

Modelling Modifiable Risk Factors of Stroke in South Africa: Classical and Bayesian Logistic Regression Models

Delson Chikobvu

University of the Free State

Lyness Matizirofa (✉ Lynessmatizirofa@gmail.com)

University of South Africa

Research Article

Keywords: Stroke, Modifiable risk factors, Classical statistics, Bayesian statistics, South Africa

Posted Date: January 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-141783/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title Page

Modelling modifiable risk factors of stroke in South Africa: Classical and Bayesian logistic regression models

Type of article: Original

Authors: Delson Chikobvu¹; Lyness Matizirofa²

Affiliation: ¹ University of the Free State, Department of Mathematical Statistics and Actuarial Science, P.O. Box 339, Bloemfontein 9300, South Africa

² University of South Africa, College of Science, Department of Statistics, Roodeport, Johannesburg 1709, South Africa

Corresponding author: Lyness. Matizirofa, Florida Campus, University of South Africa, Roodeport, South Africa. Email: lynessmatizirofa@gmail.com

Abstract

Background: Stroke is the second leading cause of death and long-term disability in South Africa (SA). Yet little is known in SA about the modelling of modifiable risk factors of stroke. Information on the relative contribution of modifiable risk factors on stroke incidence is needed for early interventions in SA. Identification of risk factors remains the principal aspect of stroke prevention. This study aims to identify and quantify the risk of stroke associated with modifiable stroke predictors in SA.

Methods: A cross-sectional hospital-based study was employed to identify and quantify the risk of stroke associated with modifiable stroke predictors using 35730 individual patient data retrieved from selected private and public hospitals between January 2014 and December 2018 in SA. Multivariate logistic regression analysis was employed to assess the effect of modifiable factors on stroke. Bayesian logistic regression analysis was employed to understand the uncertainty in the estimation of model parameters.

Results: The dominant modifiable risk factors were: hypertension, cholesterol, heart problems and diabetes and all depend on the age of an individual and the interaction of these factors. The parameters, in the simpler model, of having hypertension, cholesterol, heart problems and diabetes were positive and significant confirming greater impact on the risk of stroke distribution. For instance, the odds ratio(OR) of patients with hypertension developing stroke when compared to those without hypertension is 80% higher. The OR of diabetic patients developing stroke when compared to non-diabetic was 194% higher.

Conclusions: Most strokes are attributable to modifiable factors. Study findings will be used to raise awareness of modifiable risk factors to prevent strokes and recommend regular screening and treatment of identified risk factors. This will reduce the burden of stroke in SA.

Keywords: Stroke, Modifiable risk factors, Classical statistics, Bayesian statistics, South Africa

Background

Stroke is the major leading cause of long-term disability and death globally [1]. Stroke is becoming a challenging public health problem in Africa, yet not enough is known about the modelling of modifiable risk factors of stroke[2]. African countries are undergoing an epidemiological transition driven by socio-demographic and lifestyle changes leading to an upswing of non-communicable diseases(NCD) such as stroke[3]. Global estimates suggest that sub-Saharan Africa (SSA) has the highest incidence of 316 / 100 000 people per year suffering from stroke [1]. Paradoxically, there is insufficient information on the current epidemiology of stroke in African countries and low middle-income countries (LMIC) [3]. Reducing the burden of stroke requires identification of modifiable risk factors of stroke and the treating of medical conditions such as hypertension, cholesterol, heart problems and diabetes, that increase stroke risk [4]. This study aims to identify and quantify the modifiable stroke risk factors in SA. The study results will be used to design appropriate strategies for risk factor reduction. There is a need to know the prevalence of modifiable stroke risk factors in the South African population.

SA is undergoing epidemiologic transition leading to increases in cardiovascular disease such as stroke [5]. Not enough is known about modelling modifiable stroke risk factors in SA. Stroke is the second leading cause of death and long-term disability in SA with crude stroke incidence of

244 / 100 000 person-years and crude mortality of 144 per 100 000 person-years [5]. To reduce the burden of stroke in SA, the current study will identify and quantify the risk associated with modifiable stroke risk factors using appropriate statistical methods and the study findings will be used to raise awareness on dangers of the modifiable risk factors of stroke.

Risk factors of stroke can be classified as modifiable and non-modifiable [6]. Non-modifiable risk factors are unpreventable such as age, race and heredity whilst modifiable factors are controllable such as hypertension, cholesterol, diabetes and obesity [7]. This study identified and quantified the effect of stroke modifiable risk factors in SA using classical and Bayesian logistic regression analysis. Several studies carried out in western countries showed major modifiable stroke risk factors as hypertension, physical inactivity alcohol consumption, cholesterol, smoking, heart disease, obesity and diabetes [4]. The effect of these factors needs to be quantified in a predominantly black population where the incidences of a stroke may be more prevalent. The impact of risks such as hypertension is suspected to be greater in this population.

Based on the literature, many studies used logistic regression to investigate modifiable risk factors of stroke despite its limitation of focusing merely on the conditional mean. This study used classical and Bayesian logistic regression analysis to identify and quantify the effect of modifiable risk factors on stroke in SA. The following paragraphs describe the literature review on the research methodologies of the stroke risk factors used by many researchers in different nations.

A systematic review was conducted in South Asia to identify common stroke risk factors. This review revealed that hypertension, smoking and diabetes were the dominant modifiable risk factors of stroke [8]. A meta-analysis was conducted in China to investigate modifiable and non-

modifiable risk factors of ischemic stroke and found that cholesterol and diabetes were significantly associated with ischemic stroke [6]. Logistic regression was used to identify the modifiable and non-modifiable risk factors of stroke and the study findings showed that hypertension and diabetes were prevalent in old aged white males [7]. Cox proportional hazards model was used to investigate modifiable risk factors associated with stroke incidence in very old people in Sweden. This Swedish study showed that high blood pressure and heart problems were strongly associated with stroke incidence in elderly people [9]. A Rotterdam study calculated the population attributable risks (PARs) for individual modifiable etiological risk factors to estimate the proportion of strokes that could be preventable concluded that sex-adjusted combined PAR of hypertension, cholesterol, smoking, diabetes, overweight/obesity were the most important etiological factors [10]. Hypertension, diabetes, Atrial fibrillation(AF) and cholesterol were identified as major stroke risk factors in an American study through multivariate logistic regression [11].

A cross-section study in Saudi Arabia used descriptive statistics to recognise the prevalence of modifiable risk factors in the Saudi population. The results showed that most patients had multiple risk factors including hypertension, diabetes, obesity and cholesterol [12]. Descriptive statistics were also used to understand the knowledge on and awareness of risk factors of hypertension in a rural community in SA [13]. The SA study findings indicated that the hypertension risk factors were excessive salt intake, alcohol consumption and physical inactivity led to the high prevalence of cardiovascular diseases (CVD) such as stroke. An American study used logistic regression analysis to determine stroke risk factors and established that hypertension, diabetes and cholesterol were high in old aged black people.

Further, conditional logistic regression was used to estimate odds ratios(OR) and PARs with 95 % confidence intervals (95% CIs) for modifiable risk factors of stroke in Ghana and Nigeria. The

findings indicated that hypertension, cholesterol, smoking and diabetes were the important risk factors of stroke [1]. Multivariable logistic regression was used to identify the predictors of stroke subtypes, hypertension, family history of stroke, alcohol consumption, and heart problems were identified as the major risk factors for ischemic stroke [14]. The largely used Generalised linear models in medical research during the past decades found hypertension, cholesterol, smoking, obesity, alcohol consumption and diabetes as dominant modifiable risk factors. Li et al [15] used Bayesian networks to estimate the stroke incidence in China and constructed a variable with the combination of smoking, overweight and hypertension. This study used both the classical statistics approach and Bayesian statistics approach to identify modifiable predictors as well as quantify their effect on stroke in SA. Prevention begins with using appropriate statistical methods identifying and estimating the effect of stroke predictors on stroke.

Modelling techniques are increasingly used in medicine to help make medical diagnoses and predict stroke risk factors treatment efficacy. The Markov chain Monte Carlo (MCMC) simulation procedure is designed to fit Bayesian statistics models. The Bayesian approach is distinct concerning to both flexibilities in which prior information (expert knowledge) can be incorporated with the use of a prior statistical probability distribution. This study aims to identify and quantify the effect of modifiable stroke risk factors using the Classical and Bayesian statistics (MCMC logit) modelling approaches.

Predictors of stroke such as blood pressure, cholesterol and diabetes are normally measured with error [16]. If these measurement errors (ME) are not corrected during analysis, the effects will be biased leading to false information. ME in covariates may lead to grave biases in parameter estimates and confidence intervals of statistical models if inappropriate statistical methods are used. The most commonly used method for parameter measurement error correction approach is the

Bayesian analysis using MCMC [17]. This study used MCMC logit approach to get parameter estimates and to quantify the uncertainty in these parameters.

Methods

A cross-sectional hospital-based study was employed to identify and quantify modifiable stroke risk factors using individual patient data retrieved from selected private and public hospitals between January 2014 and December 2018 in SA.

Stroke was defined according to the World Health Organization(WHO) as a syndrome of rapidly developing clinical signs of focal or global disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death with no apparent origin other than vascular [18]. Confirmation of stroke was based on computed tomography (CT) or magnetic resonance imaging (MRI).

Measurements

Dependent variable

The response variable was confirmed stroke, coded as 0 = no to confirmed stroke and 1 = yes to confirmed stroke.

Independent variables

Diabetes was defined as a fasting glucose concentration of greater than 7.0 mmol/L, cholesterol was defined as fasting cholesterol concentration of at least 5.2 mmol/L, high-density lipoproteins (HDL) cholesterol at least 1.03 mmol/L and low-density lipoproteins (LDL) cholesterol of at least 3.4 mmol/L. Whilst hypertension was defined with cut off of 140/90 mmHg for up to 72 hours and heart problems was defined as current atrial fibrillation, heart failure, ischemic heart disease and valvular heart diseases [1].

the risk factors come as a 'combination or combo' with more than one risk factor per individual. The modifiable risk factors of stroke were hypertension, cholesterol, heart problems and diabetes coded as 0 = no if the measurement is below the defined value of interest and 1 = yes if the measurements exceeded the study definition. The main effects were modelled first before adding up to five-factor interaction variable: hypertension-yes*cholesterol-yes*heart problems-yes*diabetes-yes*age was generated also. This variable was generated to adjust for interaction bias between study modifiable variables. Often

Setting

The study was conducted in the nine provinces of SA. The inclusion criteria for the study were stroke patients aged 18 years and above and were admitted in sampled private and public hospitals between January 2014 and December 2018. There are approximately 407 public and 203 private hospitals in SA [19]. This study, randomly selected 55% of the 203 private hospitals in all provinces whilst 45% of the 407 public hospitals were randomly selected across nine provinces of SA. A stratified probability sampling technique was used to calculate the proportions accordingly. The strata being public and private hospitals. Thus, study data were retrieved from 183 public and 112 private hospitals, making a total of 295 hospitals. Although most South Africans use public hospitals, many of these institutions did not capture good quality pertinent variables while private hospitals were doing so. The proportions used were based on the availability of study variables in the public and private hospital databases and the total number of private and public hospitals in SA. Therefore, 55 % of the data were retrieved from private hospitals and 45% from public hospitals. The proportions used were based on the total number of private and public hospitals in SA.

Data collection

A validated data retrieval sheet which was developed with the help of experts in the field was used to retrieve study data. In this study, patients' medical records were reviewed to elicit all modifiable risk factors of stroke. The data retrieval sheet was formulated with all the study variables include; confirmation of stroke (CT/MRI) and modifiable stroke risk factors. The variable type of hospital that is private or public hospital was anonymous for ethical reasons which mean there was no variable specifying type of hospital admitted as agreed upon in advance. The study hospitals were sampled from nine provinces of SA (Gauteng, KwaZulu-Natal, Western Cape, Eastern Cape, North West, Free State, Limpopo, Mpumalanga and Northern Cape). The case managers for the sampled hospitals assisted with data retrieval. The total number of stroke patients including the non-confirmed strokes was 35730.

Ethical considerations

This study was approved by the committee of research on human subjects of the University of South Africa as well as the study hospitals. The ethical clearance reference number is 2017/SSR-ERC/001. All methods were carried out in accordance of with relevant guidelines of Helsinki that is patients' rights were adhered to by not using patient names, IDs in reporting study results and also hospital names were not used in the data analysis as agreed in advance during ethical clearance application process. Written informed consent forms were obtained from the hospital managers before data retrieval from patients' medical records.

Classical statistical analysis

Quantitative data analysis was done using R statistical software version 4.2.0. Frequencies and percentages were used to summarise the prevalence of categorical variables. Multivariate logistic regression analysis was employed to assess the effect of modifiable risk factors on stroke. Whilst, Bayesian logistic regression analysis was employed to understand the uncertainty in the estimation of parameters. MCMC was used in this study because it is more appropriate in many situations with complex joint statistical distributions [20].

The study logistic regression for the main effects model for modifiable predictors for stroke is

$$\text{given as: } \text{logit}(\pi) = \beta_0 + \mu_{\text{hypertension-yes}} + \mu_{\text{cholesterol-yes}} + \mu_{\text{heart problems-yes}} + \mu_{\text{diabetes-yes}} + \epsilon \quad (1)$$

where $\pi = P(y = 1)$ is the probability of developing stroke, $\mu_{\text{hypertension-yes}}$ is the

hypertension effect on stroke, $\mu_{\text{cholesterol-yes}}$ is the cholesterol effect on stroke,

$\mu_{\text{diabetes-yes}}$ is the diabetes effect on stroke, $\mu_{\text{heart-problems-yes}}$ is the heart problems effect

on stroke, ϵ is the error term and $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.

The study logistic regression model for modifiable predictors can be re-expressed in terms of the β_j 's.

$$\text{logit}(\pi) = \beta_0 + \beta_{\text{hypertension-yes}}X_{\text{hypertension-yes}} + \beta_{\text{cholesterol-yes}}X_{\text{cholesterol-yes}} + \beta_{\text{heart-problems-yes}}X_{\text{heart problems-yes}} + \beta_{\text{diabetes-yes}}X_{\text{diabetes-yes}} + \epsilon \quad (2)$$

β_0 is the intercept, β_j is the parameter, X_j are dummy variables for covariate j and ϵ is the error term.

Where $y = 1$ is the confirmed stroke for an individual, β_0 is the intercept, $\beta_1 \dots \beta_p$ are the other p unknown parameters, X_1, \dots, X_p are the known p independent covariates. It is the dummy variables associated with (hypertension-yes, cholesterol-yes, heart problems-yes and diabetes-yes, with 2 categories respectively and ϵ is the error term. The basis is a person without any of hypertension, cholesterol, heart problems and diabetes. Based on literature, modifiable variables often occur in combinations, interaction variables were also generated to adjust for interaction bias between study modifiable variables. Hence this study developed a five-factor interaction model.

The study five-factor interaction logistic regression model for modifiable predictors for stroke is given as $\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p + \epsilon \dots \dots \dots \dots \dots \dots$ (3)

β_0 is the intercept, β_j is the parameter, X_j are dummy variables for covariate j and ϵ is the error term.

Where $\pi = P(y = 1)$, $y = 1$ is the confirmed stroke for an individual, β_0 is the intercept, $\beta_1 \dots \beta_p$ are the other p unknown parameters, X_1, \dots, X_p are the known p independent covariates. It is the dummy variables associated with (hypertension-yes, cholesterol-yes, heart problems-yes, diabetes-yes, with (2 categories) respectively, interaction variables for this model are defined in table 4, for instance $x_{27} = x_1 \times x_2 \times x_3 \times x_4 \times x_5 = \text{hypertension-yes} * \text{cholesterol-yes} * \text{heart problems-yes} * \text{Diabetes-yes} * \text{age}$ is the combination of hypertension-yes, cholesterol-yes, heart problems-yes*diabetes-yes and age effect on stroke and ϵ is the error term and $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$. See table 4 for the full list of covariates(variables).

The prior, likelihood, and posterior distributions in Bayesian Statistical analysis

Often in scientific research, we have prior knowledge about a disease (expert knowledge). This prior belief is quantified by expressing a prior statistical distribution on the model parameters. Data, represented by the likelihood, is collected to update this prior knowledge. The result is an updated belief distribution on the parameters. The updated belief distribution on the parameters is the posterior distribution.

Let $X = (X_1 X_2 \dots X_p)$ = the observed stroke predictors.

The prior could be taken as the non-informative flat prior to express ignorance of any knowledge initially of the regression parameters.

$$P(\beta_0, \beta_1 \dots \beta_p) \propto \text{constant}$$

The likelihood is the product of the data density $P(X|\beta_0, \beta_1 \dots \beta_p)$

The posterior of $\beta_0, \beta_1 \dots \beta_p$ may now be found by combining the prior with the likelihood.

The posterior is approximated as:

$$\text{Posterior} \propto \text{Prior} \times \text{likelihood}$$

$$P(\beta_0, \beta_1 \dots \beta_p | X) \propto P(\beta_0, \beta_1 \dots \beta_p) \times P(X | \beta_0, \beta_1 \dots \beta_p)$$

MCMC is then used to extract the marginal distributions of $\beta_0, \beta_1 \dots \beta_p$ from this joint posterior.

The posterior means of each of the parameters can be used to calculate other statistics of interest, such as the odds ratio and credibility intervals. The MCMC logit in R is used to compute the posterior means of the $\beta_0, \beta_1, \dots, \beta_p$ and the corresponding credible intervals.

The results are presented in the next section.

Results

Table 1. The prevalence of modifiable risk factors and type of stroke(n=35730)

Variable	Number	%
Hypertension		
Yes	19756	55.3
No	15974	44.7
Cholesterol		
Yes	16923	47.4
No	18807	52.6
Heart-problems		
Yes	19453	54.4
No	16277	45.6
Diabetes		
Yes	22183	62.1
No	13547	37.9
Type of stroke		
Ischemic	27550	77.1
Hemorrhagic	8180	22.9

The prevalence of hypertension was 19756 / 100 000. The incidences of stroke due to cholesterol was 47.4%. Most of the strokes were due to diabetes 22183 / 100 000. The majority of people developed stroke because of heart problems 19453 / 100 000 A total of 27550 patients were ischemic strokes and only 22.9% were hemorrhagic (Table 1).

Table 2. Multivariate logistic regression model for modifiable risk factors of stroke

Variable	Coefficient (β_j)	Odds ratio: $\exp(\beta_j)$	P-value
Intercept	-0.90		0.000
Hypertension-yes	0.59	1.80	0.000
Cholesterol-yes	0.61	1.84	0.000
Heart problems-yes	0.25	1.28	0.000
Diabetes-yes	1.08	2.94	0.000

Table 2 shows the model with only p=4 main effects covariates. The parameters in this simple model, of having hypertension, cholesterol, heart problems and diabetes are positive and significant, thus confirming their greater impact on the risk of stroke distribution. The odds of developing stroke in this population are much less than not developing stroke evidenced by a negative intercept of -0.90. A basis is a person without hypertension, cholesterol, heart problems and diabetes. Considering this reference group, model parameters are positive. The log odds of people suffering a stroke due to hypertension is 0.59 when compared to those without hypertension. The coefficient is positive which means hypertensive individuals are more at risk of developing stroke than people without hypertension in South Africa. The odds ratio is 1.80 when comparing people with hypertension to those without. The odds for hypertensive people developing stroke are approximately 80% higher than the odds for those without hypertension.

Moreover, the odds ratio of patients with cholesterol developing stroke when compared to the basis is 84% higher. This means individuals with cholesterol are much more at risk of suffering stroke than those without cholesterol in this South African population. Further, the odds ratio for people with diabetes compared to those without diabetes is 2.94. These patients have a 194% higher risk of developing stroke than those who are not diabetic in South Africa. Study findings also indicate that the log odds of people with heart problems developing stroke is 0.25 implying that individuals

with heart problems are more at risk of developing stroke than those without heart problems in SA. All coefficients are significant, meaning the effect of these modifiable factors on stroke was significant.

Table 3. MCMC Logit model for modifiable risk factors of stroke

Variable	Mean of β_j	95% Credible Interval of β_j
Intercept	-0.92	-0.94 , -0.86
Hypertension-yes	0.62	0.60 , 0.67
Cholesterol-yes	0.64	0.62 , 0.69
Heart problems-yes	0.24	0.23 , 0.29
Diabetes-yes	1.08	1.07, 1.13

The Bayesian approach allows us to quantify the uncertainty in the calculated parameters, β_j 's. The means of the posterior parameters are very close to the classical estimates obtained in table 2 using maximum likelihood estimates. This is not entirely surprising considering that a non-informative prior was used. The 95% Credible intervals are also given showing that zero is not included in all intervals (Table 3). The uncertainty in the parameter estimates are expressed in the range of the credible intervals. The positive posterior mean for hypertension, cholesterol, heart problems and diabetes means that patients with hypertension, cholesterol, heart problems, and diabetes are more likely to develop a stroke than those without these conditions (Table 3). In table 4, the results are adjusted for interaction bias between the study modifiable covariates. Risk factors come as a 'combination or combo' with more than one risk factor per individual.

Table 4. Multivariate logistic regression model for modifiable and interaction risk factors of stroke

Variable	Coefficient (β_j)	Odds ratio: $\exp(\beta_j)$	P-value
Intercept	-1.056		0.000
X_1 =Hypertension-yes	0.590	1.80	0.000
X_2 =Cholesterol-yes	0.610	1.84	0.000
X_3 =Heart-problems-yes	0.250	1.28	0.000
X_4 =Diabetes-yes	1.080	2.94	0.000
X_5 =Age	-0.006	0.994	0.026
$X_6 = X_1 \times X_2$ =Hypertension-yes*cholesterol-yes	-1.079	0.34	0.000
$X_7 = X_1 \times X_3$ =Hypertension-yes*heart problems-yes	-0.319	0.727	0.225
$X_8 = X_1 \times X_4$ =Hypertension-yes*diabetes-yes	-1.474	0.229	0.000
$X_9 = X_4 \times X_2$ =Diabetes-yes*cholesterol-yes	-1.986	0.137	0.000
$X_{10} = X_4 \times X_3$ =Diabetes-yes*heart problems-yes	-1.233	0.294	0.000
$X_{11} = X_1 \times X_5$ =Hypertension-yes*age	-0.007	0.993	0.033
$X_{12} = X_2 \times X_5$ =Cholesterol-yes*age	-0.019	0.981	0.000
$X_{13} = X_3 \times X_5$ =Heart-problems-yes*age	-0.038	0.963	0.000
$X_{14} = X_4 \times X_5$ =Diabetes-yes*age	0.004	1.004	0.146
$X_{15} = X_1 \times X_2 \times X_3$ =Hypertension-yes*cholesterol-yes*heart-problems-yes	0.128	1.136	0.845
$X_{16} = X_1 \times X_4 \times X_3$ =Hypertension-yes*diabetes-yes*heart-problems-yes	1.619	5.046	0.000
$X_{17} = X_2 \times X_3 \times X_4$ =Cholesterol-yes*heart-problems-yes*diabetes-yes	-0.501	0.606	0.089
$X_{18} = X_1 \times X_2 \times X_5$ =Hypertension-yes*cholesterol-yes*age	0.014	1.014	0.001
$X_{19} = X_1 \times X_3 \times X_5$ =Hypertension-yes*heart-problems-yes*age	0.025	1.025	0.000
$X_{20} = X_1 \times X_4 \times X_5$ =Hypertension-yes*diabetes-yes*age	0.004	1.004	0.338
$X_{21} = X_1 \times X_2 \times X_3 \times X_4$ =Hypertension-yes*cholesterol-yes*heart-problems-yes*diabetes-yes	-3.149	0.043	0.000
$X_{22} = X_1 \times X_2 \times X_3 \times X_5$ =Hypertension-yes*cholesterol-yes*heart-problems-yes*age	-0.015	0.985	0.120
$X_{23} = X_2 \times X_3 \times X_4 \times X_5$ =Cholesterol-yes*heart-problems-yes*diabetes-yes*age	0.015	1.015	0.002
$X_{24} = X_1 \times X_3 \times X_4 \times X_5$ =Hypertension-yes*heart-problems-yes*diabetes-yes*age	-0.044	0.957	0.000
$X_{25} = X_1 \times X_2 \times X_3 \times X_4$ =Hypertension-yes*cholesterol-yes*heart-problems-yes*diabetes-yes*age	-0.015	0.985	0.120
$X_{26} = X_1 \times X_2 \times X_3 \times X_4$ =Hypertension-yes*cholesterol-yes*heart-problems-yes*diabetes-yes*age	-3.149	0.043	0.000
$X_{27} = X_1 \times X_2 \times X_3 \times X_4 \times X_5$ =Hypertension-yes*cholesterol-yes*heart problems-yes*diabetes-yes*age	0.055	1.057	0.000

The five-factor interaction model indicates that modifiable risk factors interact, evidenced by a significant positive coefficient of 0.055 for variable X_{27} see table 5. This means individuals with a combination of hypertension, diabetes, heart problems, cholesterol and older ages are much more at risk of suffering stroke than those without this combo in this South African population. Lower-level interactions and covariates become important in arriving at this 5-way interaction of covariates (Table 5).

Table 5. MCMC Logit model for modifiable risk factors of stroke

Variable	Mean of β_j	95% Credible Interval of β_j
Intercept	-1.055	-1.065 , -1.045
Hypertension-yes	1.750	1.690 , 1.760
Cholesterol-yes	2.032	2.001 , 2.052
Heart problems-yes	1.535	1.225 , 1.640
Diabetes-yes	1.440	1.100 , 1.540
Age	-0.007	-0.009 , -0.006
Hypertension-yes*cholesterol-yes	-1.080	-2.00 , -1.070
Hypertension-yes*heart problems-yes	-0.320	-0.420 , -0.150
Hypertension-yes*diabetes-yes	-1.464	-1.744 , -1.344
Diabetes-yes*cholesterol-yes	-1.990	-2.000 , -1.898
Diabetes-yes*heart problems-yes	-1.224	-1.312 , -1.115
Hypertension-yes*age	-0.009	-1.10 , -0.006
Cholesterol-yes*age	-0.018	-0.025 , -0.010
Heart problems-yes*age	-0.039	-0.059 , -0.029
Diabetes-yes*age	-0.003	-0.100 , -0.001
Hypertension-yes*cholesterol-yes*heart problems-yes	0.129	-0.110 , 0.229
Hypertension-yes*diabetes-yes*heart problems-yes	1.620	1.430 , 1.820
Cholesterol-yes*heart problems-yes*diabetes-yes	-0.500	-0.650 , -0.401
Hypertension-yes*cholesterol-yes*age	0.016	0.113 , 0.118
Hypertension-yes*heart problems-yes*age	0.027	0.017 , 0.035
Hypertension-yes*diabetes-yes*age	0.003	-0.029 , 0.006
Diabetes-yes*cholesterol-yes*age	0.032	0.022 , 0.042
Diabetes-yes*heart problems-yes*age	0.040	0.019 , 0.041
Cholesterol-yes*heart problems-yes*diabetes-yes*age	0.020	0.015 , 0.025
Hypertension-yes*heart problems-yes*diabetes-yes*age	-0.043	-0.054 , -0.033
Hypertension-yes*cholesterol-yes*heart problems-yes*age	-0.015	-0.025 , -0.010
Hypertension-yes*cholesterol-yes*heart problems-yes*diabetes-yes	-3.150	-4.150 , -2.150
Hypertension-yes*cholesterol-yes*heart problems-yes*diabetes-yes*age	0.060	0.0426 , 0.065

The means of the posterior parameters are very close to the classical estimates obtained in table 4 using maximum likelihood estimates. This is because of the use of the non-informative prior. The 95% Credible intervals are also given showing that zero is not included in all intervals except for the interaction variable hypertension-yes*cholesterol-yes*heart problems-yes (Table 5). The uncertainty in the parameter estimates are expressed in the range of the credible intervals. The parameter could be zero in the case of the parameter for the interaction variable hypertension-yes*cholesterol-yes*heart problems-yes as zero is included in the credible interval.

Discussion

The study findings showed that strokes were attributable to established modifiable risk factors and could be prevented by an early intervention such as regular screening and treatment of hypertension, cholesterol, heart problems and diabetes. The significant modifiable risk factors in SA are diabetes, hypertension, heart problems and Cholesterol. All the covariates interact and together with older age, increase the risk of developing a stroke. The results of this study are consistent with a Jordanian study which identified hypertension, diabetes and heart problems as the most common risk factors of ischemic stroke [21].

Boehme et al [4] also found that hypertension, diabetes and cholesterol as the most important modifiable risk factors for stroke due to obesity and physical inactivity and the incidence of stroke in hypertensive people increase with ageing. In SA hypertension is the most prevalent modifiable risk factor of stroke. The prevalence of hypertension increases with age in black South Africans mainly due to: high urbanisation with adoption of Westernised food and lifestyles leading to obesity, bad dietary habits and physical inactivity [22]. Other possible reasons for high prevalence in hypertension leading to more strokes could be excessive salt intake, genetic factors and alcohol consumptions in black South Africans [23]. Hornsten et al [9] also found that high blood pressure as the major risk factor for stroke in their cohort study due to social changes associated with aging. Similar findings on hypertension being the most prevalent modifiable risk factor for stroke was reported in an American study [11]. It is therefore important for early detection and treatment of hypertension in SA to reduce the burden of stroke. Future trials focusing on treating blood pressure at earlier stages are urgently needed in SA.

There were some limitations to this study. Risk factors such as stress, HIV/AIDS, smoking and alcohol consumption were not captured in patients' records. Modifiable factors such as smoking, obesity and physical inactivity were not available in the data, yet they are important predictors of stroke. The non-modifiable predictors such as race, gender and family history of stroke were not examined since the data set did not include this information. Further studies could consider including all risk factors of stroke and examining the non-modifiable risk factors for more comprehensive evaluation of stroke predictors. Nevertheless, the strengths of this hospital-based study lie in that all cases of stroke are included for the specific period 2014 to 2018. Also, recent data set has been used without missing information. To date, this is the only comprehensive study on modelling modifiable predictors of stroke using a Bayesian approach in SA.

Conclusions

The study identified diabetes, hypertension, heart problems and cholesterol as major predictors of stroke in SA. The study also established that individuals with a combination of hypertension, diabetes, cholesterol, heart problems and old ages were at higher risk of suffering stroke than those without this combo in this South African population. Identification and early treatment of modifiable predictors of stroke are needed in SA to reduce the burden of stroke.

Abbreviations

NCD: Non-communicable diseases; LMIC: Low and middle-income countries; CVD: Cardiovascular diseases; CT: Computed tomography; MRI: Magnetic resonance image; HDL: High-density lipoproteins; LDL: Low-density lipoproteins; OR: Odds ratio; ME: Measurement error; 95% CIs: 95% Confidence intervals; MCMC: Markov Chain Monte Carlo; SSA: Sub-Saharan Africa; SA: South Africa.

Acknowledgments

We would like to thank the authorities of respective hospitals which supplied their data and for their tremendous support during this study. The authors thank the National Research Foundation (NRF) for funding this project.

Authors contributions

LM contributed towards the conceptualization of the study, study design, literature search, collected data and prepared it for analysis, analysed data and interpreted the results and manuscript write-up. DC critically reviewed and corrected misconceptions in the final version of the revised manuscript and approved the final manuscript.

Funding

This research study received funding from the National Research Foundation (NRF) South Africa.

Availability of data and materials

The datasets used or analysed during the current study are not available from the corresponding author to share with the public because the hospital managers do not permit it for ethical reasons as agreed in advance.

Ethical approval and consent to participate

Ethical clearance was obtained from the committee of research on human subjects of the University of South Africa as well as the study hospitals. The ethical clearance reference number is 2017/SSR-ERC/001. Permission to retrieve data from patients' records was obtained from the hospital managers. At hospital level written informed consent was obtained from the hospital managers as well.

Consent for publication

Not applicable. No individual person's personal details, images or videos are being used in this study.

Competing interests

The authors declare that they have no competing interests.

Author Details

Department of Statistics, College of Science, Engineering and Technology, Florida Campus,
Roodeport, Johannesburg, University of South Africa, South Africa.

References

1. Owolabi MO, Sarfo F, Akinyemi R, Gebregziabher M, Akpa O, Akpalu A, et al. Dominant modifiable risk factors for stroke in Ghana and Nigeria (SIREN): a case-control study. *The Lancet Global Health*. 2018;6:e436–46.
2. Adeloye D. An Estimate of the Incidence and Prevalence of Stroke in Africa: a systematic review and meta-analysis. *PLoS ONE*. 2014;9:e100724.
3. Owolabi M, Akarolo-Anthony S, Akinyemi R, Arnett D, Gebregziabher M, Jenkins C, et al. The burden of stroke in Africa: a glance at the present and a glimpse into the future: a review article. *CVJA*. 2015;26:S27–38.
4. Boehme AK, Esenwa C, Elkind MSV. Stroke risk factors, genetics, and prevention. *Circ Res*. 2017;120:472–95.
5. Maredza M, Bertram MY, Tollman SM. Disease burden of stroke in rural South Africa: an estimate of incidence, mortality and disability-adjusted life years. *BMC Neurol*. 2015;15:54.
6. Cui Q. Modifiable and non-modifiable risk factors in ischemic stroke: a meta-analysis. *Afr H Sci*. 2019;19:2121.
7. Choudhury MJH, Chowdhury MTI, Nayeem A, Jahan WA. Modifiable and non-modifiable risk factors of stroke: a review update. *J Natl Inst Neurosci Bangladesh*. 2015;1:22–6.
8. Kulshreshtha A, Anderson LM, Goyal A, Keenan NL. Stroke in South Asia: a systematic review of epidemiologic literature from 1980 to 2010. *Neuroepidemiology*. 2012;38:123–9.
9. Hörnsten C, Weidung B, Littbrand H, Carlberg B, Nordström P, Lövheim H, et al. High blood pressure as a risk factor for incident stroke among very old people: a population-based cohort study. *Journal of Hypertension*. 2016;34:2059–65.
10. Bos MJ, Koudstaal PJ, Hofman A, Ikram MA. Modifiable etiological factors and the burden of stroke from the Rotterdam Study: a population-based cohort study. *PLoS Med*. 2014;11:e1001634.
11. Romero JR, Morris J, Pikula A. Review: Stroke prevention: modifying risk factors. *therapeutic advances in cardiovascular disease*. 2008;2:287–303.
12. Alharbi M, Alharbi A, Alamri M, Alharthi A, Alqerafi A, Alharbi M. Ischemic stroke: prevalence of modifiable risk factors in the Saudi population. *IJMDC*. 2019;:601–3.
13. Jongen VW, Lalla-Edward ST, Vos AG, Godijk NG, Tempelman H, Grobbee DE, et al. Hypertension in a rural community in South Africa: what they know, what they think they know and what they recommend. *BMC Public Health*. 2019;19:341.

14. Fekadu G, Chelkeba L, Kebede A. Risk factors, clinical presentations and predictors of stroke among adult patients admitted to stroke unit of Jimma university medical center, south west Ethiopia: prospective observational study. *BMC Neurol.* 2019;19:187.
15. Li X, Pang J, Li M, Zhao D. Discover high-risk factor combinations using Bayesian network from cohort data of National Stroke Screening in China. *BMC Med Inform Decis Mak.* 2019;19:67.
16. Muff S, Riebler A, Held L, Rue H, Saner P. Bayesian analysis of measurement error models using integrated nested Laplace approximations. *J R Stat Soc C.* 2015;64:231–52.
17. Gray C. Use of the Bayesian family of methods to correct for effects of exposure measurement error in polynomial regression models. 2018. doi:10.17037/PUBS.04649757.
18. Hatano S. Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ,* 1976;54:541–53.
19. NDoH. National Department of Health. Uniform patient fee schedule for paying patients attending public hospitals. UPFS; 2009. <http://www.doh.gov.za/programmes>. Accessed 9 Nov 2020.
20. Roy V. Convergence Diagnostics for Markov Chain Monte Carlo. *Annu Rev Stat Appl.* 2020;7:387–412.
21. Qawasmeh MA, Aldabbour B, Momani A, Obiedat D, Alhayek K, Kofahi R, et al. Epidemiology, Risk factors, and predictors of disability in a cohort of Jordanian patients with the first ischemic stroke. *Stroke Research and Treatment.* 2020;2020:1–9.
22. Wandai M, Aagaard-Hansen J, Day C, Sartorius B, Hofman KJ. Available data sources for monitoring non-communicable diseases and their risk factors in South Africa. *S Afr Med J.* 2017;107:331.
23. Seedat Y. Hypertension in black South Africans. *J Hum Hypertens.* 1999;13:97–103.