Int J Disabil Hum Dev 2020;19(2):173-185 ISSN: 2191-1231

Modelling predictors of stroke disease in South Africa: Bayesian binary quantile regression approach

Lyness Matizirofa*, **BSc, BSc Hons, MSc**
Department of Statistics, University of South Africa, South Africa

Abstract

Stroke is currently the second prevalent cause of death and disability worldwide. South Africa (SA) is experiencing an epidemiological transition due to socio-demographic and lifestyle changes leading to an increase of non-communicable diseases, which in turn may result in an upswing of stroke cases. Modifiable predictors cause most strokes. The purpose of this paper is to address two important gaps in the stroke disease literature that is identifying and modelling predictors of stroke and estimating linear quantile models when predictors are measured with error. Methods: A hospital-based cross-sectional study design was used to model the predictors of stroke incidences in SA. We estimated posterior marginal by integrated nested Laplace approximations (INLA) for latent Gaussian models. The main objective of this study is to assess the effects of predictors of stroke for different quantiles for adults stroke patients and to estimate linear quantile regression models when predictors are measured with error. We used Bayesian quantile regression (BQR) methods. BQR was applied to stroke data collected between 2014 and 2018 in SA. The study considered lower, central and upper quantiles. Results: The study findings showed that stroke and modifiable risk factors were significantly associated with ($p < 0.0001$). The prevalence of stroke increased with cholesterol, hypertension, diabetes and heart-problem (OR 1.29, 1.33, 2.92 and 1.27) respectively. Modifiable and non-modifiable predictors had significant impact on stroke across quantiles. Conclusions: Most strokes were due to modifiable risk factors. Study findings showed significant impact of each predictor on stroke across quantiles.

Keywords: Stroke, Bayesian quantile regression, modifiable and non-modifiable predictors, South Africa

Introduction

Stroke is currently the second prevalent cause of death worldwide, and the leading cause of disability in adults (1). African countries are experiencing an epidemiological transition due to sociodemographic and lifestyle changes leading to an increase of non-

* ***Correspondence:*** Lyness Matizirofa, BSc, BSc Hons, MSc, PhD Student, Department of Statistics, College of Science, Engineering and Technology, University of South Africa, South Africa.
Email: 58528555@mylife.unisa.ac.za; lynessmatizirofa@gmail.com

communicable diseases yet the dominant risk factors are not yet established (2). Global estimates of stroke burden suggest that sub-Saharan has the highest incidence of 316 per 100,000 people per year (2). In South Africa (SA) it is among the top ten leading causes of disability in the country and accounting for 25,000 deaths yearly, it is also the major contributor to morbidity and mortality in SA (3). Based on literature, majority of stroke risk factors are preventable but little if not none has been done on modelling stroke risk factors. Reducing the burden of stroke in population requires identification of modifiable risk factors (4). Hence, this study has identified and modelled stroke risk factors in an effort to try and reduce the burden of stroke in SA.

Stroke risk factors can be classified as modifiable and non-modifiable. The modifiable risk factors are controllable risk factors of stroke that are of paramount importance, as intervention strategies aimed at reducing these factors can subsequently reduce the risk of stroke (4). These factors include hypertension, diabetes, cholesterol, smoking, obesity and atrial fibrilation. Hypertension is known as the most important risk factor for stroke (4-6, 8). There is evidence that a high proportion of hemorrhagic stroke (HS) is caused by hypertension (4). There is no doubt the burden of stroke globally due to modifiable risk factors.

The primary objectives of this study are to identify, quantify and model predictors of stroke incidences in SA. The study utilizes private and public hospital data retrieved from the nine provinces of SA, namely: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West, and Western Cape. The study data was collected between January 2014 and December 2018 from the randomly selected hospitals.

Non-modifiable risk factors are not preventable and it includes age, gender, race and genetics, stroke is known as disease of aging, thus the incidence increases with age (4). Several studies also reported that the incidence of stroke doubles after the age of 55 years (4-7). Recent evidence suggests, however, that the incidence of stroke increases in different racial disparities (4, 8). Black race has been identified as a racial group strongly affected with hypertension (8-9). The racial disparities in stroke incidence is highlighting the importance of stroke prevention interventions aimed at minority groups. Further gender has been identified as significant stroke risk factor. It has been indicated that women are high risk due to hormonal factors such as use of hormonal contraceptives and also longer lifespan of women (4). However, some studies showed that men being at higher risk of stroke because of unhealthy life styles such as smoking and heavy alcohol consumption (6). Therefore, comprehensive studies need to be undertaken to raise awareness on risk factors of stroke. Having precise information of stroke incidence and its associated risk factors of stroke incidences from hospital-based data is important to develop public health interventions such as campaign awareness on risk factors and possible measures to reduce the burden of stroke. This necessitates the carrying out of modelling the predictors of stroke disease in SA to obtain such critical information.

Several statistical techniques such as logistic regression have been used to analyze risk factors for different diseases (5, 6, 10). In developed countries, research on risk factors were extensively studied whilst the dominant risk factors of stroke in Africa are not yet clearly established (3). The most plausible strategy to reduce the burden of stroke in Africa requires identification of modifiable risk factors. Logistic regression analysis focuses on the mean only. However, in order to fill this gap this study modelled the predictors of stroke disease through Bayesian quantile regression. Numerous statistical techniques have been used to examine effect of covariates in diseases occurrence, but this paper employs a specific area of application; Bayesian quantile regression modelling approach for cross-sectional data collected collected between January 2014 and December 2018.

Several studies, have done Bayesian quantile regression with continuous dependent variables using independently distributed asymmetric Laplace densities (ALD) (11-13). Little if not none has been done on binary Bayesian quantile regression for stroke disease. This study consider the standard binary regression model: $y_i = I(y_i^* \geq 0) = x_i^{\tau} \beta + \epsilon_i$, where y_i is the indicator of the *i th* individual's response determined by the underlying variable y^*, x_i is a k x 1 vector of explanatory variables, β is a *k x1* vector regression coefficients, ϵ_1 is a random error term and I = *1, n*. Binary quantile regression models was developed to identify predictors of stroke disease. Modelling the conditional quantile functions of predictors such as gender, age, race, diabetes, cholesterol, hypertension and heart-problems had been done. Understanding the association between stroke and its predictors will allow researchers and policy makers to understand points along with the distribution of intervention in SA. The purpose of this paper is to address two important gaps in the stroke disease literature that is identifying and modelling predictors of stroke and estimating linear quantile models when predictors are measured with error.

Methods

Stroke was defined according to the World Health Organisation (WHO) criteria 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 (15). Diagnosis of stroke was done according to the International classification of Diseases, Revision 10 (ICD 10) codes in the categories of 160, 161, 162, 163, and 164. Diagnosis was mainly based on Magnetic Resonance Imaging (MRI) or Computed Tomography (CT).

Measures

The response variable was confirmed stroke coded, 1 = yes and 0 = no. The independent variables were demographic characteristics (marital – status, employment status and residence), non-modifiable risk factors (age, race, and gender), modifiable risk factors (hypertension, cholesterol, heart-problems and diabetes) and type of stroke (ischemic stroke (IS) and hemorrhagic stroke (HS)).

Study setting

This study was carried out in SA. The estimate of mid-year population in South Africa is 57.73 million (17). South Africa comprises nine provinces as shown in Table 1. The following table shows the population estimates by province.

Table 1. Mid-year population estimates for South Africa by province, 2018

Province	Population
Gauteng	14 717 000
KwaZulu-Natal	11 384 700
Western Cape	6 621 100
Eastern Cape	6 522 700
Limpopo	5 797 300
Mpumalanga	4 523 900
North West	3 979 000
Free State	2 954 300
Northern Cape	1 225 600
Total	57 725 600

There are approximately 407 public hospitals and 203 private hospitals in SA (16). This study, randomly selected 55% of the 203 private hospitals and 45% of the 407 public hospitals across nine provinces of SA. A Stratified probability sampling technique was used to calculate the proportions accordingly.

Data collection

A data retrieval sheet was developed with the help of expects in the field. The data collection instrument was validated before use. The data retrieval sheet was formulated with all the study variables include; demographic characteristics; confirmation of stroke disease(CT/MRI); stroke type; mode of admission (inpatient/outpatient); risk factors of stroke; direct stroke costs and allied services such as occupational therapy, speech therapy and physio therapy.

Data were retrieved from 101 private hospitals and 203 public hospitals. The case managers for the sampled hospitals assisted with data retrieval. The total number of stroke patients including none confirmed strokes was 35 730.

Statistical analysis

A hospital based cross-sectional study design was used to model the predictors of stroke disease incidences in SA. Quantile regression was used in this study because it more appropriate in many situations than mean regression. Further, quantile regression methods provide more complete description of functional changes than focusing solely on the mean and it provides more comprehensive information on the relationship between the outcome variable and the covariates than the classical mean regression. Quantile regression models were developed because it leads to richer view of how covariates influence the response (18, 19). The quantiles used for the present study were based on Ji et al. study (20). Study quantile were lower (10%, 25%) central location (50%) and upper (75%, 95%) quantiles. This choice allowed making comparison of the effects of predictors of stroke in the lower and upper quantiles. Lastly, it can estimate the conditional quantile function of the response, it offers some extended flexibility of covariates effects and covariates may have different effects on different quantiles. In this paper, binary Bayesian quantile regression parameter estimation methods were used. Bayesian quantile regression methods because it provides complete statistical inference and decision making in the event of uncertainty, lastly works very well with dichotomous response data. Bayesian quantile regression was used because it allows the estimated coefficients to change over quantiles and it explains the impact of each independent variable on different parts of the distribution (13, 21, 22). In particular, the integrated nested Laplace approximations (INLA) for latent Gaussian models was used in this study (13). The main objective of this study is to assess the effects of predictors of stroke for different quantiles for stroke patients aged 18 years and above and to estimate linear quantile regression models when predictors are measured with error. To account for measurement error in explanatory variables, Bayesian approaches provides flexible framework. INLA provides accurate approximations by avoiding time-consuming sampling and INLA computational approach directly approximates the posterior of interest with closed form expression, thus it does not have problems of convergence and mixing (13,21). Due to its flexibility, in this paper used INLA for Bayesian binary quantile regression modelling of predictors of stroke. Bayesian analysis of stroke predictors models using INLA was carried out to data collected from private and public hospitals in SA. To date, there are no published research on on Bayesian quantile regression for binary dependent variables in particular stroke disease. This paper provides quantile inference approach through Bayesian modelling approach.

Data analysis was carried out with a statistical software R 3.5.4 (http: cran.r-project.org), using the quantreg package. Descriptive statistics were used to summarize demographic information and risk factors of stroke. Continuous variables were presented as the mean and standard deviation whilst categorical variables were expressed frequencies and percentages. The chi-square test was used to test the association between stroke and categorical variables. Additionally, R-INLA package was used for Bayesian modelling predictors of stroke, this package was used to calculate accurate approximation of the marginal posterior densities of hyper parameters. This package supports the binary quantile regression modelling and calculated the fitted probabilities for each vector of predictors.

Ethical considerations

Permission to conduct this research was obtained from the provincial departments and from individual hospitals. The research was granted permission by the committee of research on human subjects of the University of South Africa as well as the study hospitals. The ethical clearance number is 2017/SSR-ERC/001.

Results

Demographic characteristics of stroke patients are shown in Table 2. Of the 35,730 patients, 21,453 were confirmed strokes. Most of the patients were in the age group 18-54 years, 19,474/100,000 person years which is an indication of young strokes. Study findings showed more females 18,145 (50.8%) than males 17,585 (49.2). The majority of the stroke patients were married (64.4%) and approximately 70% were unemployed. The proportion of patients with hypertension was found to be higher among females (55.3%) A significantly greater proportion of coloured patients were diabetic (62.1%). A larger proportion of whites suffered stroke due to heart-problems (54.4%). The prevalence of hypertension was 19,756/100,000 person years. Lastly, the prevalence of IS (77.1%) was higher than the HS (22.9%).

The three major modifiable risk factors for this study were diabetes (62.1%), hypertension (55.3%) and heart-problems (54.4%). Table 2 indicates that hypertension, cholesterol, heart-problems, and diabetes as significantly associated with ischemic stroke ($p < 0.0001$) respectively. Differently, female gender was not a significant risk factor for ischemic ($p = 0.36$). Most Indians suffered ischemic stroke due to cholesterol (47.4%). The non-modifiable risk factors such as age groups 18-54 years and 55-75 years and black, white and Indian races were significantly associated with ischemic stroke ($p < 0.0001$) respectively. Besides modifiable and non-modifiable risk factors being associated with stroke, demographic characteristics were also significantly associated with stroke ($p < 0.0001$) respectively.

Table 2. The χ^2 test association between stroke and patient characteristics, potential risk factors

Variable	Study sample (N = 35730)	Confirmed stroke (N = 21453)	P -value
Gender			
Male	17585 (49.2)	10516 (49.0)	0.36
Female	18145 (50.8)	10937 (51.0)	
Age			
18-54	19474 (54.5)	12579 (58.6)	< 0.0001
55-75	10446 (29.2)	5845 (27.2)	
76 -98	5810 (16.3)	3029 (14.1)	
Race			
Black	10560 (29.6)	7821 (13.1)	< 0.0001
White	12243 (34.3)	4225 (19.7)	
Indian	5203 (14.6)	1082 (5.0)	
Coloured	3351 (9.4)	5203 (24.3)	0.97
Asian	4373 (12.2)	3122 (14.6)	
Marital status			
Single	6669 (18.7	2307 (10.8)	< 0.0001
Married	23009 (64.4)	15192 (70.8)	
Divorced	1278 (3.6)	660 (3.1)	
Widow	3777 (10.6)	2935 (13.7)	
Unknown	997 (2.8)	359 (1.8)	
Employment status			
Employed	8883 (24.9)	6575 (30.6)	< 0.0001
Unemployed	24879 (69.6)	14522 (67.7)	
Retired/pensioner	1968 (5.5)	356 (1.7)	
Residence			
Rural	1226 (3.4)	671 (3.1)	< 0.0001
Urban	33770 (94.5)	20493 (95.5)	
Informal settlement	734 (2.1)	289 (1.3)	
Hypertension			
Yes	19756 (55.3)	11266 (52.5)	< 0.0001
No	15974 (44.7)	10187 (47.5)	
Cholesterol			
Yes	16923 (47.4)	10698 (49.9)	< 0.0001
No	18807 (52.6)	10755 (50.1)	
Heart-problems			
Yes	19453 (54.4)	11179 (52.1)	< 0.0001
No	16277 (45.6)	10274 (47.9)	
Diabetes			
Yes	22183 (62.1)	15476 (72.1)	< 0.0001
No	13547 (37.9)	5977 (27.9)	
Type of stroke			
Ischemic	25770 (77.1)	16146 (75.3)	< 0.0001
Hemorrhagic	8180 (22.9)	5307 (24.7)	

Table 3 shows that the prevalence of ischemic stroke was insignificantly increased among females (OR = 1.31, 95% CI 1.24-1.37, p < 0.0001; OR = 1.02, 95%CI 0.98-1.06, p = 0.36). The likelihood of stroke was significantly increased among stroke patients aged between 18-54 years and 55-75 years (OR 1.68, 95% CI 1.06-1.78, p = 0.001; OR 1.17, 95% CI 1.09- 1.24, p < 0.0001) respectively. Thus, both age groups had increased risk of ischemic stroke. Additionally, the likelihood of stroke was significantly increased among Indian patients with elevated cholesterol level (OR1.29; 95% CI 1.23-1.34, p < 0.0001). The risk of ischemic stroke was significantly high among blacks with hypertension (OR 1.14, 95% CI 1.06-1.24, p < 0.0001; OR 1.33, 95% CI 1.27-1.39, p < 0.0001). Moreover, the risk of ischemic stroke was increased among whites, Indians, and coloureds (OR 1.30, 95% CI 1.20-1.33, p < 0.0001; OR 0.19, 95% CI 0.17-0.21, p < 0.0001; and OR 2.35, 95%CI 1.17-3.24, p = < 0.0001) respectively. Coloureds were likely to suffer ischemic stroke due to diabetes than the basis (Asians). Lastly the risk of ischemic stroke was significantly high among whites with heart-problems (OR 1.30, 95%CI 1.20-1.33, p < 0.0001). The present study shows that the modifiable risk factors such as hypertension, cholesterol, diabetes, and heart-problems (p < 0.0001) respectively were significantly associated with ischemic stroke. The risk of ischemic stroke was 1.3 times higher than hemorrhagic stroke.

Table 3. Bivariate logistic regression analysis of each predictor and stroke

Variables	Total	n (%)	Odds Ratio OR	95% CI for OR	p-value
Gender					
Male	17585	17585 (49.2)	Ref		
Female	18145	18145 (50.8)	1.02	0.98 ─ 1.06	0.36
Total	35730				
Hypertension					
Yes	19756	19756 (55.3)	1.33		
No	15974	15974 (44.7)	Ref	1.27 ─ 1.39	< 0.0001
Total	35730				
Cholesterol					
Yes	16923	16923 (47.4)	Ref		
No	18807	18807 (52.6)	1.29	1.23 ─ 1.34	< 0.0001
Total	35730				
Heart-problems					
Yes	19453				
No	16277	19453 (54.4)	1.27	1.21 ─ 1.32	< 0.0001
Total	35730	16277 (45.6)	Ref		
Diabetes					< 0.0001
Yes	22183	22183 (62.1)	2.92	2.80 ─ 3.10	
No	13547	13547 (37.9)	Ref		
Total	35730				
Age (years)					
18 ─ 54	19474	19474 (54.5)	1.68	1.06─ 1.78	0.001
55 ─ 75	10446	10446 (29.2)	1.17	1.09─ 1.24	< 0.0001
76 – 98	5810	5810 (16.3)	Ref		< 0.0001
Total	35730				
Race					
Black	10560	10560 (29.6)	1.14	1.06 ─ 1.24	0.001
White	12243	12243 (34.3)	1.30	1.20 ─ 1.33	< 0.0001
Indian	3351	3351 (9.4)	2.35	1.17 ─ 3.21	< 0.0001
Coloured	5203	5203 (14.6)	2.70	1.80 ─ 3.34	0.97
Asian	4373	4373 (12.2)	Ref		
Total	35730				
Type of stroke					
Ischemic	25770	25770 (77.1)	1.31	1.24 ─ 1.37	< 0.0001
Hemorrhagic	8180	8180 (22.9)	Ref		
Total	35730				

Multiple logistic regression analysis was carried out to investigate the relationship between stroke incidence and the set of seven predictors simultaneously. The predictors were gender, age, hypertension, cholesterol, heart-problems, diabetes and race. Table 4 depicts that the likelihood of ischemic stroke was significantly increased among Females with hypertension, cholesterol, heart-problems and diabetes (OR 1.11 ,95% CI 1.05-1.17, $p < 0.0001$; OR 1.69 ,95% CI 1.60-1.79, $p < 0.0001$; OR 2.21, 95% CI 2.08-2.36, $p < 0.0001$; OR 4.76, 95% CI 4.48-5.06, $p < 0.0001$) respectively had increased risk of developing ischemic stroke. Further, the likelihood of stroke was significantly increased among stroke patients aged between 18-54 years and 55-75 years, with heart-problems and elevated hypertension (OR 1.32, 95%CI 1.28-1.36, $p < 0.0001$; OR 1.35, 95%CI 1.29-1.42, $p < 0.0001$and OR 1.69, 95% CI 1.60-1.79, $p < 0.0001$) respectively.

Bayesian binary quantile regression analysis was carried out to investigate the impact of each predictor of stroke incidence in different quantiles. Table 5 shows that; age-groups 18-54 years and 55-75 years had positive and significant effect on stroke across quantiles. This is supported by 95% credible intervals not including zero (0.1 ,95% CI: 0.460 to 0.568; 0.25 ,95% CI 0.382 to 0.450; 0.50, 95% CI 0.208 to 0.331; 0.75, 95% CI 0.211to 0.364 and 0.95 95% CI 0.214 to 0.325). These findings portray that the impact of age groups 18-54 years and 55-75 years on ischemic stroke was stronger at the lower quantiles than at the upper quantiles. The magnitude of the association between ischemic stroke and age increases from low to high quantiles and was constant at the upper quantiles with posterior mean of 0.011.

Further, female gender showed that posterior means were positive across quantiles. Thus, women with elevated hypertension, cholesterol, heart-problems and diabetes were likely to develop stroke across quantiles than men in SA. Female-gender had a smaller effect on ischemic stroke at the upper quantiles and bigger effect at the lower quantiles. The estimated conditional quantile functions decreased from 0.386 at 10th quantile to 0.321 at 25th quantile 0.185 at the central location and 0.071 in the 75th quantile. Consequently, the magnitude of the association for stroke with female-gender increases from high to low quantiles. This indicates high impact on ischemic stroke from females. The impact of Indian- race on ischemic stroke was significantly stronger at the upper quantile than lower quantiles. Also the effect of black- race was smaller at the lower quantiles. Largely, the effect of non-modifiable risk factors was significant across quantiles. Hypertension was positively related with ischemic stroke across quantiles. The hypertension estimated conditional quantile functions increased from lower quantiles to upper quantiles (see Table 6).

Table 4. Multivariate logistic regression analysis for non-modifiable and modifiable stroke predictors

Variables	OR	95% CI for OR	p-value
Female-gender	1.11	1.05 – 1.17	< 0.0001
Age-group 55-75 years	1.32	1.28 – 1.36	< 0.0001
Age-group 76-98 years	1.36	1.20 – 2.36	< 0.0001
Hypertension	1.69	1.60 – 1.79	< 0.0001
Cholesterol	2.21	2.08 – 2.36	< 0.0001
Heart-problems	1.35	1.29 – 1.42	< 0.0001
Diabetes	4.76	4.48 – 5.06	< 0.0001
Black-race	2.43	2.24 – 2.64	< 0.0001
White-race	2.35	1.63 – 2.38	< 0.0001
Indian-race	1.17	1.05 – 1.20	< 0.0001
Coloured-race	5.89	2.90 – 7.90	0.97

Table 5. Binary quantile regression estimates and some predictors of stroke quantiles model

Quantiles	Constant	Age-group 18-54 years	Age-group 55-75 years	Female-gender	White-race	Black-race	Indian-race	Coloured-race
0.10	-1.03 [-1.096 ,-0.964]	0.530 [0.460 , 0.568]	0.013 [0.012 , 0.014]	0.386 [0.347, 0.424]	0.098 [0.076 , 0.120]	0.182 [0.112 , 0.253]	0.558 [0.495,0.560]	0.516 [0.444, 0.580]
0.25	-0.848 [-0.915 , -0.781]	0.447 [0.382 , 0.450]	0.012 [0.011 , 0.013]	0.321 [0.282 , 0.360]	0.053 [0.030, 0.076]	0.198 [0.127 , 0.270]	0.450 [0.387 ,0.454]	0.498 [0.427 , 0.500]
0.50	-0.621 [-0.686 , -0.555]	0.270 [0.208 , 0.331]	0.010 [0.009 , 0.011]	0.185 [0.149 , 0.149]	0.059 [0.039 , 0.080]	0.240 [0.177 , 0.303]	0.200 [0.141 , 0.230]	0.405 [0.404 ,0.410]
0.75	-0.593 [-0.720 , -0.466]	0.288 [0.211 , 0.364]	0.011 [0.010 , 0.013]	0.071 [0.021 , 0.121]	0.185 [0.164, 0.205]	0.364 [0.301 , 0.427]	0.668 [0.593 , 0.743]	0.094 [0.023 , 0.164]
0.95	-1.439 [-1.544 , -1.330	0.275 [0.214 , 0.335]	0.011 [0.010 , 0.013]	0.068 [0.024 , 0.112]	0.161 [0.144, 0.179]	0.315 [0.258 , 0.371]	0.566 [0.506 , 0.625]	0.117 [0.057 ,0.176]

Notes: Credible Confidence interval of 95%, below parameters. The number of equivalent replicates = 4444.95.

Table 6. Binary quantile regression estimates and some predictors of stroke model

Quantiles (τ)	Constant	Cholesterol	Diabetes	Heart-problems	Hypertension
0.1	-1.03 [-1.096 , -0.964]	0.116 [0.048 , 0.184]	0.069 [0.009 , 0.129]	0.410 [0.349 , 0.449]	0.182 [0.112 , 0.253]
0.25	-0.848 [-0.915 , -0.781]	0.027 [-0.017, 0.042]	0.244 [0.182 , 0.305]	0.427 [0.364 , 0.464]	0.198 [0.127 , 0.227]
0.5	-0.621 [-0.686 , -0.555]	0.244 [0.183 , 3.04]	0.642 [0.586 , 0.697]	0.427 [0.347, 0.447]	0.240 0.178 , 0.303]
0.75	-0.575 [-0.720 , -0.466]	0.575 [0.514 , 0.636]	1.140 [1.07 , 1.205]	0.400 [0347 , 0.403]	0.364 [0.301 , 0.427]
0.95	-1.439 [-1.333 , -1.544]	0.482 [0.427 , 0.537]	1.006 [0.956 , 1.055]	0.342 [0.295 , 0.368]	0.315 [0.258 , 0.371]

Notes: Credible confidence interval of 95% below parameters. The number of equivalent replicates = 4444.95.

As shown in Table 6, cholesterol has positive significant estimated quantile regression estimates in upper quantiles. However, cholesterol effect was not significant at the 25th quantile (95% CI -0.017 to 0.042), evidenced by a 95% credible interval includes zero. The magnitude of higher levels of cholesterol was bigger in the central location and upper quantiles. These results imply that cholesterol effect on ischemic stroke was much larger at the upper end and central locations than the lower end. Additionally, diabetes had positive and significant conditional quantile functions across quantiles. Thus, the Bayes estimate effect of diabetes on ischemic stroke was larger at the upper and central locations than the lower locations. Heart-problems had a negative significant estimated conditional quantile functions with ischemic stroke across all quantiles evidenced by credible intervals, which does not include zero. Overall, the effect of modifiable risk factors on ischemic stroke were statistically significant across quantiles except cholesterol at the 25th quantile.

Discussion

Based on evidence from literature, this is the most recent research on Bayesian quantile regression analysis of modelling risk factors associated with stroke in SA. Study showed that the prevalence of ischemic stroke (77.1%) was higher than hemorrhagic stroke (22.9%). This is in line with a study by Gan et al. (6) that reported more ischemic strokes compared to hemorrhagic strokes. The dominant modifiable risk factors for this study were diabetes (62.1%), hypertension (55.3%) and heart-problems (54.4%). Several studies also reported hypertension as the major stroke risk factor (5, 8, 26-28,31). Similar findings were reported in a study done in Ghana and Nigeria that found that cholesterol, blood pressure and diabetes were associated with stroke (1). In this study the prevalence of hypertension was slightly higher in women than in men, but the difference was not statistically significant. This could be due to use of hormonal contraceptives by women (33). A study in Japan also found more female strokes due to sex hormones (30). Study findings showed that the prevalence rates of cholesterol, hypertension, heart-problems and diabetes was significantly associated with ischemic stroke in lower and upper quantiles. Largely, the study findings show that modifiable risk factors were significantly associated with stroke. This could be due to poverty in South Africa, which leads people buying cheap and unhealthy food and physical inactivity, also heavy smoking and alcohol consumption at a younger age (9, 22-23, 31). In agreement with this finding, Bos et al. (5) identified modifiable etiological factors such as high blood pressure, cholesterol, diabetes, obesity and smoking were found to be major stroke risk factors.

Despite, hypertension, cholesterol, and diabetes being the common risk factors of stroke from several studies, age, gender and race were identified as common non-modifiable risk factors of stroke (1). The major non-modifiable risk factors for present study were female gender, age groups 18-54 years and 55-75 years and lastly black, white, Indian and coloured races. The risk of stroke was high in young age group 18-54 years, possibly due to job stress since there is high rate of unemployment in South African young and they start smoking and drinking alcohol at young ages. Patients aged 18-54 years were 1.32 times likely to suffer ischemic stroke compared to 76-98 years' age group. Wang et al. (25) also reported young aged patients. It could be, the trend of old age stroke is now shifting to young people (22, 26-27, 31). Further studies are needed to understand the shift. Young South African suffer stroke due to hypertension, possibly due to a major problem of high blood pressure in South African children (28). Stroke patients 55 -75 years were 1.32 times increased risk of ischemic stroke. Similarly, numerous studies found that the risk of stroke doubles after 55 years of age (4, 28, 33). Since stroke is a heart disease, this could be due to cumulative effects of aging on the cardiovascular system.

Further, the risk of ischemic stroke was 1.02 times higher in females compared to males. Although men have a higher risk of stroke than women do in other populations, female gender was not a significant risk factor for stroke in this study. This finding is different from Chinese and American studies which reported increased risk of stroke in men than women (18, 28). In agreement, several studies found more prevalent strokes in men than in women (4, 28, 33). Men have higher risk of stroke compared to women, this is possibly attributed to bad behavioural lifestyles of heavy alcohol consumption, smoking and physical inactive in men. Interestingly, female black patients were 1.14 times likely to suffer ischemic stroke than other racial groups. Consistently, higher stroke rates were noted in Africans Americans, Hispanic Americans and the black race compared to white Americans (28). In the present study, Blacks with elevated hypertension were associated with increased risk of ischemic stroke (OR 1.14). This could be due to co-existing risk factors in races such as hypertension in blacks, cholesterol in Indians, heart-problems in whites and diabetes in coloureds. Other factors that may influence racial differences in stroke risk include other social determinants of disease and nativity (6). The racial disparities in stroke incidence is highlighting the importance of stroke prevention interventions aimed at minority groups.

Another important study finding were obtained from logistic regression analysis. Patients with hypertension were 1.33 times likely to suffer ischemic stroke than patients without hypertension. Congruently, a Chinese community – based cross-sectional study also found that patients with hypertension were 1.47 times more likely to suffer stroke than those without hypertension (6, 26). This study found hypertension as an important risk factor of ischemic stroke. As such, efforts to treat patients with elevated hypertension are necessary to reduce the risk of stroke. Additionally, stroke induced by heart-problems was 1.27 times higher than that of stroke patients without heart-problems. Lastly, patients who suffer stroke due to diabetes were 2.92 times higher than that of patients without diabetes. Reducing the burden of stroke requires identification of modifiable risk factors and raise awareness on their impact on ischemic stroke.

Several studies were done on modelling risk factors of stroke through logistic regression (1, 4, 6, 14), found stroke to be significantly increased in people with hypertension, smoking, bad dietary habits, diabetes mellitus, and cardiac causes. A recent study done in China also found that risk factors significantly associated with prevalence of stroke were increasing age, hypertension physical inactivity and male gender (6, 32). Differently, Hornsten et al (28) found that hazard ratios (HR) from univariate models were increased among women, people with high systolic blood pressure SBP, diabetes and current smokers. A study by Bos et al (5) established that people with stage II hypertension, diabetes mellitus, heavy smokers and obesity had increased hazard risk of stroke.

Notwithstanding several studies modelling risk factors of stroke through ordinary least squares logistic regression (OLS). The current study further modelled the predictors of stroke by using Bayesian binary quantile regression analysis. BQR analysis was carried out to investigate the impact of each predictor of stroke incidence in different quantiles. The magnitude of the association for stroke with age increases from low to high quantiles. This outcome entails that the age effect on stroke is smaller at the lower and upper end than the central locations. Thus the effect of age groups 18-54 years and 55-75 years was significant in both lower and upper quantiles. Thus, the effect of population aging has been recognized as a serious problem in several studies using logistic regression (25, 28, 33). Further, female gender had bigger effect on stroke at the lower quantiles and smaller effect at the upper quantiles. Supported by several OLS findings that revealed women with higher risk of stroke than men (6, 29, 32). The impact of black, white and Indian race on ischemic stroke was significantly stronger at the upper quantile than lower quantiles. Thus, the effect of black racial group is smaller at the lower quantiles. The Study results imply a major impact of female-gender and coloured race on stroke was higher at lower quantiles than upper quantiles, whilst age-groups 18-54 years and 55-75 years, black-race white-race and Indian –race were significantly stronger at the upper quantiles. This finding is consistent with other studies (6, 32).

Besides non-modifiable risk factors on stroke, modifiable factors showed significant impact on stroke across quantiles. The magnitude of higher levels of cholesterol was bigger in the central location and upper quantiles. These results imply that cholesterol effect on ischemic stroke was much larger at the upper end and central locations than the lower end. However, the impact of cholesterol was not statistically significant in the 25th quantile. Additionally, diabetes had positive and significant conditional quantile functions across quantiles. Thus, the Bayes estimate effect of diabetes on ischemic stroke was larger at the upper and central locations than the lower locations. In agreement with significant effect of modifiable risk factors on stroke, numerous OLS reported increased of stroke in patients with hypertension, cholesterol, heart-problems and diabetes (1, 4, 7-8, 26). Moreover, heart-problems had positive significant estimated conditional quantile functions with ischemic stroke across all quantiles. Lastly, the hypertension estimated conditional quantile functions increased from lower quantiles to upper quantiles since there are no studies on modelling risk factors of stroke with QR analysis, it's difficult to compare current study with previous studies. However, comparing present study BQR model results with OLS results similar risk factors of stroke were found to be significantly associated with stroke at all quantiles. Thus, richer information was obtained in the distribution of the response variable.

Conclusion

This study provides a comprehensive quantile regression analysis for the stroke predictors in South Africa. In summary, hypertension, cholesterol, diabetes, and heart-problems, were found to be significant modifiable risk factors of stroke at all quantiles. Whilst, female-gender, age groups 18-54 years and 55-75 years black, white, Indian and coloured races were significant non-modifiable risk factors across quantiles. The present study identified high prevalence of stroke among young adults aged between 18 and 54 years. However, modifiable risk factors such, as hypertension, cholesterol, heart-problems and diabetes were identified as the major risk factors in this study. These modifiable predictors may be responsible for these high incidences of stroke, and these factors are appropriate targets for population based stroke prevention campaigns. Health policy makers should place more emphasis on the effective prevention and control of stroke and its predictors. Further, these factors can be managed by avoiding gaining excessive weight, reducing salt, and sugar intake, and by exercising regularly. Hypertension was identified as the important risk factor for cardiovascular diseases such as stroke (26). Health programmes can increase awareness of hypertension and its causes, but hinge on the knowledge and perception of the targeted community. Therefore, this study investigated the impact of risk factors such as hypertension across quantiles for stroke in South Africa with the aim to increase awareness of hypertension and cardiovascular disease in the local population.

These study findings call for programs and policies meant to curb unhealthy lifestyles, promote physical activity and health nutrition among children. We recommend setting up of wellness centers at schools, tertiary institutions, work places and communities where people can be tested regularly for hypertension, cholesterol and diabetes level. Health professionals should raise awareness of the dangers of these risk factors. In addition, government may increase value added tax on cigarettes, alcohol and other unhealthy foods in an effort to discourage consumption of unhealthy food and smoking (9, 23, 26). This study also revealed that, the majority of stroke incidences were due to diabetes 22.18/1,000 person-years, hypertension 19.7/1,000 person -years and heart-problems 19.45/1,000 person-years.

Furthermore, we have demonstrated significant effects at all quantiles of stroke disease variation in the incidence of stroke with consideration of its predictor's factors of cholesterol, hypertension, and diabetes were very high across quantiles. The magnitude and intensity of the coefficients on predictors of stroke changes across the quantiles. Further studies based on these findings are needed to enhance our understanding of stroke and its association with predictors such as gender, hypertension, cholesterol race, heart-problems and age group 18 to 54 years.

Acknowledgments

I would like to thank the authorities of respective hospitals which supplied their data and for their tremendous support during this study. I am also thankful to Dr E Mhlanga, Mr T Mbara and Associate Professor E Ruhode for their editorial support and lastly Mr L Matizirofa for additional referencing support. The author declared no potential conflicts of interest with respect to the research, authorship, and or publication of this article. This research study received funding from the National Research Foundation (NRF) South Africa.

Ethical compliance

The authors have stated all possible conflicts of interest and all sources of funding for this work. If this work involved human participants, informed consent was received from each individual and it was conducted in accordance with the 1964 Declaration of Helsinki. If this work involved experiments with humans or animals, it was conducted in accordance with the related institutions' research ethics guidelines.

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. Lancet Glob Health 2018;6(4):e436–46.

[2] 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: review article. Cardiovasc J Afr 2015; 26(2):S27-38.

[3] 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.

[4] Boehme AK, Esenwa C and Elkind MSV. Stroke risk factors, genetics, and prevention. Circ Res 2017;120(3): 472-95.

[5] 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(4):e1001634.

[6] Gan Y, Wu J, Zhang S, Li L, Yin X, Gong Y, et al. Prevalence and risk factors associated with stroke in middle-aged and older Chinese: a community-based cross-sectional study. Sci Rep 2017;7(9501):1-7.

[7] Bahonar A, Khosravi A, FariborzKhorvash F, Maracy M, Saadatnia M. Seasonal and monthly variation in stroke and its subtypes-10year hospital-based study. Mater Sociomed 2017;29(2):119-23.

[8] O'Neal WT, Judd SE, Limdi NA, McIntyre WF, Kleindorfer DO, Cushman M, et al. Differential impact of risk factors in blacks and whites in the development of atrial fibrillation: the reasons for geographic and racial differences in stroke (REGARDS) Study. J Racial Ethn Health Disparities 2017;4(4):718-24.

[9] 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):1-10.

[10] Akahoshi K, Michiaki Kai M. Multiple logistic regression analysis on the health hceckup data and the lifestyle habits of medicated residents: A population-bohort based study. J Commun Public Health Nurs 2016;2:132.

[11] Benoit DF, Van den Poel D. Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution. J Appl Econom 2012;27(7):1174–1188.

[12] Muff S, Riebler A, Rue H., Saner P, Held L. Bayesian analysis of measurement error models using INLA. Methodology (stat.ME) 2013. arXiv:1302.3065 [stat. ME]:1-37.

[13] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol 2009;71(2):319–92.

[14] Divajeva D, Marsh T, Logstrup S, Kestens M, Vemer P, Kriaucioniene V, et al. Economics of chronic diseases protocol: cost-effectiveness modelling and the future burden of non-communicable disease in Europe. BMC Public Health 2014;14:456.

[15] Hatano S. Experience from a multicentre stroke register: a preliminary report. Bull World Health Organ 1976; 54(3550)541-53.

[16] Makombo, T. Fast facts: public health sector in need of an antidote". Fast Facts 2016;6(298):6.

[17] Stas SA. Mid-year population estimates - Media Release, 2018. URL: http://www.statssa.gov.za/?p = 11341.

[18] Koenker R, Hallock KF. Quantile regression. J Econ Perspect 2001;15(4):143-56.

[19] Koenker, R. Quantile regression. Cambridge: Cambridge University Press, 2005. URL: http://ebooks.cambridge.org/ref/id/CBO9780511754098.

[20] Ji Y, Lin N, Zhang B. Model selection in binary and tobit quantile regression using the Gibbs sampler. Comput Stat Data Anal 2012;56(4):827-39.

[21] Yu K, Moyeed RA. Bayesian quantile regression. Stat Probab Lett 2001;54(4):437–47.

[22] Reddy P, Meyer-Weitz A and Yach D. Smoking status, knowledge of health effects and attitudes towards tobacco control in South Africa. SAMJ 1996;86:1389-93.

[23] Ramsoomar L and Morojele NK. Trends in alcohol prevalence, age of initiation and association with alcohol-related harm among South African youth: Implications for policy. S Afr Med J 2012; 102: 609-12. doi:10.7196/SAMJ.5766.

[24] Wang G, Zhang Z, Ayala C, et. al. Costs of Hospitalisation for stroke patients aged 18-64 years in the United States. J Stroke Cerebrovasc Dis 2014;23: 861-8.

[25] Romero JR, Morris J, Pikula A. Review: Stroke prevention: modifying risk factors. Ther Adv Cardiovasc Dis 2008;2(4):287-303.
[26] 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):1-10.
[27] Matizirofa L, Kuhudzai AG. Predictors of high blood pressure in South African children: quantile regression approach. IJSMBR 2017;6:1-8.
[28] 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. J Hypertens 2016;34 (10):2059-65.
[29] van Alebeek ME, Arntz RM, Ekker MS, Synhaeve NE, Maaijwee NA, Schoonderwaldt H, et al. Risk factors and mechanisms of stroke in young adults: The FUTURE study. J Cereb Blood Flow Metab 2018;38(9): 1631-41.
[30] Hiraga A. Gender differences and stroke outcomes. Neuroepidemiology 2017;48(1-2):61-2.
[31] 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(4):331-7.
[32] Allen CL, Bayraktutan U. Risk Factors for ischaemic Stroke. Int J Stroke 2008;3(2):105-16.
[33] Goldstein LB, Adams R, Becker K, Furberg CD, Gorelick PB, Hademenos G, et al. Primary prevention of ischemic Stroke: a statement for healthcare professionals from the stroke Council of the American Heart Association. Stroke 2001;32(1):280-99.

Submitted: April 20, 2020. *Revised:* May 05, 2020. *Accepted:* May 10, 2020.