]), HBP(for no vs yes,HR(95%CI)=2.72[1.18-6.27]), BMI(for each 1 kg/m^2 increase, HR(95%CI)=1.13[1.02-1.25]) and HDL-C(for each 1 mmol/L increase, HR(95%CI)=1.58[1.19-2.09]).

In stratified analysis, HBP, elevated HDL-C level, ever smoking, and ever or current drinking were associated with an increased risk of overall stroke for females than for males. However, in sensitivity analyses using imputed dataset, HDL-C was not identified to be associated with overall stroke in different gender (Supplementary material 1, Fig .S1, Fig .S2).

Table 2 Univariate cox regression analysis of stroke incidence

Covariate	Overall stroke		Ischemic stroke		Hemorrhagic stroke	
	HR(95%CI)	P Value	HR(95%CI)	P Value	HR(95%CI)	P Value
Age(year)	1.08[1.05-1.10]	<.001	1.08[1.04-1.12]	<.001	1.08[1.04-1.13]	<.001
Gender,female vs male	1.57[1.04-2.36]	<.001	1.73[0.90-3.34]	<.001	3.04[1.27-7.27]	<.001
HBP,no vs yes	3.35[2.23-5.05]	<.001	2.94[1.54-5.60]	<.001	4.18[1.88-9.29]	<.001
BMI(kg/m^2)	1.05[0.99-1.11]	<.001	1.03[0.94-1.12]	<.001	1.13[1.02-1.25]	<.001
LDL-C(mmol/L)	1.22[1.05-1.41]	<.001	1.31[1.09-1.57]	<.001	1.08[0.75-1.55]	<.001
HDL-C(mmol/L)	1.24[0.90-1.71]	<.001	1.14[0.62-2.09]	<.001	1.54[1.12-2.13]	<.001
TC(mmol/L)	1.28[1.06-1.56]	<.001	1.47[1.09-1.97]	<.001	1.12[0.76-1.65]	<.001
Smoking status,no vs ever	2.74[1.31-5.72]	<.001	2.48[0.75-8.22]	<.001	2.76[0.63-12.07]	<.001
Smoking status,no vs current	0.99[0.62-1.57]	<.001	0.85[0.40-1.82]	<.001	1.26[0.53-2.97]	<.001
Drinking status,no vs ever/current	1.22[0.80-1.85]	<.001	1.32[0.68-2.54]	<.001	1.52[0.69-3.34]	<.001
Diabetes,no vs yes	1.48[0.88-2.51]	<.001	2.14[1.01-4.53]	<.001	0.91[0.27-3.03]	<.001
Site,urban vs rural	0.95[0.61-1.47]	0.804	0.85[0.43-1.69]	0.644	1.29[0.52-3.24]	0.584

Table 3 Stepwise multiple cox regression analysis of stroke incidence(P<0.1)

Covariate	Overall stroke		Ischemic stroke		Hemorrhagic stroke	
	HR(95%CI)	P Value	HR(95%CI)	P Value	HR(95%CI)	P Value
Age(year)	1.07[1.04-1.09]	<.001	1.07[1.03-1.11]	<.001	1.08[1.03-1.13]	<.001
Gender,female vs male	1.66[1.09-2.52]	0.018	1.85[0.96-3.58]	0.068	3.49[1.43-8.55]	0.006
HBP,no vs yes	2.63[1.73-3.98]	<.001	2.29[1.19-4.40]	0.013	2.72[1.18-6.27]	0.019
BMI(kg/m^2)	NA	NA	NA	NA	1.13[1.02-1.25]	0.021
LDL-C(mmol/L)	NA	NA	1.27[1.07-1.51]	0.006	NA	NA
HDL-C(mmol/L)	NA	NA	NA	NA	1.58[1.19-2.09]	0.002
TC(mmol/L)	1.26[1.03-1.53]	0.024	NA	NA	NA	NA

3.3. Nomogram construction and validation

These independently associated risk factors were used to construct an overall stroke, ischemic stroke and hemorrhagic stroke risk estimation nomogram (Fig. 1, Fig. 2). Moreover, nomograms were also built using imputed dataset (Supplementary material 1, Fig. S3, Fig. S4). A score was assigned on the point scale for each subtype within these variables. By counting the total score and locating it on the total points scale, we could easily draw a straight line down to determine the estimated probability of stroke over 6 years.

The resulting model was internally validated using the bootstrap validation method. The nomogram demonstrated good accuracy in estimating the risk of overall stroke, ischemic stroke and hemorrhagic stroke, with an unadjusted C-index of 0.74(95% CI: 0.72–0.76), 0.74(95% CI: 0.71–0.77) and 0.81(95% CI: 0.78–0.84), respectively; a bootstrap-corrected C-index of 0.73, 0.72 and 0.78 respectively. The calibration plots for 6-year overall stroke-free, ischemic stroke-free and hemorrhagic stroke-free probability showed good agreement between the predicted possibility and the actual observation(Fig. 3, Fig. 4).

We also assessed the models by imputing missing values using an imputational regression model. Before data imputation, there were 3969 cases. HBP was missing in 28 cases(0.7%); BMI was missing in 152 cases(3%); LDL-C was missing in 397 (10%) cases; HDL-C was missing in 397 (10%) cases; TC was missing in 397 (10%) cases; smoking status was missing in 26 (0.7%) cases; drinking status was missing in 27 (0.7%) cases; diabetes was missing in 19 (0.5%) cases. Finally, the missing values

were imputed using a single conditional imputation method (with the transcan function in R). Our findings remained robust with imputed data, with an unadjusted C-index of 0.73(95% CI: 0.70–0.76), 0.70(95% CI: 0.68–0.72), and 0.83(95% CI: 0.81–0.85)for overall stroke, ischemic stroke and hemorrhagic stroke, respectively; a bootstrap-corrected C-index of 0.72, 0.68 and 0.81 for overall stroke, ischemic stroke and hemorrhagic stroke, respectively. Moreover, the calibration plots for 6-year overall stroke-free, ischemic stroke-free and hemorrhagic stroke-free probability showed good agreement between the predicted possibility and the actual observation (Supplementary material 1, Fig. S5, Fig. S6). These results indicated that the nomogram could accurately predict 6-year risk of overall stroke, ischemic stroke and hemorrhagic stroke in middle-aged and elderly Chinese population.

4. Discussion

We used data from the CHNS study, which is a nationally representative prospective cohort in nine provinces around China, enabling us to provide a general method to estimate the risk of stroke in middle-aged and elderly Chinese population. We created a practical nomogram based on age, gender, HBP, and TC to predict the risk of overall stroke; a nomogram based on age, gender, HBP, and LDL-C to predict the risk of ischemic stroke; a nomogram based on age, gender, HBP, BMI, and HDL-C to predict the risk of hemorrhagic stroke(Fig. 1, Fig. 2). Discrimination was supported by the unadjusted C-index of 0.74(95% CI: 0.72–0.76), 0.74(95% CI: 0.71–0.77) and 0.81(95% CI: 0.78–0.84), respectively; bootstrap-corrected C-index of 0.73,0.72 and 0.78 respectively. In addition, among the three different models we established, the hemorrhagic stroke model showed the best discrimination predictive ability (unadjusted C -index =0.81). Calibration curves for 6-year overall, ischemic and hemorrhagic stroke free probability demonstrating the good agreements between predicted and observed probability(Fig. 3, Fig. 4). In addition, the sensitivity analysis using imputed data did not change the results substantially (Supplementary material 1, Fig. S3, Fig. S4, Fig. S5, Fig. S6).

Many stroke risk prediction models aiming to predict the risk of stroke have been developed previously, for example, the new FSRP[6], the QStroke[7], and others[8]. The new FSRP represented the current status of stroke predictors in the United States and France[6]; however, its performance among Chinese residents has not been

evaluated. In China, Wang et al[9] developed an age- and sex-specific lifetime risk chart of stroke including 6 risk factors (blood pressure, non-HDL-C, HDL-C, BMI, diabetes and smoking), which allowed for stratification of lifetime risk of stroke. In addition, Chien et al[11]developed a model from a community cohort of 3513 Chinese participants aged ≥35 years old in Taiwan incorporating 6 risk factors including age, gender, BP, family history of stroke, atrial fibrillation, and diabetes (C-index of 0.77). However, these risk prediction models are only available for the total stroke risk. Rare stroke risk assessment tool is available for either ischemic stroke and hemorrhagic stroke in China. Therefore, compared with current existing models, we think that better understanding of the association of risk factor with subtype specific stroke and stroke model derived from the CHNS population would be more enlightening for stroke prevention among Chinese adults.

HBP is an important and modifiable predictive risk factor for stroke[18][19]. Results from the INTERSTROKE study show that HBP, using a definition: history of HBP or blood pressure >160/90mm Hg, was a highly important risk factor for stroke in developing countries and accounts for 35% of all stroke[20]. Our result confirmed the association between HBP and risk of ischemic and hemorrhagic stroke, and a more potent association for hemorrhagic stroke than for ischemic stroke(Table 3). In China, the prevalence of HBP rapidly increased in the past 30 years. However, awareness, treatment and control of HBP declined or remained unchanged in China from 2000 to 2010 [4][21], while they had increased significantly in the developed countries[22], these may affect the incidence of stroke.

The association between cholesterol level and stroke is complex, with an increased risk for ischemic stroke with elevated TC[23][24][25], and a decreased risk for hemorrhagic stroke with elevated TC[26][27]. LDL-C showed positive associations with ischemic stroke, and no association with hemorrhagic stroke[23]. HDL-C showed positive associations with hemorrhagic stroke in a recent meta-analysis[28], but not by a recent retrospective cohort study[23]. In our study(Table 3), the association with the risk of overall stroke, ischaemic stroke and hemorrhagic stroke was positive for TC, LDL-C and HDL-C, respectively. Further, our analysis indicated that BMI exerted an impact on the onset of hemorrhagic stroke. A notable correlation between BMI and hemorrhagic stroke has also been reported[29][30], although conflicting studies exists [31][32]. Moreover, our result found that the risk of stroke was higher for ever smokers than for current smokers without adjustment for covariates(Table 2). However, our research lacked detailed data about quantity or intensity of smoking, and the proportion of heavy smokers among the ever smokers might be higher than current smokers in our cohort.

We also found that HBP, elevated HDL-C level, ever smoking, ever or current drinking were more strongly associated with overall stroke in females than for males. However, in sensitivity analysis with imputed data, HDL-C was not identified to be associated or showed only weak relationships with different gender(Supplementary material 1, Fig. S1, Fig. S2). These results were similar to a previous observational study[33].

A nomogram is an excellent visual tool that is convenient, easy to use, and

effective prognoses, which can easily estimate the risk of stroke by adding the score of each predicting variable to predicted probability. In addition, early treatment can be given to patients at high risk of early stroke. In contrast, these at low risk of stroke patients can be managed expectantly to prevent the potential treatment. Therefore, patients and doctors can use the nomogram to assist the level 1 prevention of stroke and provide key information to reduce the incidence and burden of diseases. Furthermore, non-professionals can learn about the nomogram through the Internet, TV or newspaper in a short period of time, thereby improving their understanding of the risk factors of diseases.

Our study has several limitations. First, for the development of a prediction model, the sample size is relatively low. The prediction of risk factors for stroke in China still requires more data to improve the prediction model. Second, the long-term, >6 years, incidence of stroke for these risk factors are unknown. In future studies, we project to prospective follow-up participants, gather data about mortality and incidence of stroke in participants. Finally, the nomogram has not been externally validated. However, we used a 1000 bootstrap resampling strategy for internal validation and the nomogram showed good performance in terms of calibration and discrimination for predicting risk of stroke in middle-aged and elderly Chinese population.

5. Conclusion

In summary, the nomogram here developed can be conveniently used to facilitate the individualized prediction of 6-year risk of stroke in middle-aged and elderly Chinese population.

Conflict of Interest

The authors declare that there is no conflict of interest.

Statement of Ethics

Data of this study was from the China Health and Nutrition Survey (CHNS). The Institutional Review Committees of the North Carolina at Chapel Hill, and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention approved the survey protocols and instruments and the process for obtaining informed consent for the survey. All participants (or their parents or guardians) agreed to participate in the survey and provided written informed consent.

Authors' contributions

Qi Yu conceptualized and designed the study, carried out the initial analyses, drafted the initial manuscript, and reviewed and revised the manuscript; Yuanzhe Wu, Qingdong Jin, Yanqing Chen, Qingying Lin and Xinru Liu critically reviewed and revised the manuscript; and all authors approved the final manuscript for submission.

Acknowledgement

This research uses data from CHNS. We thank the National Institute of Nutrition and Food Safety, China Centre for Disease Control and Prevention, Carolina Population Centre, the University of North Carolina at Chapel Hill, and the Fogarty

International Centre, National Institute of Health for financial support for the CHNS data collection and analysis files.

Funding Sources

None.

Patient and Public Involvement

No patient involved.

Data availability

The datasets generated and/or analysed during the current study are publically available in the China Health and Nutrition Survey repository: http://www.cpc.unc.edu/projects/china/data/datasets.

Reference

[1]Feigin VL, Norrving B, Mensah GA. Global Burden of Stroke. Circ Res. 2017;120(3):439-448. doi:10.1161/CIRCRESAHA.116.308413

[2]Writing Group Members, Mozaffarian D, Benjamin EJ, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association [published correction appears in Circulation. 2016 Apr 12;133(15):e599]. Circulation. 2016;133(4):e38-e360. doi:10.1161/CIR.0000000000000350

[3]Feigin VL, Forouzanfar MH, Krishnamurthi R, et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010 [published correction appears in Lancet. 2014 Jan 18;383(9913):218]. Lancet. 2014;383(9913):245-254. doi:10.1016/s0140-6736(13)61953-4

[4]Guan T, Ma J, Li M, et al. Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013. Neurology. 2017;89(1):53-61. doi:10.1212/WNL.0000000000004056

[5]Wang Z, Hu S, Sang S, Luo L, Yu C. Age-Period-Cohort Analysis of Stroke Mortality in China: Data From the Global Burden of Disease Study 2013. Stroke. 2017;48(2):271-275. doi:10.1161/STROKEAHA.116.015031

[6]Dufouil C, Beiser A, McLure LA, et al. Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. Circulation. 2017;135(12):1145-1159. doi:10.1161/CIRCULATIONAHA.115.021275

[7]Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ. 2013;346:f2573. Published 2013 May 2. doi:10.1136/bmj.f2573

[8]Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study [published correction appears in Am J Epidemiol. 2004 Nov 1;160(9):927]. Am J Epidemiol. 2004;160(3):259-269. doi:10.1093/aje/kwh189

[9]Wang Y, Liu J, Wang W, et al. Lifetime risk of stroke in young-aged and middle-aged Chinese population: the Chinese Multi-Provincial Cohort Study. J Hypertens. 2016;34(12):2434-2440. doi:10.1097/HJH.0000000000001084

[10]Xing X, Yang X, Liu F, et al. Predicting 10-Year and Lifetime Stroke Risk in Chinese Population. Stroke. 2019;50(9):2371-2378. doi:10.1161/STROKEAHA.119.025553

[11]Chien KL, Su TC, Hsu HC, et al. Constructing the prediction model for the risk of stroke in a Chinese population: report from a cohort study in Taiwan. Stroke. 2010;41(9):1858-1864. doi:10.1161/STROKEAHA.110.586222

[12]Boehme AK, Esenwa C, Elkind MS. Stroke Risk Factors, Genetics, and Prevention. Circ Res. 2017;120(3):472-495. doi:10.1161/CIRCRESAHA.116.308398

[13]Popkin BM, Du S, Zhai F, Zhang B. 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(6):1435-1440. doi:10.1093/ije/dyp322

[14]Yan S, Li J, Li S, et al. The expanding burden of cardiometabolic risk in China: the China Health and Nutrition Survey. Obes Rev. 2012;13(9):810-821. doi:10.1111/j.1467-789X.2012.01016.x

[15]Efron B, Tibshirani R J. An introduction to the bootstrap. CRC press, 1994.

[16]Harrell F E . Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer, 2015.

[17]Hanley J A, McNeil B J. The meaning and use of the area under a receiver

operating characteristic (ROC) curve. Radiology, 1982, 143(1): 29-36.

[18]Chobanian AV, Bakris GL, Black HR, et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003;42(6):1206-1252. doi:10.1161/01.HYP.0000107251.49515.c2

[19]Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies [published correction appears in Lancet. 2003 Mar 22;361(9362):1060]. Lancet. 2002;360(9349):1903-1913. doi:10.1016/s0140-6736(02)11911-8

[20]O'Donnell MJ, Xavier D, Liu L, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet. 2010;376(9735):112-123. doi:10.1016/S0140-6736(10)60834-3

[21]Wu S, Wu B, Liu M, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. Lancet Neurol. 2019;18(4):394-405. doi:10.1016/S1474-4422(18)30500-3

[22]Mills KT, Bundy JD, Kelly TN, et al. Global Disparities of Hypertension

Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. Circulation. 2016;134(6):441-450. doi:10.1161/CIRCULATIONAHA.115.018912

[23]Gu X, Li Y, Chen S, et al. Association of Lipids With Ischemic and Hemorrhagic Stroke: A Prospective Cohort Study Among 267 500 Chinese. Stroke. 2019;50(12):3376-3384. doi:10.1161/STROKEAHA.119.026402

[24]Kurth T, Everett BM, Buring JE, Kase CS, Ridker PM, Gaziano JM. Lipid levels and the risk of ischemic stroke in women. Neurology. 2007;68(8):556-562. doi:10.1212/01.wnl.0000254472.41810.0d

[25]Wannamethee SG, Shaper AG, Ebrahim S. HDL-Cholesterol, total cholesterol, and the risk of stroke in middle-aged British men. Stroke. 2000;31(8):1882-1888. doi:10.1161/01.str.31.8.1882

[26]Iso H, Jacobs DR Jr, Wentworth D, Neaton JD, Cohen JD. Serum cholesterol levels and six-year mortality from stroke in 350,977 men screened for the multiple risk factor intervention trial. N Engl J Med. 1989;320(14):904-910. doi:10.1056/NEJM198904063201405

[27]Zhang X, Patel A, Horibe H, Wu Z, Barzi F, Rodgers A, MacMahon S,

Woodward M; Asia Pacific Cohort Studies Collaboration. Cholesterol, coronary heart disease, and stroke in the Asia Pacific region. Int J Epidemiol. 2003 Aug;32(4):563-72. doi: 10.1093/ije/dyg106. PMID: 12913030.

[28]Wang X, Dong Y, Qi X, Huang C, Hou L. Cholesterol levels and risk of hemorrhagic stroke: a systematic review and meta-analysis. Stroke. 2013;44(7):1833-1839. doi:10.1161/STROKEAHA.113.001326

[29]Bazzano LA, Gu D, Whelton MR, et al. Body mass index and risk of stroke among Chinese men and women. Ann Neurol. 2010;67(1):11-20. doi:10.1002/ana.21950

[30]Song YM, Sung J, Davey Smith G, Ebrahim S. Body mass index and ischemic and hemorrhagic stroke: a prospective study in Korean men. Stroke. 2004;35(4):831-836. doi:10.1161/01.STR.0000119386.22691.1C

[31]Yonemoto K, Doi Y, Hata J, et al. Body mass index and stroke incidence in a Japanese community: the Hisayama study. Hypertens Res. 2011;34(2):274-279. doi:10.1038/hr.2010.220

[32]Strazzullo P, D'Elia L, Cairella G, Garbagnati F, Cappuccio FP, Scalfi L. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2

million participants. Stroke. 2010;41(5):e418-e426.
doi:10.1161/STROKEAHA.109.576967

[33]Peters SAE, Carcel C, Millett ERC, Woodward M. Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study. Neurology. 2020;95(20):e2715-e2726.
doi:10.1212/WNL.0000000000010982

Legend

Fig. 1 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population.

Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. 2 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population(A. ischemic stroke, B. hemorrhagic stroke).

Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. 3 Calibration curves for the nomogram(overall stroke).

Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line

indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

Fig. 4 Calibration curves for the nomogram(A. Ischemic stroke, B. Hemorrhagic stroke).

Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

Fig. S1 Multivariate cox regression analysis of overall stroke incidence, by sex

Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L).

Fig. S2 Multivariate cox regression analysis of overall stroke incidence using imputed dataset, by sex

Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L).

Fig. S3 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population using imputed dataset.

Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and

TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. S4 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population using imputed dataset(A.ischemic stroke, B.hemorrhagic stroke). Measurement: Age(year), BMI(kg/m^2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

Fig. S5 Calibration curves for the nomogram using imputed dataset(overall stroke). Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

Fig. S6 Calibration curves for the nomogram using imputed dataset(A.ischemic stroke, B. hemorrhagic stroke).

Nomogram-predicted probability and observed frequency over 6 years for stroke

among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

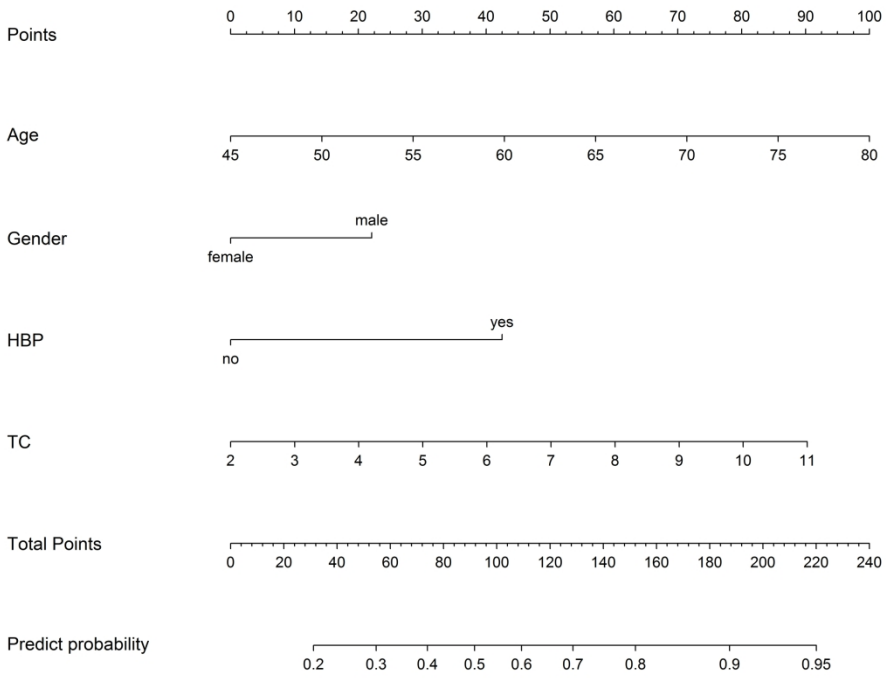


Fig. 1 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population.
Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

237x203mm (300 x 300 DPI)

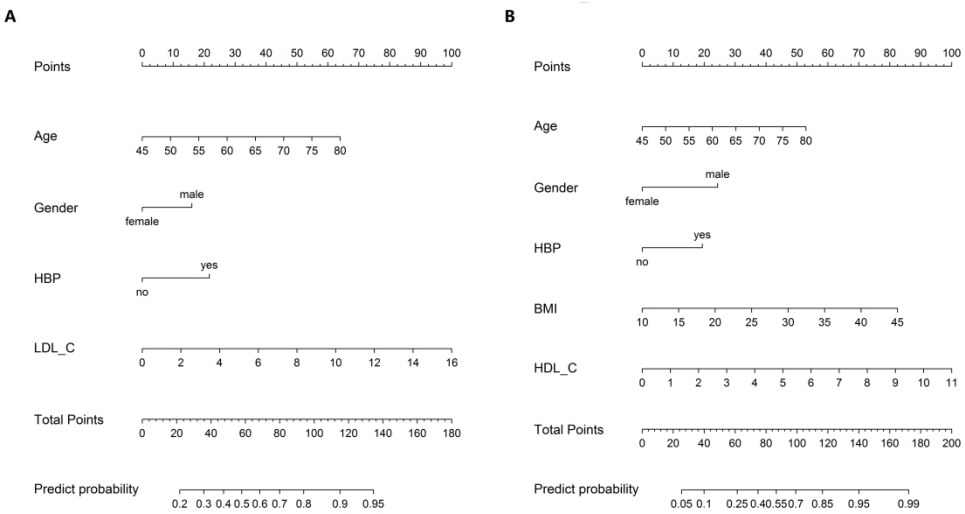


Fig. 2 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population(A. ischemic stroke, B. hemorrhagic stroke).
Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

321x203mm (300 x 300 DPI)

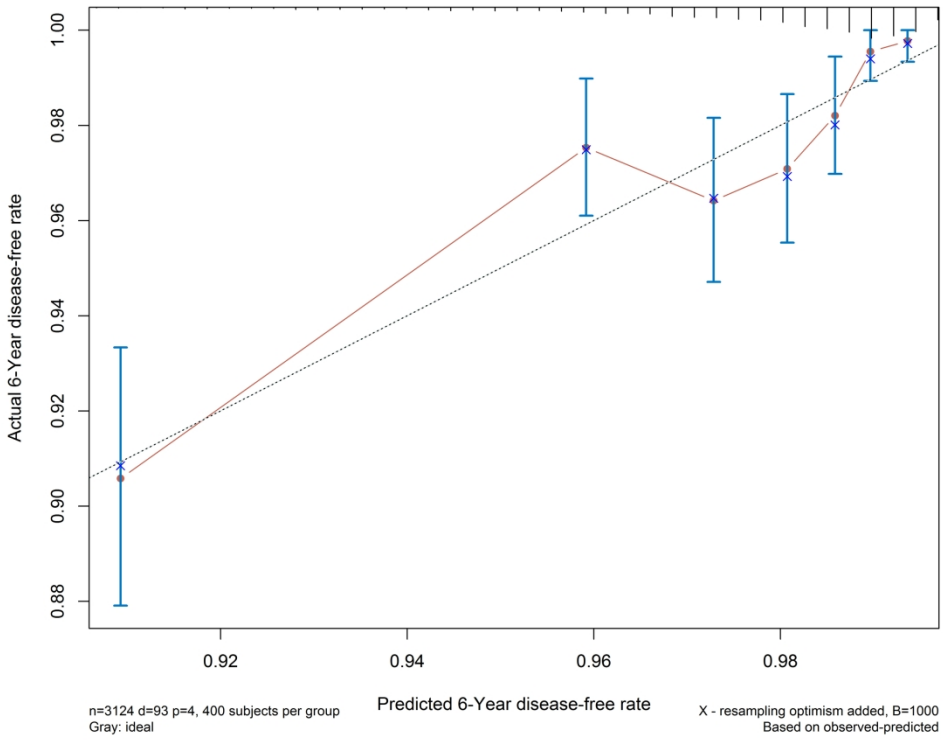


Fig. 3 Calibration curves for the nomogram(overall stroke).
Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

237x203mm (300 x 300 DPI)

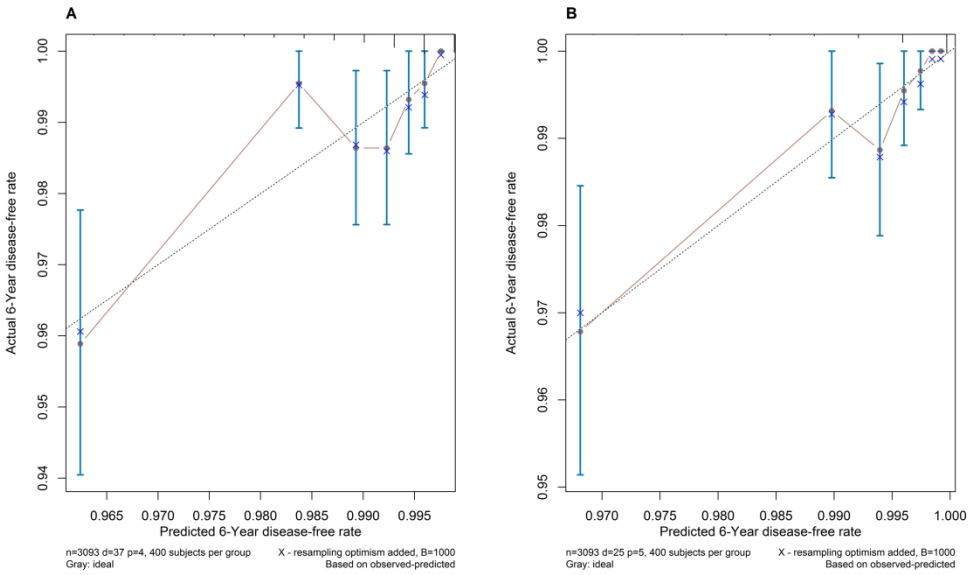


Fig. 4 Calibration curves for the nomogram(A. Ischemic stroke, B. Hemorrhagic stroke). Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

321x203mm (300 x 300 DPI)

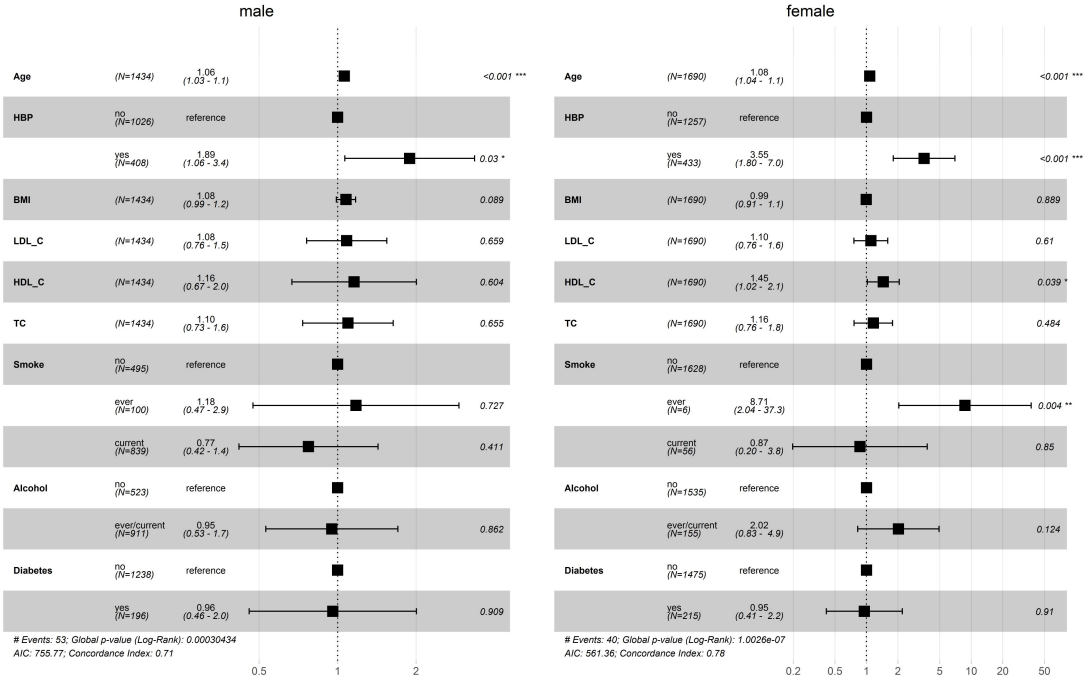


Fig. S1 Multivariate cox regression analysis of overall stroke incidence, by sex.

Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L).

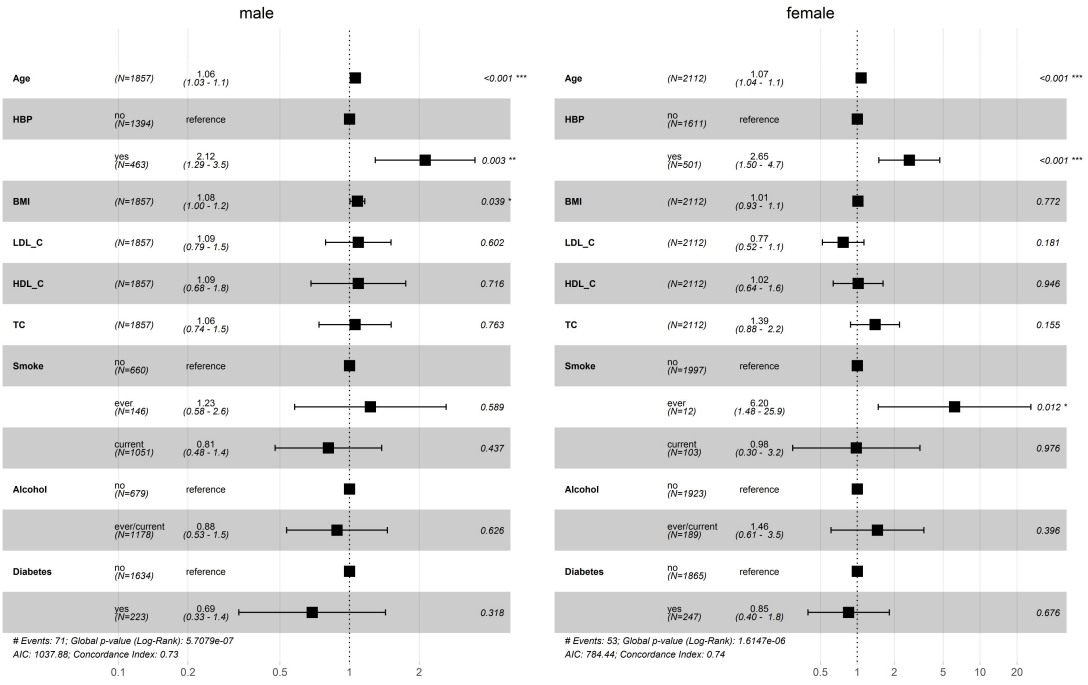


Fig. S2 Multivariate cox regression analysis of overall stroke incidence using imputed dataset, by sex.

Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L).

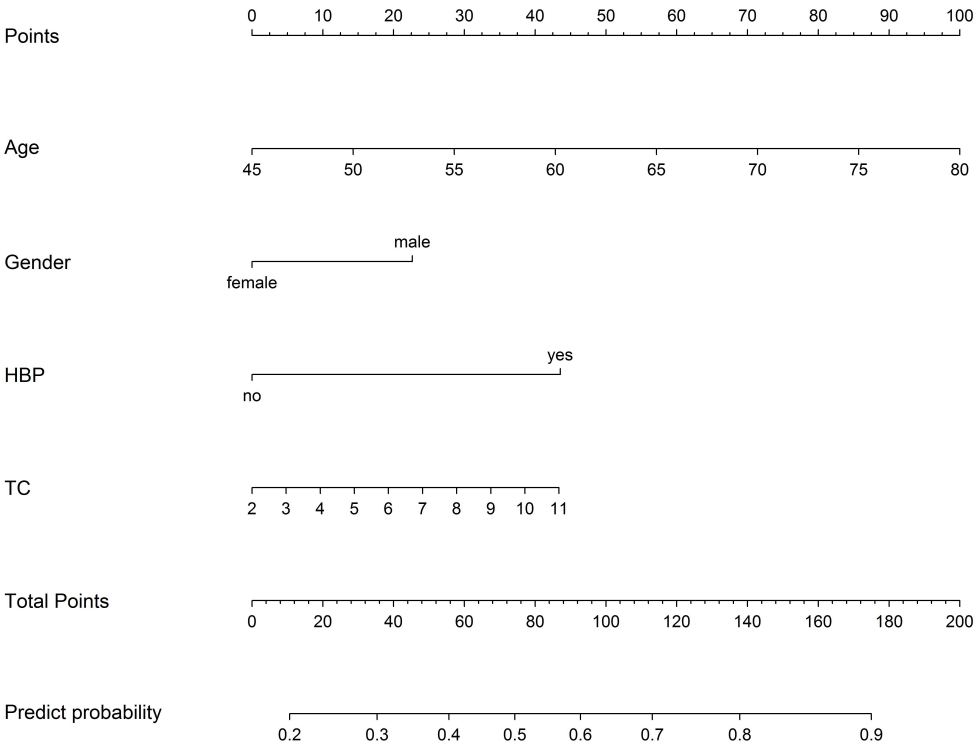


Fig. S3 Nomogram for predicting 6-year risk of overall stroke for middle-aged and elderly Chinese population using imputed dataset.

Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

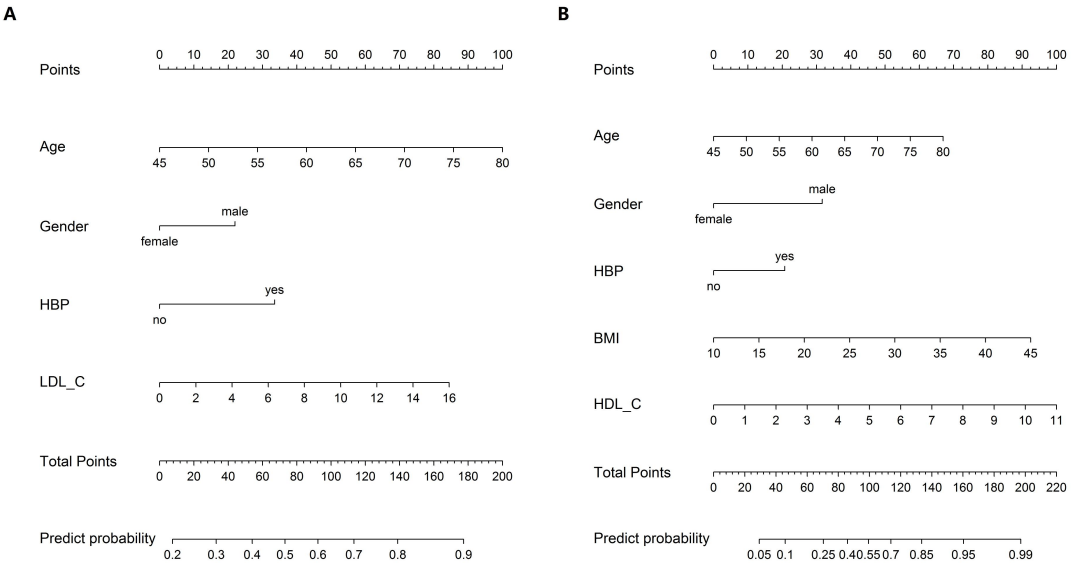


Fig. S4 Nomogram for predicting 6-year risk of stroke for middle-aged and elderly Chinese population using imputed dataset(A.ischemic stroke, B.hemorrhagic stroke).

Measurement: Age(year), BMI(kg/m2), LDL-C(mmol/L), HDL-C(mmol/L) and TC(mmol/L). The scores corresponding to each factor are listed on the "Points" axis. To estimate the 6-year probability of disease, calculate the sum of the scores of each factor and locate the sum on the "Total Points" axis, then read the probability on the "Predict probability" axis.

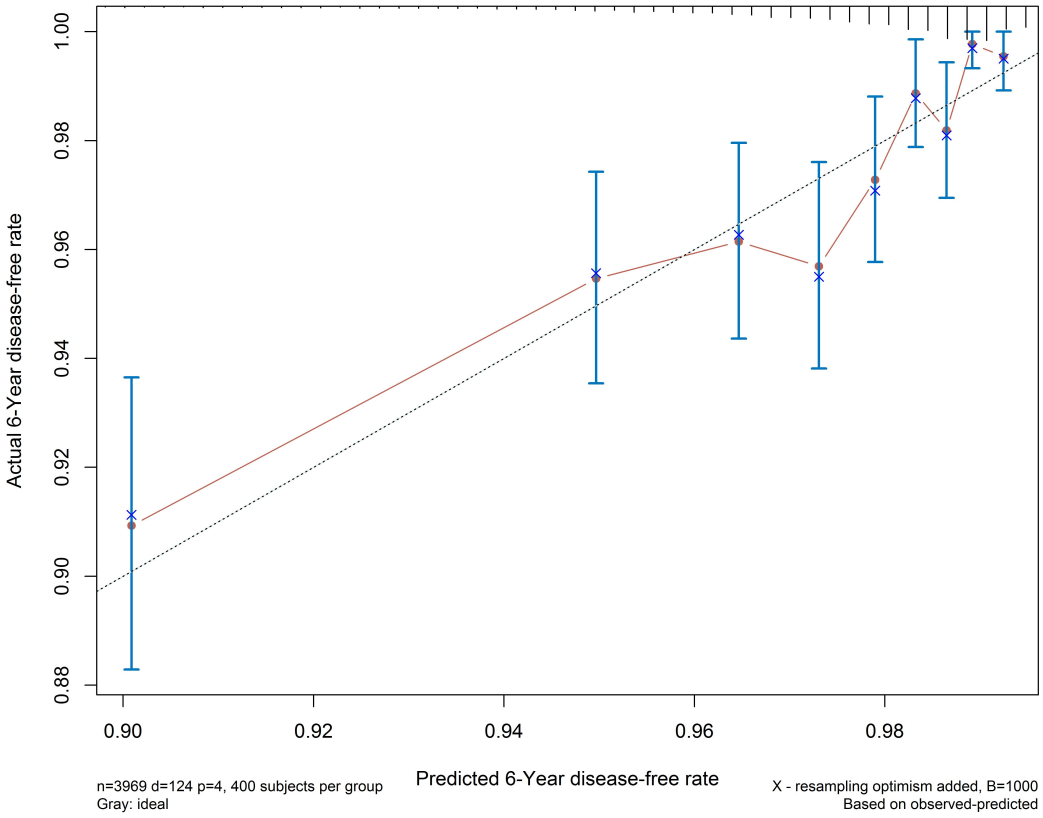


Fig. S5 Calibration curves for the nomogram using imputed dataset(overall stroke).

Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

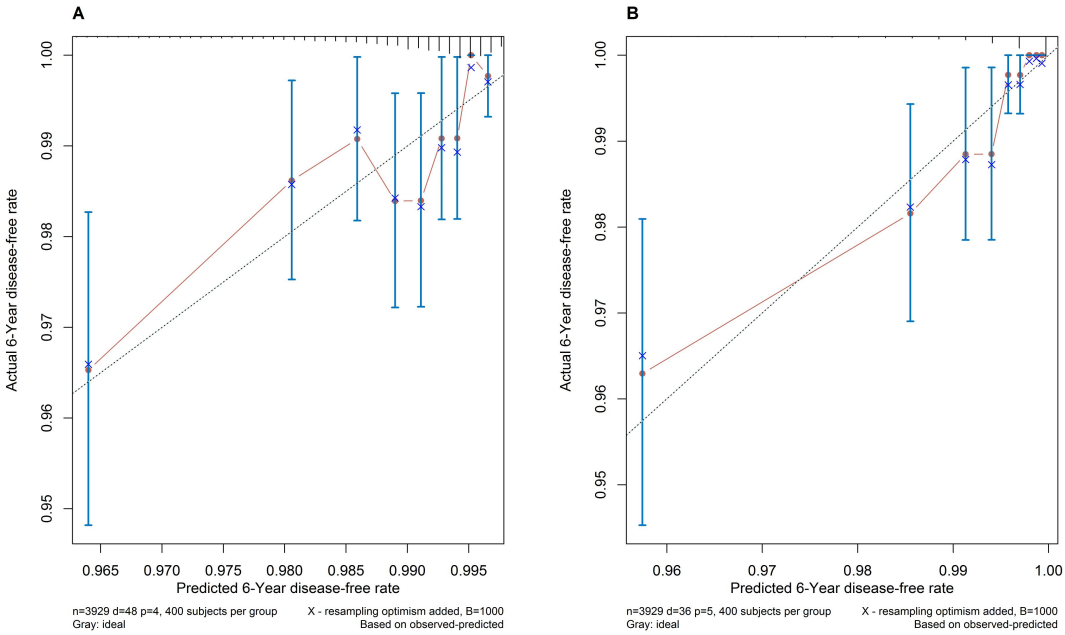


Fig. S6 Calibration curves for the nomogram using imputed dataset(A.ischemic stroke, B. hemorrhagic stroke).

Nomogram-predicted probability and observed frequency over 6 years for stroke among participants were plotted in the x- and y-axis, respectively. The gray line indicates the ideal plot for the calibration curve, where the nomogram predicted probabilities perfectly match the observed probabilities in all subgroups.

TRIPOD Checklist: Prediction Model Development and Validation

Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Item			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	D;V	Describe eligibility criteria for participants.	6
	5c	D;V	Give details of treatments received, if relevant.	-
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	-
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	-
Sample size	8	D;V	Explain how the study size was arrived at.	6
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	7
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8
	10c	V	For validation, describe how the predictions were calculated.	8
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	-
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	11
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	11
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model developme	14a	D	Specify the number of participants and outcome events in each analysis.	11

TRIPOD Checklist: Prediction Model Development and Validation

nt	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	12
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	12
	15b	D	Explain how to the use the prediction model.	16
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	16
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	18
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	16
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	-

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.