Stroke
JOURNAL OF THE AMERICAN HEART ASSOCIATION
American Stroke Association
A Division of American Heart Association

Constructing the Prediction Model for the Risk of Stroke in a Chinese Population. Report From a Cohort Study in Taiwan

Kuo-Liong Chien, Ta-Chen Su, Hsiu-Ching Hsu, Wei-Tien Chang, Pei-Chun Chen, Fung-Chang Sung, Ming-Fong Chen and Yuan-Teh Lee

Stroke published online Jul 29, 2010;
DOI: 10.1161/STROKEAHA.110.586222
Stroke is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 0039-2499. Online ISSN: 1524-4628

Constructing the Prediction Model for the Risk of Stroke in a Chinese Population

Report From a Cohort Study in Taiwan

Kuo-Liong Chien, MD, PhD; Ta-Chen Su, MD, PhD; Hsiu-Ching Hsu, PhD; Wei-Tien Chang, MD, PhD; Pei-Chun Chen, PhD; Fung-Chang Sung, PhD; Ming-Fong Chen, MD, PhD; Yuan-Teh Lee, MD, PhD

Background and Purpose—Prediction rules for the risk of stroke have been proposed. However, most studies were conducted with whites or for secondary prevention, and it is not clear whether these models apply to the Chinese population. The purpose of this study was to construct a simple points-based clinical model for predicting incident stroke among Chinese adults in Taiwan.

Methods—We estimated the 10-year risk of stroke in a cohort study of middle-aged and elderly participants who were free from stroke at baseline. Multivariate Cox model-derived coefficients were used to construct the simple points-based clinical and biochemical model and the prediction measures using the area under the receive operating characteristic curve, net reclassification improvement, and integrated discrimination improvement statistics were applied.

Results—Of the 3513 participants without stroke at baseline, 240 incident cases of stroke were documented for a median 15.9-year follow-up. Age (8 points), gender (1 point), systolic blood pressure (3 points), diastolic blood pressure (2 points), family history of stroke (1 point), atrial fibrillation (3 points), and diabetes (1 point) were found to significantly predict stroke events. The estimated area under the receive operating characteristic curve for this clinical points-based model was 0.772 (95% CI, 0.744 to 0.799). The discrimination ability of this clinical model was similar to the coefficients-based models and better than available stroke models.

Conclusions—We have constructed a model for predicting 15-year incidence of stroke in Chinese adults and this model may be useful in identifying individuals at high risk of stroke. **(*Stroke*. 2010;41:00-00.)**

Key Words: cohort study ■ prediction model ■ stroke

Prevention of stroke and its associated risk factors has been an important public health priority worldwide. Recent clinical trials demonstrated that identifying high-risk individuals and treating primary prevention can substantially prevent the risk of a stroke event,[1,2] providing a rationale for the identification of a high-risk population as well as implementing lifestyle intervention strategies to prevent stroke. The prediction models for the risk of stroke have been helpful to guide screening and interventions and to predict stroke event (Supplemental Table I; available at http://stroke.ahajournals.org). However, participants in the previous prediction models were recruited among hospital-based patients under specific treatment for conditions such as Type 2 diabetes,[3,4] hypertension,[5,6] and atrial fibrillation[7] or from the cross-sectional survey[6,8,9] and those with a history of cardiovascular diseases.[4,10–12] In addition, the associated covariates included uncommon biochemical measures such as proinsulin, lipoprotein(a),[13] or spot urine albumin–creatinine ratio,[4] and the outcomes were only ascertained by self-validation.[14] Moreover, the performance measures using the area under the receiver operating characteristics curve (AUC) were only available for some studies[4,14,15] and were limited to whites,[12,14,16] and the performance was not compared with other models. Furthermore, constructing a stroke prediction model was motivated by a high stroke risk and hypertension prevalence among the Asia Pacific countries.[17] Therefore, we constructed clinical as well as biochemical prediction models for stroke risk in a prospective community-based cohort study. We aimed to further clarify the performance measures of these models, which provided a points-based chart and nomogram from clinical measures, and compare whether these models outperform the Framingham[16] and the Prospective Cardiovascular Munster (PROCAM)[14] stroke models.

Methods

Study Design and Study Participants

Details of this cohort study have been published previously.[18–20] Briefly, the Chin-Shan Community Cohort Study began in 1990 by recruiting 1703 men and 1899 women of Chinese ethnicity aged ≥35 years from the town of Chin-Shan, 30 km north of metropolitan Taipei, Taiwan. Information about lifestyle and medical conditions and anthropometric measures was assessed by interview question-

Received April 1, 2010; accepted May 5, 2010.

From the Institute of Preventive Medicine (K.-L.C.), College of Public Health, National Taiwan University, Taiwan; the Departments of Internal Medicine (T.-C.S., H.-C.H., P.-C.C., M.-F.C., Y.-T.L.) and Emergency Medicine (W.-T.C.), National Taiwan University Hospital, Taiwan; and China Medical University Hospital (F.C.S., Y.-T.L.), Taichung, Taiwan.

Correspondence to Yuan-Teh Lee, China Medical University Hospital, Taichung, Taiwan, 400. Email ytlee@ntu.edu.tw

Stroke **is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.110.586222**

naires and physical examinations in 2-year cycles for the initial 6 years; the validity and reliability of the collected data and measurements have been reported in details elsewhere.[20,21]

Measurement of Clinical, Lifestyle, and Biochemical Markers

The procedures for clinical and biochemical measures have been reported elsewhere.[20–22] In brief, blood pressure was measured twice in the right arm using a mercury sphygmomanometer with the subject seated comfortably and arms supported and positioned at the level of the heart. The average of the blood pressure measurements was used as described previously.[23–25] Family history of stroke was coded as a stroke event among first relatives. Smoking habit was defined by current smoking status. Drinking history was defined as a binary variable using the frequency of drinking habits.

All venous blood samples drawn after a 12-hour overnight fast were immediately refrigerated and transported within 6 hours to the National Taiwan University Hospital. Serum samples were then stored at −70°C before batch assay for levels of total cholesterol, triglyceride, and high-density lipoprotein cholesterol. Standard enzymatic tests for serum cholesterol were used (Merck 14354 and 14366, respectively). Blood samples for glucose analysis were drawn into glass test tubes each containing 80 mol/L fluoride/oxalate reagent after centrifugation by 1500 *g* for 10 minutes; glucose levels were measured on supernatant by enzymatic assay (Merck 3389 commercial kit) in an Eppendorf 5060 autoanalyzer. The peripheral blood cell analysis was measured using a blood cell counter (Sysmex Cell Counter NE-8000; TOA Medical Electronics Co Ltd, Kobe, Japan).

Follow-Up Strategy and Outcome Ascertainment

Procedures for our documentation of incident stroke have been previously described and validated.[18,19,26] Participants with a baseline history of stroke events (n=89) were excluded from this study. Incident stroke cases were ascertained according to the following criteria: a sudden neurological symptom of vascular origin that lasted >24 hours with supporting evidence from the image study; fatal stroke cases were included. Transient ischemic attacks were not included in this study. Deaths were identified from official certificate documents and further verified by house-to-house visits. The cases were confirmed by cardiologists and neurologists; the National Taiwan University Hospital Committee Review Board approved the study protocol.

Statistical Analysis

We used the multivariate Cox proportional hazards model to establish a parsimonious model for predicting risk of stroke, and specified clinical and biochemical variables in the model. This clinical model included 7 significant predictors: age, gender, systolic and diastolic blood pressure, family history of stroke, diagnosis of atrial fibrillation, and diabetes mellitus (fasting glucose ≥126 mg/dL or with hypoglycemic medication history) derived from a multivariate Cox model in which we specified the stepwise method for best subset selection by entering in or removing from the model using a significance level of 0.05. The proportionality assumption was not rejected in the model. We did not include smoking, alcohol intake, or body mass index in the model because of their nonsignificance in the multivariate model. Because smoking, alcohol intake, and history of coronary heart disease are clinically important predictors for the risk of stroke,[5,11,14,15] we examined the incremental predictive value of adding these variables to this model; however, the likelihood ratio test suggested that adding these variables into the model did not improve prediction beyond the clinical model. Thus, our final clinical model did not include smoking, alcohol intake, or history of coronary heart disease. We also tested the goodness of fit of the model using the Hosmer and Lemeshow test[27] and the results showed adequate fit. In addition, we constructed another biochemical model by adding total cholesterol, white blood cell, and fasting glucose level in this clinical model. To check the impact of competing risks of stroke and nonstroke event in this cohort, we followed the methods provided by Lau et al[28] for the cause-specific hazards estimate and the results were similar to the original Cox model.

We constructed the categorization point model according to the clinical and biochemical covariates using the methods suggested by Sullivan and colleagues.[29] First, we categorized each continuous variable into meaningful groups and assigned the point values according to the coefficients from the Cox model; then we summed up the total points by adding 6 variables together for each individual. Second, we determined the probability of diabetes during the following 10 years based on the following formula, $\text{Risk}=1-S_0(t)^{\exp(\Sigma\beta X-\Sigma\beta\bar{X})}$, where $S_0(t)$ is the average survival rate at time t (eg, $t=10$ years) at the mean values of the risk factor, βs are the Cox regression coefficients, Xs are the individual's values on the variables, and X-bar is the means or proportions of the variables. Third, we conducted the internal validation of the prediction model and obtained a bias corrected estimate of the AUC using a 5-fold cross-validation procedure.[30] We randomly split the data into 5 equal parts. For k=1, …, to 5, we used the k^{th} part as the validation data set and the remaining 4 parts as the training data set. To adjust for overoptimism in model fitting, we performed variable selection and coefficient estimates for each iteration, and the coefficients were then fixed when tested on the validation data set. Because the results were consistent between validation and all data sets, we reported the estimated coefficients and points systems from the data set. Fourth, we compared the performance of the proposed prediction models with 2 available stroke models from the Framingham[16] and PROCAM[14] studies (the coefficients assigned in Supplemental Table II). In addition, we provided several additional statistics, including integrated discrimination improvement (IDI) and net reclassification improvement (NRI)[31] for the comparison between clinical and biochemical models because the AUC value is not the best discriminatory statistics for prediction power.[31–33] The details are described in the Supplemental Materials. Finally, we plotted the Bland-Altman plot for the differences and the averages of predicted risk from the clinical and biochemical models to evaluate the comparability of these 2 models.[34,35] In brief, the slope of this plot is the comparison of predicted risk: when the slope is zero, the 2 measures have equal discrimination. We also provided the nomogram for the clinical model in predicting the risk of stroke according to the Harrell algorithm[36] implemented in R software.

All statistical tests were 2-sided with a Type I error of 0.05, and probability values of <0.05 were considered statistically significant. Analyses were performed with SAS Version 9.1 (SAS Institute, Cary, NC), Stata Version 9.1 (Stata Corporation, College Station, Texas), and R Version 2.9.0 (The R Foundation for Statistical Computing).

Results

For the cohort follow-up from 1990 to the end of 2007 (a total of 49 281 person-years, median 15.9 years, interquartile range: 12.8 to 16.9 years), we documented 240 incident cases of stroke (including 189 cases of ischemic and unclassified type and 51 cases of hemorrhagic type). The means and proportions of statistically significant clinical and biochemical measures in the study participants and the Cox multivariate results that included age, gender, systolic and diastolic blood pressure, family history, atrial fibrillation, and diabetes mellitus are shown in Table 1. Biochemical measures, including total cholesterol, white blood cell count, and fasting glucose, were also included in the biochemical model.

We developed a simple points system to estimate the stroke risk using the baseline survival function at 10 years and the coefficients of the clinical model (Table 2). Depending on age, up to 8 points were assigned; gender, 1 point; systolic blood pressure, 3 points; diastolic blood pressure, 2 points; family history of stroke, 1 point; atrial fibrillation, 3 points; and diabetes, 1 points. This approach allowed manual esti-

Table 1. Basic Characteristics and Estimated Coefficient, SE, Relative Risk, and 95% CI for the Clinical and Biochemical Prediction Models in the Study Participants (n=3513)

Covariates	Mean or Proportion	Estimated Coefficient	SE	RR	95% CI		P
Clinical model							
Age, +1 year	54.6	0.071	0.007	1.07	1.06	1.09	<0.0001
Gender, 1 for men and 2 for women	1.53	−0.430	0.133	0.65	0.50	0.85	0.001
Systolic BP, mm Hg	125.1	0.016	0.004	1.02	1.01	1.03	<0.0001
Diastolic BP, mm Hg	77.0	0.017	0.008	1.02	1.00	1.03	0.031
Family history of stroke, 1 for yes, 0 for no	0.20	0.463	0.146	1.59	1.19	2.11	0.002
Atrial fibrillation, 1 for yes, 0 for no	0.01	1.134	0.317	3.11	1.67	5.78	0.0003
Diabetes mellitus, 1 for yes, 0 for no	0.13	0.376	0.160	1.46	1.07	1.99	0.019
Biochemical model							
Age, +1 year	54.6	0.073	0.007	1.08	1.06	1.09	<0.0001
Gender, 1 for men and 2 for women	1.53	−0.458	0.136	0.63	0.48	0.83	0.001
Systolic BP, mm Hg	125.1	0.016	0.004	1.02	1.01	1.02	0.000
Diastolic BP, mm Hg	77.0	0.017	0.008	1.02	1.00	1.03	0.037
Family history of stroke, 1 for yes, 0 for no	0.20	0.457	0.146	1.58	1.19	2.10	0.002
Atrial fibrillation, 1 for yes, 0 for no	0.01	1.247	0.318	3.48	1.87	6.49	<0.0001
Total cholesterol, mg/dL	197.8	0.003	0.001	1.00	1.00	1.01	0.029
White blood cell, 10^3/cc	6.3	0.081	0.040	1.08	1.00	1.17	0.042
Fasting glucose, mg/dL	109.9	0.005	0.001	1.01	1.00	1.01	0.0002

Baseline survival function at 10 years, S (10), was 0.9774 for clinical model and 0.9783 for biochemical model.
RR indicates relative risk; BP, blood pressure.

mation of 10-year risk of stroke event for each individual as shown in Table 2. Similarly, the biochemical model including total cholesterol (1 point) and fasting glucose (1 point) was constructed and the estimated risk for stroke event was assigned in Supplemental Table II. By using the simple points system, we determined that 23.5% of the sample had <1% risk, 49.0% had a 1% to 5% risk, 25.4% had a 5% to 20% risk, and 2.2% had a 20% of higher risk of incident stroke during a 10-year follow-up interval. This clinical points-based model has good discrimination ability with an AUC of 0.772 (95% CI, 0.744 to 0.799), which was similar to the biochemical points-based model (AUC, 0.773; 95% CI, 0.746 to 0.800) and the coefficient-based clinical model (0.778; 95% CI, 0.751 to 0.804) and coefficient-based biochemical model (0.779; 95% CI, 0.752 to 0.806); the test statistics were not significant among these 4 models ($P=0.25$; Supplemental Table IV). The estimated coefficients in Framingham and PROCAM for stroke were derived in Supplemental Table III. The AUCs from the Framingham and PROCAM stroke models (0.754; 95% CI, 0.726 to 0.783 for Framingham and 0.748; 95% CI, 0.722 to 0.777 for PROCAM) were significantly lower than that from our points-based clinical model ($P=0.001$; Figure 1). The optimal cutoff values for the points-based clinical and biochemical models were set as 8 with a sensitivity of 0.65 and a specificity of 0.75 for the clinical model and a sensitivity of 0.71 and a specificity of 0.69 for the biochemical model (Supplemental Table V). In addition, the correctly classified proportion of the clinical model was larger than that of the biochemical model (75% versus 69%), whereas both had a similar Youden index value.

Supplemental Figure I shows the Bland-Altman plot for the predicted risk between the clinical and biochemical points-based models, and the negative slope (−0.011, $P=0.02$) indicated that the clinical model underestimated risk for the high-risk subjects than the biochemical model. We also found that the NRI and IDI values between the clinical and biochemical models were not significant (NRI, −0.1%, $P=0.66$; IDI, −0.4%, $P=0.77$), and the reclassification table according to these 2 models showed no marked change for the risk stratification (Table 3), implying the similar discrimination performance between the clinical and biochemical points-based model for the risk of stroke. We provided a nomogram including the 7 clinical covariates for additional calculation tool (Figure 2).

Discussion

Using a community-based cohort study, we developed a points-based clinical model to predict 10-year risk of stroke event in Chinese population based on 7 variables: age, gender, systolic and diastolic blood pressure, family history of stroke, atrial fibrillation, and diabetes mellitus. These variables could be relatively easily obtained in clinical practice and the points system we developed is simple to use. The availability of a simple clinical tool to predict future risk of disease, as has been the case for prediction of coronary heart disease[37] and Type 2 diabetes,[21] will improve the prediction of stroke risk, identify high-risk populations, and enhance preventive strategies.

Several stroke prediction models have previously been developed in various populations (Supplemental Table I). Cross-sectional studies conducted in Spanish hypertensive

Table 2. The Simple Points System According to the Clinical Model and the Total Points and Absolute Risk Function in the Study Participants (n=3513)

Risk Factor	Category	Point
Age, years	35–39	0
	40–44	1
	45–49	2
	50–54	3
	55–59	4
	60–65	5
	65–69	6
	70–74	7
	≥75	8
Gender	Men	1
	Women	0
SBP, mm Hg	<110	0
	110–129	1
	130–149	2
	≥150	3
DBP, mm Hg	<75	0
	75–94	1
	≥95	2
Family history of stroke	No	0
	Yes	1
Atrial fibrillation	No	0
	Yes	3
Diabetes	No	0
	Yes	1

Point Total	Estimated Risk
0	0.003
1	0.004
2	0.005
3	0.007
4	0.011
5	0.015
6	0.021
7	0.030
8	0.043
9	0.060
10	0.085
11	0.119
12	0.165
13	0.227
14	0.308
15	0.408
≥16	0.527

SBP indicates systolic blood pressure; DBP, diastolic blood pressure.

women[8] and Korean hypertensive adults[6] provided a 10-year stroke risk based on the Framingham risk chart. Among cohort studies, clinical covariates, including age, gender, blood pressure, atrial fibrillation, and diabetes, have provided substantial predictive power for the risk of stroke. Among

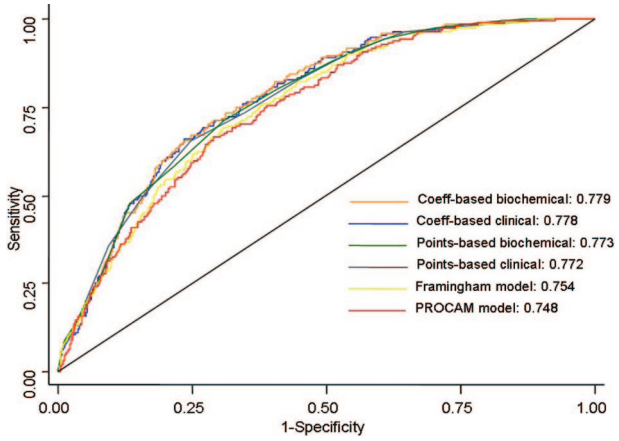

Figure 1. Receiver operating characteristic curves for various models applied to the study participants (n=3513).

705 new-onset atrial fibrillation participants recruited from the Framingham cohort and without anticoagulant therapy, age, gender, systolic blood pressure, diabetes mellitus, and previous stroke or ischemic attack history predicted 5-year stroke event significantly.[7] In addition, based on 7457 men aged 47 to 55 years in Sweden with >20 years follow-up, age, diabetes, blood pressure, and history of transient ischemic stroke, atrial fibrillation, and smoking were associated with a stroke event.[38] Asian reports on a stroke prediction model are also available. First, a Japanese prediction model including all stroke and coronary events for 19 years on 4098 men and 5255 women, aged 50 years, for a 19-year follow-up period showed age, smoking, systolic blood pressure, and serum cholesterol and glucose were included in the prediction model.[39] Second, a study based on 7209 Hong Kong Chinese with Type 2 diabetes showed that age, glycohemoglobin, spot urine albumin–creatinine ratio, and history of coronary heart disease predicted the stroke risk significantly.[4] Third, from the 9903 Chinese adults in 1 cohort data, Wu and colleagues constructed a gender-specific points-based prediction chart for 10-year risk of ischemic cardiovascular disease, in which age, systolic blood pressure, total cholesterol, diabetes mellitus, and smoking were included in the model.[40] Gender-specific effects of diabetes, cholesterol, and smoking on the risk of ischemic stroke were found among Chinese adults.[40] However, a prediction model specifically designed for the risk of stroke in Chinese adults is not currently available. Our clinical and biochemical models provide a feasible tool for identifying the high-risk individuals at the risk of stroke.

Using gender-specific age, systolic blood pressure, antihypertensive medication history, diabetes, smoking, history of cardiovascular disease, atrial fibrillation, and left ventricular hypertrophy, Wolf and colleagues developed a Framingham model to predict stroke incident during 10 years of follow-up in a white population.[16] Assmann and colleagues also constructed the PROCAM model, which included gender, age, systolic blood pressure, smoking, and diabetes status to predict stroke risk in another white population during a 10-year follow-up period and the AUC of this model reached 0.78 for stroke.[14] Using the participants in our study, the AUCs for Framingham and PROCAM models were signifi-

Table 3. Reclassification Tables for the Points-Based Clinical and Biochemical Prediction Models on the Study Participants (n=3513)

	Biochemical Model			
Clinical Model	Lowest Risk, <0.01	Low Risk, 0.01–0.049	Medium Risk, 0.05–0.20	High Risk, ≥0.20
Predicted risk				
Stroke(−)				
Lowest risk, <0.01	863	175	1	0
Low risk, 0.01–0.049	19	1409	120	0
Medium risk, 0.05–0.20	0	17	608	21
High risk, ≥0.20	0	0	6	34
Stroke(+)				
Lowest risk, <0.01	10	2	0	0
Low risk, 0.01–0.049	0	74	18	0
Medium risk, 0.05–0.20	0	1	115	0
High risk, ≥0.20	0	0	1	19

NRI and IDI values were statistically nonsignificant: NRI: −0.1%, P=0.66; IDI: −0.4%, P=0.77.

cantly lower than our clinical model. The underperformance may be attributed to regional and ethnic variation in predicting the risk of stroke.[41] In addition, our model contained family history of stroke and atrial fibrillation, which improved the prediction ability than other models. Finally, all stroke subtypes as the outcome would make feasible in screening for a general population.

Most of the variables included in our model were similar to those in previous risk functions. We found that white cell count, a marker of chronic inflammation, was incorporated into our coefficient-based biochemical model. However, the incremental effect of white cell count did not substantially increase the points-based score. The additional contribution by total cholesterol and glucose was minimal compared with age and blood pressure in our points-based model. Therefore, we argued that the points-based clinical model is sufficient for predicting the risk of stroke in Chinese adults. We also found that the nomogram using these clinical measures was consistent with the chart calculated from the points-based clinical model. In addition, we did not include information on lifestyle factors such as physical activity and smoking because they were not significant predictors of diabetes after adjusting for other risk factors. The role of smoking with regard to stroke prediction has been reported in previous literature.[5,15,40] In addition, left ventricular hypertrophy was incorporated into the Framingham stroke model.[16] Our prediction models did not include smoking, left ventricular hypertrophy, or history of coronary heart disease or transient ischemic stroke. These covariates did not increase significantly the likelihood ratio test value in the multivariate Cox model. Some lifestyle and socioeconomic factors such as physical activity,[12] household income,[10] and depression[38,42]

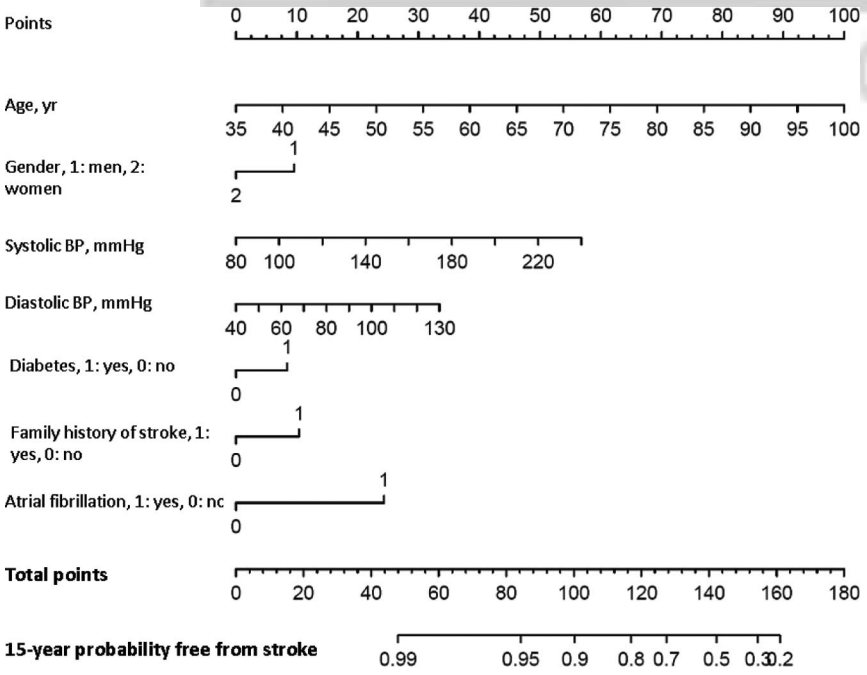

Figure 2. Nomogram for the risk of stroke incidence in the study participants (n=3513). The nomogram is used by first locating a participant position on each predictor variable scale. Each scale position has corresponding prognostic points (top axis). All points were summed to a total points value and a vertical line drawn from the total points axis straight down to the 15-year probability of free from stroke events.

were associated with the risk of stroke. However, we included family history of stroke and diastolic blood pressure measures, which increased the prediction power significantly. Previous studies also demonstrated that family history of stroke[43] and diastolic blood pressure[13,41] added to the predictive ability for stroke. Our clinical model had a similar performance measure as the laboratory model. Considering issues of practicality, cost, and feasibility, the clinical model is recommended for the first-line tool for screening high-risk individuals in a primary prevention setting.[44] Furthermore, to simplify risk prediction in office-based practices without blood sampling, the clinical model can be used as a prediction tool for stroke prevention.[37]

The AUC value depends on case mix and disease severity and may be sensitive among those with more severe disease. In addition, the AUC value was based on ranks and may not be sensitive in choosing between models.[32] Our study showed that the AUC values between the clinical and biochemical models were not significant (0.772 versus 0.773). Discrimination and calibration statistics, including net reclassification and integrated discrimination statistics, may be helpful for discriminating the predictive measures between models.[31]

To our knowledge, this is the first stroke prediction model specifically developed for a community-based Chinese population. Because of the large sample size, the estimates from our prediction models were found to be stable as demonstrated by the internal validation study. Also, the use of a community-based population reduced the possibility of selection bias. However, several potential limitations of this study should be mentioned: First, we included all types of stroke in this study and did not specify ischemic and hemorrhage subtypes, which indicated different profiles of risk factors. Second, we did not include extensive lifestyle, psychosocial, or biomarkers as well as subclinical disease data in the model. Lifestyle and psychosocial factors were helpful for primary prevention. Second, adding biomarkers such as lipoprotein(a) and subclinical diseases may improve the discriminatory ability.[18,45] However, these biochemical variables are more difficult to measure and interpret in clinical practice. Next, a lack of an independent population in Asia Pacific countries for external validation may not provide an accurate estimate for the model performance, because the internal validation used in this study may overoptimize the model fitting measures. Then, we did not check the temporal validation for the model. However, according to Altman and Royston,[46] incorporating all the available data because of the large sample size may be a better choice, and we may need to wait several years to accrue an adequate number of events in another cohort. Therefore, using all available data is optimal for the prediction model construction. In addition, the healthcare system for stroke in this community-based cohort was consistent, and we ascertained the stroke cases according to medical history and hospitalization record supplemented by image data. The improving diagnostic tools may have a modest influence on stroke ascertainment. Finally, our study population is rather older (average 55 years), and the age effect on stroke explained most of the risk of stroke (8 points), which diminished the magnitude of other important risk factors such as white blood cell, fasting glucose, and total cholesterol values. Further stratified analysis based on a younger age group may be warranted.

In conclusion, we have constructed a simple points-based clinical model for predicting 10-year incidence of stroke risk, and this model performed significantly better than existing Framingham and PROCAM prediction models within an ethnic Chinese group. This simple clinical tool should help identify high-risk populations and improve prevention and treatment strategies for the Chinese population.

Acknowledgments

We thank the staff of the Department of Internal Medicine, National Taiwan University Hospital, and the participants of the Chin-Shan Community Cardiovascular Cohort study for their valuable contributions.

Sources of Funding

This study was partly supported by the National Science Council (NSC 97-2314-B-002-130-MY3, 97-3112-B-002-034).

Disclosures

None.

References

1. Cholesterol, diastolic blood pressure, and stroke: 13 000 strokes in 450 000 people in 45 prospective cohorts. Prospective studies collaboration. *Lancet*. 1995;346:1647–1653.
2. Hebert PR, Gaziano JM, Chan KS, Hennekens CH. Cholesterol lowering with statin drugs, risk of stroke, and total mortality. An overview of randomized trials. *JAMA*. 1997;278:313–321.
3. Davis TM, Millns H, Stratton IM, Holman RR, Turner RC. Risk factors for stroke in type 2 diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS) 29. *Arch Intern Med*. 1999;159:1097–1103.
4. Yang X, So WY, Kong AP, Ho CS, Lam CW, Stevens RJ, Lyu RR, Yin DD, Cockram CS, Tong PC, Wong V, Chan JC. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care*. 2007; 30:65–70.
5. Almgren T, Persson B, Wilhelmsen L, Rosengren A, Andersson OK. Stroke and coronary heart disease in treated hypertension—a prospective cohort study over three decades. *J Intern Med*. 2005;257:496–502.
6. Park CG, Choi CU. Probability of stroke in Korean hypertensive patients visiting community-based hospitals: using a risk profile from the Framingham study. *J Hum Hypertens*. 2009;23:252–258.
7. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, Larson MG, Kannel WB, Benjamin EJ. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *JAMA*. 2003;290:1049–1056.
8. Coca A, Redon J, Cea-Calvo L, Lozano JV, Navarro J, Fernandez-Perez C, Bonet A, Gonzalez-Esteban J. Estimated risk of a first stroke and conditioning factors in Spanish hypertensive women. The RIMH study. *Blood Press*. 2006;15:237–244.
9. Bestehorn K, Wahle K, Kirch W. Stroke risk screening of adults with hypertension: prospective cross-sectional study in primary care. *Clin Drug Investig*. 2008;28:281–289.
10. Boysen G, Nyboe J, Appleyard M, Sorensen PS, Boas J, Somnier F, Jensen G, Schnohr P. Stroke incidence and risk factors for stroke in Copenhagen, Denmark. *Stroke*. 1988;19:1345–1353.
11. Kammersgaard LP, Olsen TS. Cardiovascular risk factors and 5-year mortality in the Copenhagen Stroke Study. *Cerebrovasc Dis*. 2006;21: 187–193.
12. Lumley T, Kronmal RA, Cushman M, Manolio TA, Goldstein SA. Stroke prediction score in the elderly: validation and web-based application. *J Clin Epidemiol*. 2002;55:129–136.
13. Wiberg B, Sundstrom J, Arnlov J, Terent A, Vessby B, Zethelius B, Lind L. Metabolic risk factors for stroke and transient ischemic attacks in middle-aged men: a community-based study with long-term follow-up. *Stroke*. 2006;37:2898–2903.

14. Assmann G, Schulte H, Cullen P, Seedorf U. Assessing risk of myocardial infarction and stroke: new data from the Prospective Cardiovascular Munster (PROCAM) study. *Eur J Clin Invest*. 2007;37:925–932.
15. Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2004;160:259–269.
16. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke*. 1991;22:312–318.
17. Ueshima H, Sekikawa A, Miura K, Turin TC, Takashima N, Kita Y, Watanabe M, Kadota A, Okuda N, Kadowaki T, Nakamura Y, Okamura T. Cardiovascular disease and risk factors in Asia: a selected review. *Circulation*. 2008;118:2702–2709.
18. Chien KL, Su TC, Jeng JS, Hsu HC, Chang WT, Chen MF, Lee YT, Hu FB. Carotid artery intima-media thickness, carotid plaque and coronary heart disease and stroke in Chinese. *PLoS ONE*. 2008;3:e3435.
19. Chien KL, Sung FC, Hsu HC, Su TC, Lin RS, Lee YT. Apolipoprotein A1 & B, and stroke events in a community-based cohort in Taiwan: report of Chin-Shan Community Cardiovascular Study. *Stroke*. 2002;33:39–44.
20. Lee YT, Lin RS, Sung FC, Yang CY, Chien KL, Chen WJ, Su TC, Hsu HC, Huang YC. Chin-Shan Community Cardiovascular Cohort in Taiwan: baseline data and five-year follow-up morbidity and mortality. *J Clin Epidemiol*. 2000;53:836–846.
21. Chien K, Cai T, Hsu H, Su T, Chang W, Chen M, Lee Y, Hu FB. A prediction model for type 2 diabetes risk among Chinese people. *Diabetologia*. 2009;52:443–450.
22. Chien KL, Hsu HC, Su TC, Chen MF, Lee YT, Hu FB. Fasting and postchallenge hyperglycemia and risk of cardiovascular disease in Chinese: the Chin-Shan Community Cardiovascular Cohort study. *Am Heart J*. 2008;156:996–1002.
23. Chien KL, Hsu HC, Su TC, Chen MF, Lee YT, Hu FB. Apolipoprotein B and non-high-density lipoprotein cholesterol and risk of coronary heart disease in Chinese. *J Lipid Res*. 2007;48:2499–2505.
24. Chien KL, Hsu HC, Su TC, Lee YT. Consistency in genetic inheritance mode and heritability patterns of triglyceride vs. high density lipoprotein cholesterol ratio in two Taiwanese family samples. *BMC Genet*. 2003;4:7–16.
25. Chien KL, Hsu HC, Chen PC, Su TC, Chang WT, Chen MF, Lee YT. Urinary sodium and potassium excretion and risk of hypertension in Chinese: report from a community-based cohort study in Taiwan. *J Hypertens*. 2008;26:1750–1756.
26. Chen PC, Chien KL, Hsu HC, Su TC, Sung FC, Lee YT. Metabolic syndrome and C-reactive protein in stroke prediction: a prospective study in Taiwan. *Metabolism*. 2009;58:772–778.
27. Hosmer DW Jr, Lemeshow S. *The Multiple Logistic Regression Model. Applied Logistic Regression,* 1st ed. New York: John Wiley & Sons; 1989:25–37.
28. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–256.
29. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med.* 2004;23:1631–1660.
30. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* London: Chapman & Hall/CRC; 1994.
31. Pencina MJ, D'Agostino RBS, D'Agostino RBJ, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172.
32. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928–935.
33. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA*. 2007;297:611–619.
34. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346:1085–1087.
35. Martarelli D, Martarelli B, Pompei P. Body composition obtained from the body mass index: an Italian study. *Eur J Nutr.* 2008;47:409–416.
36. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York: Springer; 2001.
37. D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117:743–753.
38. Harmsen P, Lappas G, Rosengren A, Wilhelmsen L. Long-term risk factors for stroke: twenty-eight years of follow-up of 7457 middle-aged men in Goteborg, Sweden. *Stroke*. 2006;37:1663–1667.
39. NIPPON DATA80 Research Group. Risk assessment chart for death from cardiovascular disease based on a 19-year follow-up study of a Japanese representative population. *Circ J*. 2006;70:1249–1255.
40. Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, Rao X, Zhou B, Detrano R, Liu K. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation*. 2006;114:2217–2225.
41. Tanne D, Yaari S, Goldbourt U. Risk profile and prediction of long-term ischemic stroke mortality: a 21-year follow-up in the Israeli Ischemic Heart Disease (IIHD) Project. *Circulation*. 1998;98:1365–1371.
42. Simons LA, McCallum J, Friedlander Y, Simons J. Risk factors for ischemic stroke: Dubbo Study of the elderly. *Stroke*. 1998;29:1341–1346.
43. Jousilahti P, Rastenyte D, Tuomilehto J, Sarti C, Vartiainen E. Parental history of cardiovascular disease and risk of stroke. A prospective follow-up of 14 371 middle-aged men and women in Finland. *Stroke*. 1997;28:1361–1366.
44. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I follow-up study cohort. *The Lancet*. 2008;371:923–931.
45. Chien KL, Hsu HC, Su TC, Sung FC, Chen MF, Lee YT. Lipoprotein(a) and cardiovascular disease in ethnic Chinese: the Chin-Shan Community Cardiovascular Cohort Study. *Clin Chem.* 2008;54:285–291.
46. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.