

Stroke Prediction Models: A Systematic Review

Jeena R S, Dr SukeshKumar A

Abstract—Stroke is one of the foremost causes of adult disability in many of the developing countries. Diagnosis of stroke during the initial stages is crucial for timely prevention and cure. Various conventional statistical methods and computational intelligent models like Framingham model have been developed for predicting the risk of stroke. These risk prediction models can aid in clinical decision making and help patients to have an improved and reliable risk prediction. This paper explores the various prediction models developed so far for the assessment of stroke risk.

Index Terms— Stroke, Prediction models, Framingham model.

INTRODUCTION

Stroke is a cerebro-vascular ailment affecting the normal blood supply to the brain. Without proper supervision, it leads to death or long term disability. Stroke can be either ischemic or hemorrhagic. There is a possibility for the co-occurrence of both ischemic and hemorrhagic strokes. Stroke[1] caused due to a clot in the blood vessel is known as Ischemic stroke and that due to a rupture of blood vessel is referred to as Hemorrhagic stroke. The deficiency of oxygen to the cerebral nervous system is referred to as ischemia which finally leads to their death. About 85 % of all strokes belong to ischemic category. Its frequency is accelerating in developing countries like India due to unhealthy lifestyles.

Stroke is largely driven by demography and enhanced by increasing frequency of modifiable risk factors. Several risk factors contribute to the risk of stroke. Lifestyle risk factors include diet, cigarette smoking [3], obesity, physical inactivity, alcohol consumption, family and genetic factors, age, sex, drug usage, race, geographic location, climate and socioeconomic factors whereas medical conditions include Atrial fibrillation, Blood pressure [2], Diabetes mellitus, Cholesterol, Mitral valve disease, C reactive protein, Sickle cell disease, Hyperlipidemia, and Transient ischemic attack (TIA). High blood pressure, heart disease and diabetes are the highest risk factors of stroke. But, they often do not cause symptoms in their early stages. Total count of risk factors is directly related to probability of stroke occurrence. Research work shows that physiological parameters can be considered as risk factors for predicting the near-term occurrence of stroke.

The aim of this paper is to review the presently existing and widely used cardiovascular risk assessment models and to examine the evidence available on new biomarkers and the nonclinical measures in improving the risk prediction. Identification of individuals at risk of cardiovascular disease (CVD) is of main concern. Cardiovascular risk prediction has progressed with the development and refinement of risk prediction models based upon well established clinical factors, discovery of biomarkers and social factors may give extra details on the risk of disease. Risk charts and risk score, based on global absolute risk are the key tools for CVD risk- assessment and determine the likelihood of developing the disease.

FRAMINGHAM STUDY

A health risk assessment function has been developed for the prediction of stroke using the Framingham Study cohort [4]. The stroke risk factors included in the profile are age, systolic blood pressure, the use of antihypertensive therapy, diabetes mellitus, cigarette smoking, prior cardiovascular disease, atrial fibrillation, and left ventricular hypertrophy by electrocardiogram. Probability of stroke was determined in subjects aged 55-84 years and free of stroke at the time of two examination cycles and is based on 10 years of follow-up from each of these examinations. Stroke probabilities were computed using the Cox proportional hazards model for each sex based on a point system. Its added advantage is that it permits the computation of stroke probabilities for variable lengths of follow-up, whereas the logistic model allows estimation for only a pre specified length of follow-up. Stroke risk can directly be estimated from a routine physical examination and an individual's risk can be related to the average risk of stroke for persons of the same age and sex. Regression coefficients and relative risks for significant risk factors in cox proportional model for stroke profiles in subjects aged 55-84 years and free of stroke are calculated. From these, a linear function is computed.

The general form of the function is

$$L = \text{Reg coefficient}_{\text{age}} \times \text{Age} + \text{Reg coefficient}_{\text{SBP}} \times \text{SBP} + \text{Reg coefficient}_{\text{HypRx}} \times \text{Hyp Rx} + \text{Reg coefficient}_{\text{DM}} \times \text{DM} + \text{Reg coefficient}_{\text{Cigs}} \times \text{Cigs} + \text{Reg coefficient}_{\text{CVD}} \times \text{CVD} + \text{Reg coefficient}_{\text{AF}} \times \text{AF} + \text{Reg coefficient}_{\text{LVH}} \times \text{LVH}.$$

(1)

where SBP is systolic blood pressure, Hyp Rx is use of antihypertensive therapy, DM is presence of diabetes mellitus, Cigs is cigarette smoking, CVD is previously diagnosed coronary heart disease, AF is presence of atrial fibrillation, and LVH is left ventricular hypertrophy. This function is next evaluated at the values of the means for each variable. Let this value be M.

$$M = \text{Reg coefficient}_{\text{age}} \times \text{Mean (yrs)} + \text{Reg coefficient}_{\text{SBP}} \times \text{Mean(mm Hg)} + \text{Reg coefficient}_{\text{HypRx}} \times \text{Mean}_{\text{hyp}} + \text{Reg coeffi-}$$

$$\text{coefficient}_{DM} \times \text{Mean}_{DM} + \text{Reg coefficient}_{Cigs} \times \text{Mean}_{cigs} + \text{Reg coefficient}_{CVD} \times \text{Mean}_{CVD} + \text{Reg coefficient}_{AF} \times \text{Mean}_{AF} + \text{Reg coefficient}_{LVH} \times \text{Mean}_{LVH}$$

(2)

This value of M is next subtracted from the general function L to produce a function A.

$$A = L - M \quad (3)$$

Exponential of the function A is then taken.

$$B = e^A \quad (4)$$

Choose the number of years (1 to 10) for which predictions are required. For each selection there is a value that is the estimated probability of surviving without a stroke for individuals whose risk factor values are equal to the mean values of those observed in the data. Let these values be represented by z(t), where the z is to indicate survival without a stroke and t is to index the number of years. Using cox regression model, the values of z(t) is computed on the basis of a point system.

The predicted probability that a person with a selected set of risk factors will develop a stroke within t years is given by

$$p = 1 - (z(t))^B \quad (5)$$

It may also help to identify persons at considerably increased stroke risk resulting from borderline levels of multiple risk factors. Different risk factors have been included in following versions of the score [5][6] and extensions to evaluate the risk of ten-year atrial fibrillation [7], ten-year CHD [8], 30-year CVD [9], eight-year diabetes [10], four-year hypertension [11] and ten year stroke [12] have also been proposed.

PROCAM

This model is based on 325 acute coronary events occurring within 10 years of follow-up among 5389 men 35 to 65 years of age at recruitment into the Prospective Cardiovascular Munster (PROCAM) study [13]. A Cox proportional hazards model with 8 independent risk factors, ranked in order of significance: age, LDL cholesterol, smoking, HDL cholesterol, systolic blood pressure, family history of premature myocardial infarction, diabetes mellitus, and triglycerides was developed. A simple point scoring system based on the regression coefficients of this model has also been implemented. Advantages of this model include variable duration of follow-up, censoring of subjects, proportionality of event occurrence, and time-to-event. To convert the results of the Cox model into absolute risk estimates, survival within the defined population has been calculated using Kaplan-Meier statistics.

GLOBAL VASCULAR RISK SCORE (GVRS)

A model has been developed using the NOMAS (Northern Manhattan Study) to improve currently available global cardiovascular disease risk prediction tools by incorporating both traditional and behavioral risk factors and to determine the incremental benefit of adding these risk factors [14]. A prediction tool has been developed from a racially and ethnically diverse cohort that would be useful for African-American and Hispanic people at risk for vascular disease. A cox proportional hazard model was developed by including all the traditional risk factors from the Framingham model. The traditional categorical risk attributes were replaced with continuous variables to enhance the fit of the model. Basic socio-demographic variables like age, sex, and race-ethnicity and selectively added other vascular risk factors that could be ascertained through history or blood tests were retained. Variables that were assessed included sibling history of stroke, waist circumference, BMI, waist-to-hip ratio, alcohol consumption, physical activity, peripheral vascular disease, atrial fibrillation, heart disease, homocysteine, white blood cell count, and creatinine levels. All 2-way interaction terms were examined and the terms contributing significantly to the fit by the likelihood ratio criterion are included in the final model. A global vascular risk score (GVRS) was computed by summing a product of a linear predictor from the final model. Kaplan-Meier curves for survival free of stroke, MI, or vascular death were plotted for the quartiles of the GVRS.

REYNOLDS RISK SCORE

To improve the cardio-vascular disease risk prediction, the Reynolds risk score (RRS) was derived in a cohort of 25,000 healthy U.S. women [11]. The Reynolds Risk Score includes traditional risk attributes used in the Framingham Risk Model and adds parental history of premature Coronary Heart Disease and high-sensitivity C-reactive protein. This study along with RRS calibrated for men, provided superior prediction of CVD events compared with the FRS model in 2 studies [15][16]. RRS was developed and tested in predominantly non-Hispanic white populations.

MUCA

A MUCA [17] ischemic CVD risk model was developed for the Chinese population [54]. Data from the USA–People’s Republic of China (PRC) Collaborative Study of Cardiovascular and Cardiopulmonary Epidemiology (the USA-PRC Study) cohort was used as derivation cohort to generate the prediction models and data from independent China Multicenter Collaborative Study of Cardiovascular Epidemiology (MUCA) Cohort II was used as the validation cohort to test whether the derived

models are applicable to recent Chinese populations. Risk factors include Weight, height, 3 consecutive blood pressure readings, 12-hour fasting blood sample, and blood lipids to measure total cholesterol.

ASSIGN

The ASSIGN score was derived from cardiovascular outcomes in the Scottish Heart Health Extended Cohort (SHHEC) [18]. It was tested against the Framingham cardiovascular risk score in the same database. Classic risk factors including cigarette smoking status, plus deprivation and family history but not obesity were significant factors in constructing ASSIGN score for each sex. The agreement between ASSIGN and Framingham was explored by rank correlations, kappa statistics, and by comparing results of similar thresholds for treatment, and equal-sized high-risk groups.

CUORE

A CUORE equation was developed in Italy for a low coronary incidence population [19] for the prediction of coronary and cerebrovascular events in 10 years. Risk factors analyzed in this model include age, systolic blood pressure, total cholesterol, HDL-cholesterol, smoking habit, diabetes and hypertension treatment. Multivariate analysis was done to assess the relation of risk factors to the 10-year CVD risk. Starting from a basic Cox model that included age, SBP and smoking, the other cardiovascular risk factor variables that were statistically significant in the univariate analysis were progressively included. The 10-year CVD individual risk score models were implemented including continuous values of risk factors. The main advantage of this project is that the cohorts had a high number, were enrolled in relatively recent time, included women, covered various geographical regions of the country and were followed up for total and cause specific mortality and non-fatal coronary as well as cerebrovascular events, which were validated using standardized techniques.

QRISK

QRISK is likely to provide more appropriate risk estimates to help identify high risk patients on the basis of age, sex, and social deprivation. The derivation cohort consisted of 1.28 million patients, aged 35-74 years, registered between January 1995 - April 2007 and who were free of diabetes and existing cardiovascular disease. The validation cohort consisted of 0.61 million patients from 160 practices. Risk factors include age, sex, smoking, systolic blood pressure, ratio of total serum cholesterol to high density lipoprotein, body mass index, family history of coronary heart disease aged less than 60, area measure of deprivation, and existing treatment with antihyperten-

sive agent. Cox proportional hazards model was used in the derivation dataset to estimate the coefficients associated with each risk factor for the first ever recorded diagnosis of cardiovascular disease gender wise. The 10 year cardiovascular disease risk for each patient was computed in the validation dataset, replacing missing values for continuous variables with mean values obtained from the derivation dataset by five year age-sex bands and assuming patients were non-smokers when smoking status was not recorded. To assess calibration, the mean predicted risk of cardiovascular disease was calculated at 10 years and the observed risk at 10 years was obtained using the 10 year Kaplan-Meier estimate. QRISK was followed by a reorganized model (QRISK2) in 2008, which included ethnic origin and additional risk factors like type 2 diabetes, rheumatoid arthritis, atrial fibrillation, and chronic renal disease. [20]

SCORE

The SCORE equation has been recommended by the Fifth Joint European Task Force on cardiovascular prevention [21]. SCORE chart for risk estimation was based on data from 12 European cohort studies which included 205178 subjects examined at baseline between 1970-1988 with 2.7 million years of follow-up and 7934 cardiovascular deaths. SCORE risk function has also been externally validated. Special risk charts based on SCORE were produced for both low and high risk countries and gained wide approval throughout Europe. Priorities were proposed at four levels: patients with established disease, asymptomatic individuals at high risk of CVD mortality, first-degree relatives of patients with premature CVD, and other individuals encountered in clinical practice. Ten-year risk of fatal cardiovascular disease was calculated using a Weibull model in which age was used as a measure of exposure time to risk rather than as a risk factor. The model has two sections: one section models the shape of the baseline survival function and the other computes the relative risks linked with the risk factors. The model was stratified on cohort and sex that is, separate hazard functions were computed for men and women in each of the component cohorts, but risk factor coefficients were computed from the entire database. This approach assumes that risk factors do not vary in their effect from country to country and are the same in men and women. Risk of cardiovascular death was calculated by combining two separate risk estimations: a model for coronary heart disease and a model for all non-coronary atherosclerotic cardiovascular disease.

Development of a Stroke risk prediction model in National Health insurance services personal health record in Korean Population

Another 10-year stroke prediction model was devel-

oped to categorize probability of stroke using the Korean national health examination data [22]. It developed an algorithm to provide a personalized warning on the basis of each user's stage of stroke risk for Korean population. It sorted the user's individual probability of stroke into five classes –normal, slightly high, high, risky, very risky, based on the five ranges of average probability of stroke in comparison to total population –less than 50 percentile, 50–70, 70–90, 90–99.9, more than 99.9 percentile. The finally selected risk factors by gender includes age, hypertension, past history of heart disease, past history of stroke, family history of heart disease, exercise, alcohol drinking, total cholesterol, smoking, hypertension, diabetes and body mass index for male whereas age, past history of hypertension, heart disease and stroke, family history of heart disease and stroke, exercise, alcohol drinking, total cholesterol, smoking, hypertension, diabetes and body mass index were considered for female. Cox proportional hazard regression model was used for the prediction model, similar to that of Framingham study [4]. The half of the data was used for model building and the other half for external validation. The model estimation and the internal validity test were conducted with the model construction data, and the external validity test was conducted on the validation data with the constructed model. AUC, which estimated the ROC curve area was used to compute the discrimination ability in this model.

CONCLUSION

Various risk prediction models are used in clinical assessment and are used to aid patients make an informed choice about their treatment. This paper attempted to deliver the existing picture of the cardiovascular risk assessment models which aids in stroke prediction. Key to the usefulness of determining the likelihood of stroke by means of a risk profile is an indication that modification of several powerful risk attributes will surely reduce stroke probability.

REFERENCES

1. Barry L. Zaret, M.D., Marvin Moser, M.D., Lawrence S. Cohen.: Chapter 18 Stroke - Lawrence M. Brass, M.D. (pgs 215-234)
2. MacMahon S, Rodgers A.: The epidemiological association between blood pressure and stroke: implications for primary and secondary prevention. *Hypertens Res.* 1994;17(suppl 1):S23-S32.
3. Shinton R, Beevers G.: Meta-analysis of relation between cigarette smoking and stroke. *BMJ.*
4. P.A. Wolf , R.B. D'Agostino , A.J. Belanger , W.B. Kannel , Probability of stroke: a risk profile from the framingham study, *Stroke* 22 (1991) 312–318 .
5. Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. *American heart journal.* 1991;121(1 Pt 2):293-8.
6. D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.*2008;117(6):743-53.
7. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet.* 2009;373(9665):739-45.
8. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults. (Adult Treatment Panel III). *Jama.* 2001;285(19):2486-97.
9. Pencina MJ, D'Agostino RB, Sr., Larson MG, et al. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation.* 2009;119(24):3078-84.
10. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Archives of internal medicine.* 2007;167(10):1068-74.
11. Parikh NI, Pencina MJ, Wang TJ, et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Annals of internal medicine.* 2008;148(2):102-10.
12. D'Agostino RB, Wolf PA, Belanger AJ, et al. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke; a journal of cerebral circulation.* 1994;25(1):40-3.
13. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation.*2002;105(3):310-5.
14. White H, Boden-Albala B, Wang C, Elkind MS, Rundek T, Wright CB, et al. Ischemic stroke subtype incidence among whites, blacks, and Hispanics: the Northern Manhattan Study. *Circulation.*2005;111:1327–1331.
15. P.M. Ridker, J.E. Buring, N. Rifai, N.R. Cook, Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds risk score *JAMA*, 297 (2007), pp. 611-619.
16. P.M. Ridker, N.P. Paynter, N. Rifai, J.M. Gaziano, N.R. Cook, C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds risk score for men *Circulation*, 118 (2008), pp. 2243-2251.
17. Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation.* 2006;114(21):2217-25.
18. Woodward M, Brindle P, Tunstall-Pedoe H, et al. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart.* 2007;93(2):172- 21.
19. Ferrario M, Chiodini P, Chambless LE, et al. Prediction of coronary events in a low incidence population. Assessing accuracy of the CUORE Cohort Study prediction equation. *International journal of epidemiology.* 2005;34(2):413-415.
20. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *Bmj.* 2008;336(7659):1475-82.
21. Perk J, De Backer G, Gohlke H, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice. *European heart journal.* 2012;33(13):1635-701.
22. Jae-woo Lee a , Hyun-sun Lim b , Dong-wook Kim b , Soon-ae Shin c , Jinkwon Kim d , Bora Yoo e , Kyung-hee Cho e , The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record , *Elsevier Computer Methods and Programs in Biomedicine* 153 (2018) 253–257.