Journal of Multidisciplinary Healthcare

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Analysis of Health-Related Quality of Life in Elderly Patients with Stroke Complicated by Hypertension in China Using the EQ-5D-3L Scale

Shuai Wang[1], Caiyu Shen[2], Shu Yang[2]

[1]School of Public Health, Bengbu Medical University, Bengbu, Anhui, 233000, People's Republic of China; [2]School of Health Management, Bengbu Medical University, Bengbu, Anhui, 233000, People's Republic of China

Correspondence: Shu Yang, Bengbu Medical University, No. 2600, Donghai Avenue, Longzihu District, Bengbu, Anhui, People's Republic of China, Email sy.bbmu@foxmail.com

Purpose: To evaluate the health-related quality of life(HRQoL)status of elderly patients with hypertensive stroke, to understand the factors influencing it, and to provide a basis for the development of health intervention policies.

Patients and Methods: This study used the EQ-5D-3L scale to assess the HRQoL among elderly patients who experienced a stroke related to high blood pressure. Various analytical methods were employed to examine the factors that influenced the patient's quality of life. Univariate analysis, Tobit regression, random forest, and XGBoost models were applied to analyze the HRQoL of the patients. Furthermore, to interpret the machine learning results, the SHAP method was utilized. This method involved assessing the importance of each feature, examining the effect of each feature on the prediction result of a single sample, and determining the impact of individual features on the overall prediction.

Results: The study found that the median health utility value for elderly patients with hypertensive stroke was 0.427, with an interquartile range of 0.186 to 0.745. The results of the Tobit regression model, Random Forest, and XGBoost model were compared. The results of the model evaluation show that the performance of the machine learning model and the Tobit regression model are not very different. The XGBoost model performs slightly better relative to the random forest model. The factors that strongly influenced the health utility value of patients included BMI, social activities, smoking, education level, alcohol consumption, urban/rural residence, annual income, physical activity level, and hours of sleep at night.

Conclusion: Health-related quality of life in hypertensive stroke patients is influenced by a variety of factors. Health-related quality of life can be positively influenced by modifying these factors and making lifestyle adjustments. Maintaining a healthy weight, being socially active, quitting smoking, improving living conditions, increasing physical activity levels and getting enough sleep are recommended. Lifestyle changes need to be developed for each individual on a case-by-case basis and by medical advice.

Keywords: HRQoL, hypertensive stroke patients, machine learning, SHAP

Introduction

China is facing the world's largest stroke challenge. According to the results of the Global Disease Burden Study, in 2019, there were 3.94 million new stroke cases, 28.76 million epidemic cases, and 2.19 million deaths in China.[1] Stroke is the second leading cause of death and the third leading cause of disability globally, and the burden of stroke is increasing rapidly in low- and middle-income countries.[2] Although the standardized incidence and standardized mortality rates of stroke have declined since 1990, the disease burden of stroke in China remains very high.[3] This may be the result of China's aging population. In this context, aligning healthcare initiatives with sustainable development goals becomes imperative.

The 2022 Chinese Clinical Practice Guidelines for Hypertension recommend lowering the diagnostic criteria for hypertension to systolic blood pressure (SBP) ≥130 mmHg and/or diastolic blood pressure (DBP) ≥80 mmHg in Chinese adults, more people will now categorized as hypertensive and can be controlled through lifestyle and other measures.[4] It

Received: 21 February 2024
Accepted: 23 April 2024
Published: 30 April 2024

is estimated that nearly 50% of strokes can be attributed to hypertension.[5] Both diastolic and isolated systolic hypertension are important predictors of primary or recurrent stroke, and even a slight decrease in blood pressure can reduce the risk of stroke.[6] As hypertension persists, the probability of stroke risk increases.[7] The study of HRQoL in stroke patients has become a hot issue. There are studies focused on investigating the health-related quality of life of stroke patients, and there are studies dedicated to exploring the health-related quality of life profile of hypertensive patients.[8,9] National and international studies have shown that hypertension is a controllable and important risk factor for stroke, but through the literature, no specific papers have been found on health-related quality of life analysis in the population of elderly patients with stroke who suffer from hypertension. This suggests that there is a certain knowledge gap in the field of health-related quality of life research in this specific population. Although hypertension and stroke are common in older adults, further research is needed to gain a deeper understanding of the impact on quality of life in patients with both conditions coexisting.

The EQ-5D-3L is a commonly used HRQoL assessment tool to measure an individual's health status and their perception and assessment of health.[10] The higher the health utility value, the better their HRQoL. In this study, we propose to evaluate the health utility value of patients based on the EQ-5D-3L scale for a specific group of elderly people suffering from hypertensive stroke, to explore the influencing factors affecting the HRQoL of elderly patients with hypertensive stroke, and to discuss their influencing factors and importance. To provide a reference basis for improving the HRQoL of elderly patients with hypertensive stroke.

Methods

Data Sources

China Health and Retirement Longitudinal Study(CHARLS) is a longitudinal survey of the middle-aged and elderly population representative, targeting people over 45 years old and including data on basic personal information, health status, lifestyle, healthcare utilization, pension status, economic status, family situation, social support, etc.[11] CHARLS is a publicly available dataset, and following the regulations on dataset use to apply for data use has been granted a license to use the CHARLS 2015 dataset. The CHARLS program was reviewed and approved by the Ethics Review Board of Peking University (approval number:IRB00001052-11015). All participants or their surrogate respondents signed an informed consent form.

Data Extraction and Variable Assignment

Data cleaning was performed according to the purpose of the study. A total of 21,095 patients were surveyed during the CHARLS 2015 follow-up period, and after combining basic information with health status and functional data, patients with lack of health status were excluded, leaving 20,966 patients. Patients who did not have a stroke were excluded, leaving 564 patients. Finally, patients who did not suffer from hypertension were excluded, leaving 301 study participants, as shown in Figure 1.

Corresponding information on hypertensive stroke patients included in the study was extracted from the CHARLS dataset. The depression scores were calculated according to the CES-D 10 scale, and the 10-item streaming depression self-assessment scale had high reliability and validity, and was able to effectively measure the level of depression in the elderly population.[12] The regional variable is divided into seven geographical divisions: East China, South China, North China, Central China, Southwest China, and Northeast China. Age was calculated based on the true age provided by the patient. Marital status is categorized as married/unmarried. Educational attainment is the highest level of education attained by the patient at the time of the interview, categorized as no education, primary and below, middle school, college and above. Sleep duration was calculated based on the number of hours of sleep at night and was categorized as not enough (<7h/d), normal (≥7h/d and ≤8h/d), and excessive length (>8h/d).[13] Drinking behavior was differentiated by drinking status in the past year, with no drinking in the past year being classified as "non-drinking". Smoking behavior was categorized as still smoking, having quit smoking, or never having smoked. BMI <18.5 kg/m^2 was considered too low, 18.5 kg/m^2≤BMI<24kg/m^2 was considered normal, 24kg/m^2≤BMI<28kg/m^2 was considered overweight, and BMI≥28kg/m^2 was considered obese.[14] Waist circumference ≥90cm in men and ≥85 cm in women is considered to be

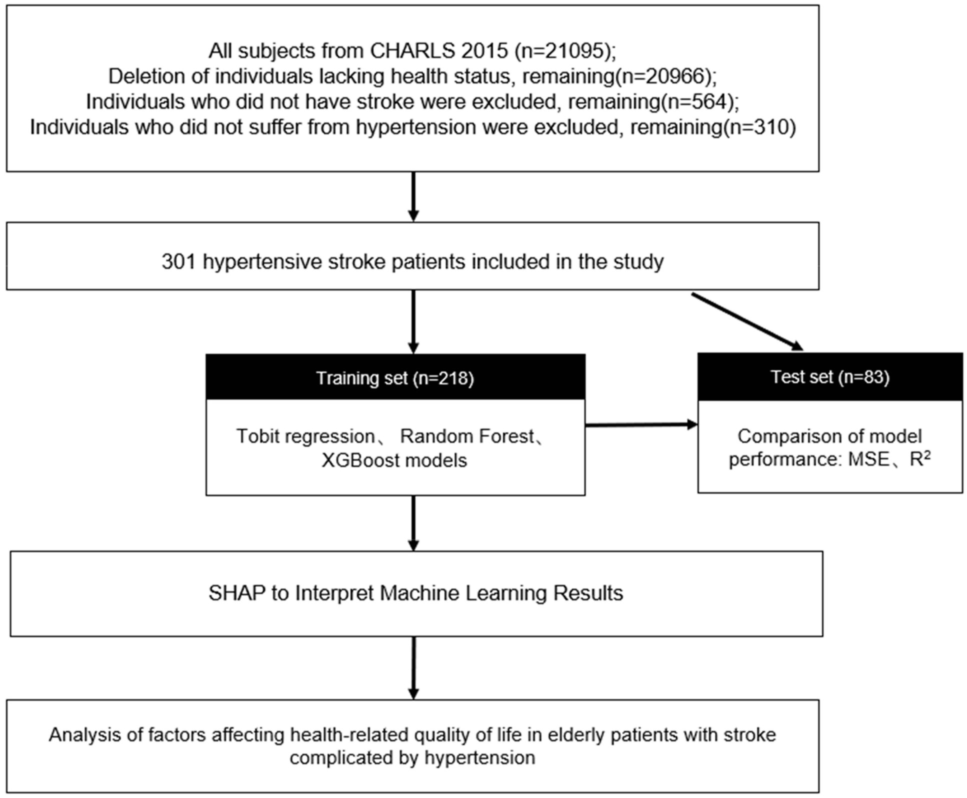

Figure 1 Flowchart for inclusion of research subjects and analytical methods. MSE, Mean Square Error; R^2, R-squared.

abdominal fat accumulation.[15,16] Physical activity levels are assessed according to the International Physical Activity Quotient (IPAQ), which allows for the categorization of a week's physical activity energy expenditure into three levels of intensity: lower, medium and highly.[17] Social activities were categorized as yes/no.

HRQoL Evaluation

The EQ-5D-3L scale was used to evaluate the HRQoL of the patients, as the most common international scale, its ease of use, the amount of data accumulated, and other advantages have made it a widely used assessment tool in the medical field, and the higher the value of health utility, the better their HRQoL.[18,19]

5D refers to the five dimensions of the EQ-5D-3L scale: mobility, self-care, daily activities, pain/discomfort, and anxiety/depression. This study selected the question "Do you have difficulty bending, bending knees, or squatting down? " From the CHARLS questionnaire as a measure of the respondents' mobility ability; "May I ask if you have difficulty cooking due to health and memory reasons? "As a measure of self-care for respondents; "May I ask if you have difficulty doing household chores due to health and memory reasons? "As a measure of the respondents' daily activities; "Do you often feel distressed by physical pain? Have you taken measures to alleviate the pain? "As a measure of pain/discomfort for respondents; According to the CES-D10 scale, depression scores were calculated as a measure of anxiety/depression in elderly respondents.[20]

3L refers to the three levels of each dimension of the EQ-5D-3L scale: Level 1, which indicates that the individual has no problems with the dimension; Level 2, which indicates that the individual has mild or moderate problems with the dimension; and Level 3, which indicates that the individual has severe problems or complete limitations with the dimension. The EQ-5D-3L scale is widely available and easy to construct, making cross-sectional comparisons between diseases and longitudinal comparisons of histories easier to make.[21] This also supports health economics and clinical decision-making.

Table 1 shows the EQ-5D health utility value calculated using a comprehensive utility integration system obtained by Liu et al, which combines rural and urban populations. It is currently a more suitable integration system for Chinese people.[22] Formula: U = 1 - C - N3 - (mobility + self-care + daily activities + pain/discomfort + anxiety/depression).

Table 1 The EQ-5D-3L Scale Based on the Health Utility Value Point System of Chinese Residents

Variable	Define	Coefficients
C	At least one dimension is at level 2, 3	0.067
N3	At least one dimension is at level 3	0.016
Mobility		
MO1	The dimension of action ability is at level 1	0.000
MO2	The dimension of action ability is at level 2	0.101
MO3	The dimension of action ability is at level 3	0.275
Self-Care		
SC1	Self-Care dimension at level 1	0.000
SC2	Self-Care dimension at level 2	0.103
SC3	Self-Care dimension at level 3	0.239
Usual Activities		
UA1	Daily activity dimension at level 1	0.000
UA2	Daily activity dimension at level 2	0.086
UA3	Daily activity dimension at level 3	0.217
Pain/Discomfort		
PD1	Pain/Discomfort dimension at level 1	0.000
PD2	Pain/Discomfort dimension at level 2	0.110
PD3	Pain/Discomfort dimension at level 3	0.232
Anxiety/Depression		
AD1	Anxiety/Depression dimension at level 1	0.000
AD2	Anxiety/Depression dimension at level 2	0.074
AD3	Anxiety/Depression dimension at level 3	0.172

Where C is a constant term and N3 indicates at least one dimension level of 3. MO1, SC1, UA1, PD1, AD1 indicates that the value 1 is assigned when level 1 is selected among the five dimensions, and 0 in other cases; MO2, SC2, UA2, PD2, AD2 indicates that the value 1 is assigned when level 2 is selected among the five dimensions, and 0 in other cases; MO3, SC3, UA3, PD3, AD3 indicates that the value 1 is assigned when level 3 is selected among the five dimensions, and 0 in other cases. Overall health state U is highest at 1 if all five dimensions have a level of 1; if all five dimensions have a level of 3, overall health state U is lowest at −0.218.

Statistical methods

Tobit Regression

Descriptive analyses of demographics, economic status, health behaviors, health problems, hypertension-related health problems, and quality of life of elderly hypertensive stroke patients were performed using frequency counts, constitutive ratios, medians, and interquartile spacing.

Statistical analyses were performed applying R4.2.2 with a test level of $\alpha = 0.05$. Health utility values were tested for normality and the results were skewed and statistically described using median and interquartile spacing [M (P_{25}, P_{75})]. One-way analyses were performed using the Wilcoxon rank sum test and the Kruskal–Wallis (K - W) test. Multifactor analysis was performed using Tobit regression model, which has been widely used by a large number of experts and scholars to study the evaluation of health scales, the analysis of factors affecting the quality of life, and the analysis of factors affecting health.[23] The health utility value as the dependent variable has a certain limitation in the range of [−0. 218~1], which has the characteristic of being intercepted and meets the conditions for the use of Tobit regression model.

Machine Learning Methods

Random Forest is an integrated learning algorithm proposed by Breiman that combines Bagging and decision tree modeling.[24,25] It combines the simplicity, efficiency, and easy interpretability of decision trees with the accuracy and stability of integrated learning. XGBoost is an integrated learning algorithm based on Boosting.[26] In the process of building a decision tree, XGBoost calculates the splitting gain based on the gradient and second-order derivative of each sample, which is used to determine the optimal features and thresholds for splitting. The optimal split is selected by a greedy algorithm to generate multiple decision trees and combine their predictions by weighting them to build a stronger model.[27,28]

In this study, machine learning models (including Random Forest and XGBoost) were used to predict the health utility values of elderly hypertensive stroke patients, and the importance of each feature in the prediction was analyzed. In view of the small sample size of the study, deep learning models may be overfitted with small samples, so machine learning models are chosen instead of deep learning models in this study. In this case, XGBoost is used as an example, and the model is optimized through techniques such as parallelization, pruning, regularization, and weighted quantum algorithm to improve training speed and prediction performance. These optimization methods can help improve the performance and generalization of machine learning models. In this paper, XGBoost is manually tuned to avoid the risk of overfitting.

Explanatory Models

Predictive analytics and explanatory analytics are one of the important tasks of machine learning. Predictive analytics helps us to make predictions about unknown data, while explanatory analytics helps us to understand the decision-making process of the model and the reasons behind the results.[29] The interpretability methods for machine learning models are Partial Dependence Plot (PDP), Individual Conditional Expectation (ICE), Permuted Feature Importance, Global Surrogate, Local Surrogate (LIME), Shapley Value (SHAP), and so on.[30,31] In order to better evaluate the role of influencing factors, the SHAP interpretation model is introduced to interpret the XGBoost results. In the field of medicine, researchers have successfully used SHAP values for machine learning model interpretation.

SHAP is a method for improving the interpretability of "black-box" machine learning models. SHAP can be used to calculate the contribution of each observation to the model predictions by combining different values of the features into different subsets of features and calculating the contribution of each feature, and then calculating the contribution of each observation to the model predictions by assigning these contributions. This gives the importance of each feature to the overall model prediction, and SHAP provides a comprehensive review of how each feature affects the model prediction. One of the major advantages of SHAP is that it can reflect the influence of the feature in each sample and the positive or negative impact of that influence on the final prediction.[32,33]

Results

Descriptive Statistics

The present study had a valid sample of 301 cases, of which the highest number of males was 169 (56.15%). The largest number of people living with their families was 266 (88.37%), and the largest number of people earning between ¥50,000 and ¥100,000 a year was 218 (72.43%). The highest number of individuals with health insurance was 273 (90.7%). The highest number of individuals with blood pressure controlled to normal was 235 (78.07%). The highest number of individuals with a low level of weekly physical activity was 216 (71.76%). The highest number of individuals with no monthly social activities was 152 (50.50%). The highest number of individuals with insufficient night sleep was 131 (43.52%). The highest number of people were still smoking was 147 (48.84%). The highest number of people not drinking was 227 (75.42%). The highest number of people had a nap during the day was 188 (62.46%). The highest number of people aged 64–74 years was 125 (41.53%). The highest number of married people was 237 (78.74%). The highest number of people with education level of primary school and below was 137 (45.51%). The highest number of people living in rural areas was 156 (51.83%). The highest number of people living in East China was 81 (26.91%). The highest number of people with overweight body mass index (BMI) was 109 (36.21%). The highest number of people with abdominal fat accumulation was 222 (73.75%).

The results of the one-way analysis of variance showed that annual income, weekly physical activity level, monthly social activities, nightly sleep duration, smoking status, age, education level, urban/rural residence, alcohol consumption, BMI, and abdominal fat accumulation had a statistically significant effect on the HRQoL of elderly patients with hypertensive stroke. ($P < 0.05$). (Table 2)

Table 2 Basic Information, Health Utility Values and Univariate Analysis in Hypertensive Stroke Patients(n=301)

Features	N (%)	Health Utility value [M (P_{25}, P_{75})]	*P* value
Gender			0.909
Male Female	169(56.15) 132(43.85)	0.458(0.186, 0.745) 0.419(0.186, 0.745)	
State of residence			0.650
Live alone Living with family	35(11.63) 266(88.37)	0.360(0.186, 0.646) 0.451(0.186, 0.745)	
Annual income			<0.001
Less than 50, 000 50, 000–100, 000 Larger than 100, 000	22(7.31) 218(72.43) 61(20.27)	0.747(0.513, 0.859) 0.463(0.220, 0.745) 0.186(0.186, 0.360)	
Medical insurance			0.219
Yes None	273(90.70) 28(9.30)	0.427(0.186, 0.745) 0.480(0.186, 0.859)	
Blood pressure control			0.836
Control to normal Uncontrolled	235(78.07) 66(21.93)	0.455(0.186, 0.748) 0.396(0.258, 0.745)	
Physical activity level			<0.001
Lower Medium Highly	216(71.76) 52(17.28) 33(10.96)	0.360(0.186, 0.651) 0.641(0.423, 0.859) 0.611(0.412, 0.745)	
Social events			<0.001
Yes None	149(49.50) 152(50.50)	0.510(0.238, 0.758) 0.360(0.112, 0.662)	
Length of sleep at night			<0.001
Not enough Normal Excessive length	131(43.52) 78(25.91) 92(30.56)	0.455(0.203, 0.745) 0.643(0.412, 0.859) 0.233(0.186, 0.462)	
Smoking			0.003
Never smoked Still smoking Have quit smoking	38(12.62) 147(48.84) 116(38.54)	0.326(0.186, 0.482) 0.412(0.186, 0.745) 0.517(0.186, 0.839)	
Naps			0.824
Yes None	188(62.46) 113(37.54)	0.458(0.186, 0.745) 0.360(0.186, 0.749)	
Age			<0.001
<55 55~64	39(12.96) 79(26.25)	0.642(0.386, 0.754) 0.604(0.281, 0.766)	

(*Continued*)

Table 2 (Continued).

Features	N (%)	Health Utility value [M (P_{25}, P_{75})]	*P* value
64~74	125(41.53)	0.422(0.186, 0.648)	
≥74	58(19.27)	0.205(0.129, 0.461)	
Marital status			0.156
Unmarried	64(21.26)	0.360(0.186, 0.617)	
Married	237(78.74)	0.455(0.186, 0.746)	
Level of education			0.004
No education	70(23.26)	0.360(0.175, 0.548)	
Primary and below	137(45.51)	0.422(0.186, 0.672)	
Middle school	86(28.57)	0.615(0.194, 0.758)	
College and above	8(2.66)	0.690(0.502, 0.859)	
Urban and Rural residence			0.011
Rural	156(51.83)	0.360(0.186, 0.637)	
Urban	145(48.17)	0.510(0.186, 0.758)	
Area of residence			0.449
Northern China	48(15.95)	0.396(0.186, 0.619)	
North-eastern	27(8.97)	0.427(0.156, 0.745)	
Eastern China	81(26.91)	0.458(0.186, 0.758)	
Central China	57(18.94)	0.510(0.223, 0.745)	
South China	26(8.64)	0.438(0.186, 0.834)	
Southwestern	43(14.29)	0.424(0.203, 0.745)	
Northwestern	19(6.31)	0.326(0.083, 0.535)	
Drinking			0.001
Non-alcoholic	227(75.42)	0.360(0.186, 0.671)	
Drinking	74(24.58)	0.626(0.332, 0.758)	
BMI			<0.001
Normal	81(26.91)	0.604(0.326, 0.773)	
Overweight	109(36.21)	0.513(0.274, 0.758)	
Obese	89(29.57)	0.223(0.186, 0.465)	
Underweight	22(7.31)	0.186(0.179, 0.447)	
Fat accumulation in the abdomen			0.040
Yes	222(73.75)	0.412(0.186, 0.745)	
None	79(26.25)	0.482(0.273, 0.758)	

Abbreviations: [M (P_{25}, P_{75})], median and interquartile spacing; BMI, Body Mass index.

Multifactorial Analysis of Factors Influencing the Health Utility Value of Stroke Patients

Variables that were statistically significant in the univariate analysis were included in the Tobit intercept regression model. In particular, patients with an annual income of less than 50, 000 were used as a reference, and patients with an annual income of >50,000 had lower health utility values ($p<0.05$). Patients with a low level of weekly physical activity were used as a reference, and those with a medium or high level of physical activity had higher health utility values ($p<0.05$). Patients with a normal number of hours of sleep at night were used as a reference, and those with insufficient nightly sleep hours had lower health utility values ($p< 0.05$). Using patients with monthly social activities as a reference, patients without monthly social activities had lower health utility values ($p<0.05$). Using patients who were still smoking as a reference, patients who had quit smoking had higher health utility values ($p<0.05$). Using patients younger than 55 years as a reference, patients older than 65 years and younger than 74 years had lower health utility values ($p<0.05$). Using patients with a normal body mass index as a reference, patients who were obese and underweight had lower health utility values ($p<0.05$). (Table 3)

Table 3 Tobit Regression Modelling of Factors Influencing HRQoL in Hypertensive Stroke Patients(n=301)

Independent Variable	**β**	**SE**	**Z value**	***P* value**	**95% CI**
Annual income (Reference value = less than 50,000)					
50,000–100,000	−0.162	0.065	−2.476	0.013	(−0.289, −0.034)
Larger than 100,000	−0.179	0.083	−2.163	0.031	(−0.342, −0.017)
Physical activity level (reference value = low)					
Medium	0.118	0.045	2.589	0.010	(0.029, 0.207)
Highly	0.161	0.056	2.865	0.004	(0.051, 0.272)
Fat accumulation in the abdomen (reference value = yes)					
None	0.036	0.048	0.756	0.450	(−0.058, 0.130)
Social events (reference value = yes)					
None	−0.141	0.035	−4.078	<0.001	(−0.209, −0.073)
Length of sleep at night (reference value = normal)					
Not enough	−0.114	0.041	−2.781	0.005	(−0.194, - 0.034)
Excessive length	−0.046	0.058	−0.797	0.426	(−0.159, 0.067)
Smoking (reference value = still smoking)					
Never smoked	−0.058	0.052	−1.105	0.269	(−0.161, 0.045)
Have quit smoking	0.140	0.036	3.924	<0.001	(0.070, 0.210)
Drinking (reference value = no alcohol)					
Drinking	0.057	0.039	1.441	0.150	(−0.020, 0.134)
Age (reference <55 years)					
55~64	−0.054	0.056	−0.960	0.337	(−0.165, 0.056)
65~74	−0.104	0.052	−1.977	0.048	(−0.206, −0.001)
>74	−0.093	0.063	−1.479	0.139	(−0.216, −0.030)
Level of education (reference value = not educated)					
Primary school and below	−0.053	0.044	−1.205	0.228	(−0.140, 0.033)
Secondary Schools	−0.021	0.051	−0.406	0.685	(−0.121, 0.079)
Tertiary and above	0.092	0.111	0.832	0.406	(−0.125, 0.309)
Urban and Rural residence (reference value = Rural)					
Urban	0.041	0.034	1.187	0.235	(−0.027, 0.108)
BMI (reference value = Normal)					
Overweight	−0.025	0.050	−0.495	0.621	(−0.122, 0.073)
Obesity	−0.209	0.054	−3.892	<0.001	(−0.314, −0.104)
Underweight	−0.189	0.074	−2.572	0.010	(−0.334, −0.045)

Model Performance

The dataset is divided into training and test sets in the ratio of 7:3 for model training and performance evaluation. In the R language xgboost package, the core hyperparameters are set as follows: nrounds = 50, eta = 0.1, gamma = 0, subsample = 1, colsample_bytree = 1, max_depth = 3, min_child_weight = 3. At the same time, the number of random number of seeds to 1234 to facilitate reproduction of the results.

The Tobit regression model, random forest, and XGBoost were used for model training on the training set, and the trained models were tested on the test set based on the two metrics of mean square error (MSE) and coefficient of determination (R^2). The model performance of the three single models is judged. (Table 4)

The results of analyzing the metrics show little difference in performance between the machine learning model and the Tobit regression model. The XGBoost model performs slightly better relative to the random forest model. There is no significant difference in the performance of machine learning models constructed with or without all features.

Machine Learning Based Feature Importance Analysis

The coefficients of the Tobit regression model, indicate the direction and degree of influence of a particular characteristic on the results, but they cannot be used to directly compare the degree of influence between individual characteristics. This is because the Tobit regression model estimates the parameters by maximum likelihood estimation, which takes into account the effect of truncated data. The loss of information due to truncated data may result in biased estimates. As a result, the coefficients of different features may be affected by data truncation to varying degrees, making it difficult to compare the degree of influence between individual features. This study aims to compare the degree of influence of individual features to provide a more comprehensive and intuitive understanding and to help assess the relative importance of different features on the results. This study analyses two machine learning results, Random Forest and XGBoost.

Random Forest Feature Significance Analysis

In this study, we use the Random Forest function of the Random Forest package Random Forest in the R language to construct random forests and the caret package for hyperparameter tuning. In the establishment of random forests, which do not have access to the regression coefficients of the independent variables, the effect of the independent variables on the dependent variable is assessed by the mean reduction in mean squared error (%IncMSE), with a larger value indicating a greater importance of the variable. Plotting the importance of random forest features using the varImpPlot function.[34]

As shown in Figure 2, the top ten in terms of the importance of Random Forest characteristics are length of sleep at night, social activity, level of education, BMI, annual income, smoking, age, drinking, physical activity level, and urban/rural residence.

Table 4 Comparison of Model Performance

	Training set		Test Set	
	MSE	**R2**	**MSE**	**R2**
Tobit regression	0.0753	0.3004	0.0691	0.2926
Random Forest1	0.0402	0.6262	0.0710	0.2726
Random Forest2	0.0376	0.6502	0.0759	0.2224
XGBoost1	0.0519	0.5182	0.0653	0.3313
XGBoost2	0.0459	0.5731	0.0663	0.3210

Notes: The number of people in the training set is 218, The number of people in the test set is 83. Random Forest1 and XGBoost1: Include in the model those characteristics that were statistically significant in the one-way analysis of variance. Random Forest2and XGBoost2: Include all characteristics in the model.

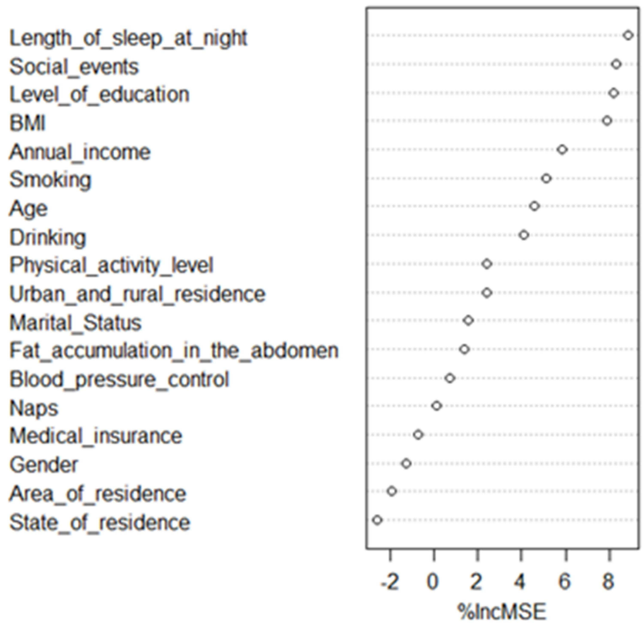

Figure 2 Importance ranking of Random Forest outcome features(n=218).

Importance Analysis of XGBoost Features

In this study, XGBoost is constructed using the xgboost package xgb.train function in R. Hyperparameter tuning is performed using the caret package. In XGBoost, the importance of features can be ranked by calculating the Gain score for each feature. The Gain score indicates how much each feature contributes to the model performance when split in the tree. A higher Gain score means that the feature contributes more to the performance of the model.[35] Plotting XGBoost feature importance using the xgb. importance function.

In terms of the importance of XGBoost characteristics, the top ten were BMI, social activities, age, annual income, drinking, level of education, area of residence, smoking, medical insurance, urban/rural residence, see Figure 3.

SHAP-Based Model Interpretation

This study uses the shapviz package in RStudio to visualize XGBoost results.

Contribution of Single-Sample Characteristics to Predicted Values

The graphs were drawn using the sv_force function, see Figure 4. Select a training sample. Demonstrate the SHAP model for a single sample, with the color and size of the nodes indicating the SHAP value of the feature. Larger nodes indicate higher SHAP and smaller nodes indicate lower SHAP values. The red color represents a positive effect, when the patient has a normal BMI, there is a positive effect on the health utility value, with a SHAP value of 0.0507; The blue color represents the negative impact, when the patient has no monthly social activities, it is a negative impact on the health utility value, with a SHAP value of −0.0588, which is the highest impact. The most important variables in this sample are social activity, age, and BMI. Other variables are less important. In this sample, the variables of social activity, age, smoking, urban/rural residence, education level, drinking, and length of sleep at night have a negative effect on the value of health utility, ie, on their HRQoL, while the rest of the variables have a positive effect on the value of health utility, ie, on their HRQoL.

Importance Analysis of SHAP Explanatory Model Features

A bar chart of feature importance based on SHAP values is made using the sv_importance function. The vertical coordinate corresponds to the feature items and the horizontal axis is the mean absolute value of the SHAP values, which reflects the importance of each feature in the prediction. The top 10 important of features are BMI, social activities,

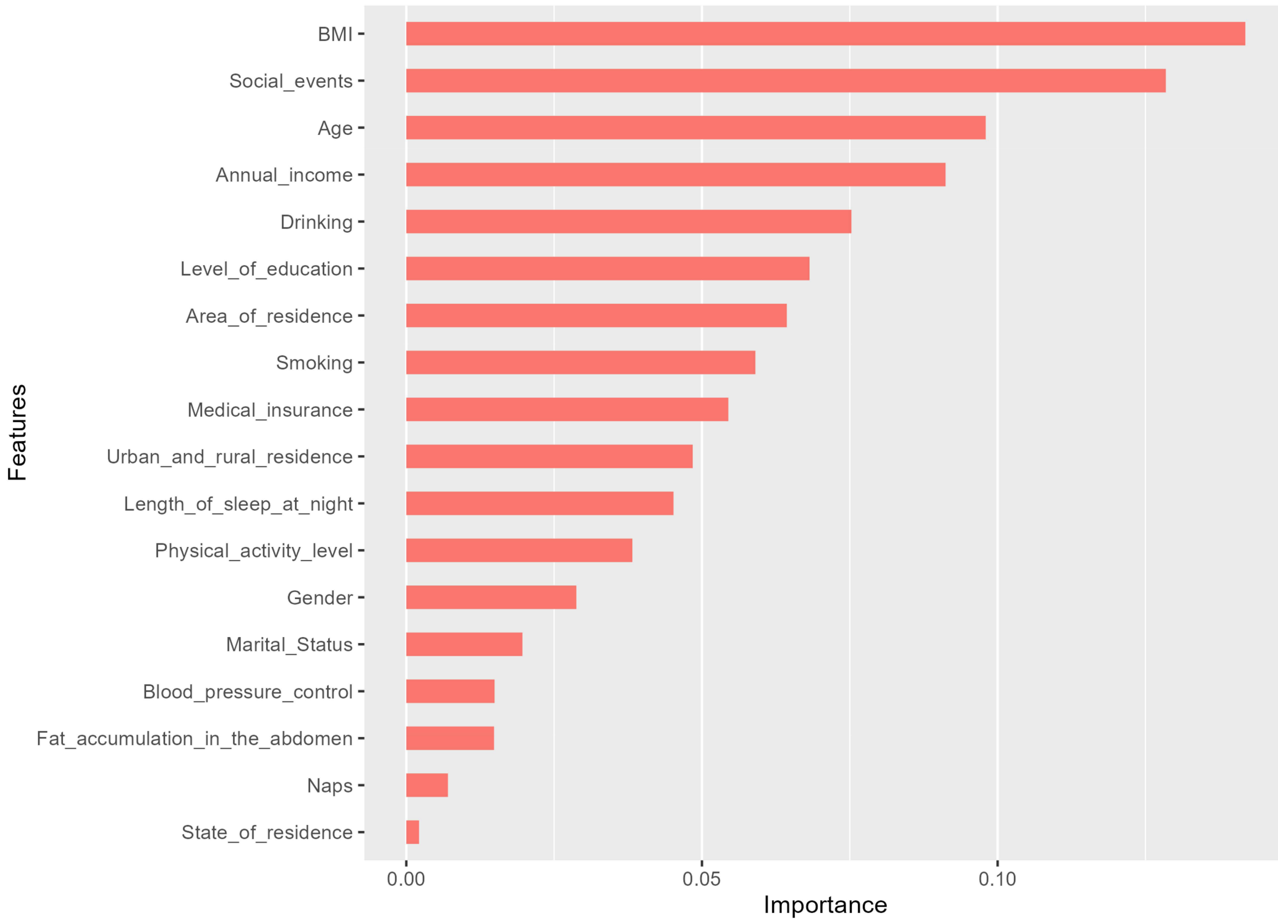

Figure 3 Chart representing importance ranking of features obtained from result of XGBoost classifier(n=218).

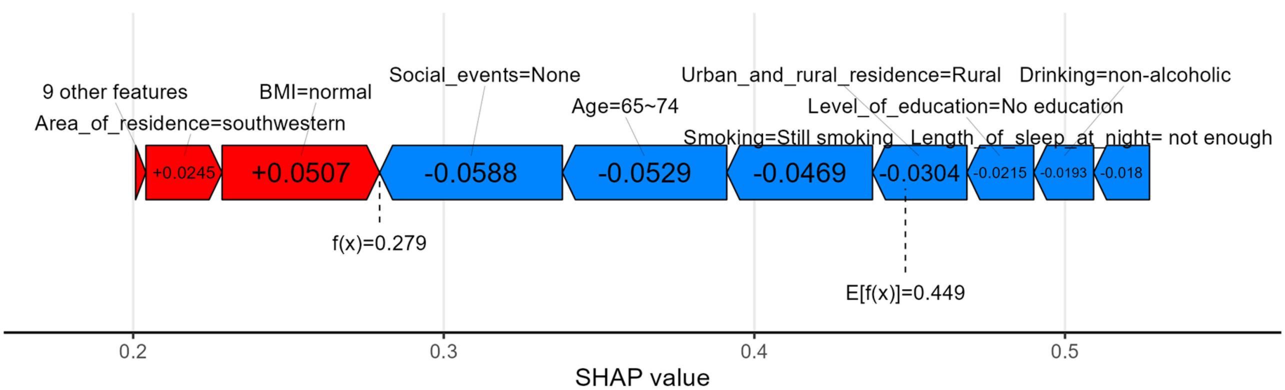

Figure 4 Contribution of each characteristic to the predicted value for a single sample.

smoking, age, level of education, urban/rural residence, length of night sleep, region of residence, drinking and gender, as shown in Figure 5.

Assumptions about the importance of features based on Random Forest, XGBoost and SHAP interpretation models. Different models and methods may yield slightly different results. The Random Forest Feature Importance Plot measures the impact of features on the model by calculating the percentage increase in Mean Square Error (MSE) after a random permutation of each feature relative to the original case. The relative importance ranking of features is obtained by calculating the %IncMSE value for each feature.[36] The XGBoost Feature Importance Map calculates feature importance

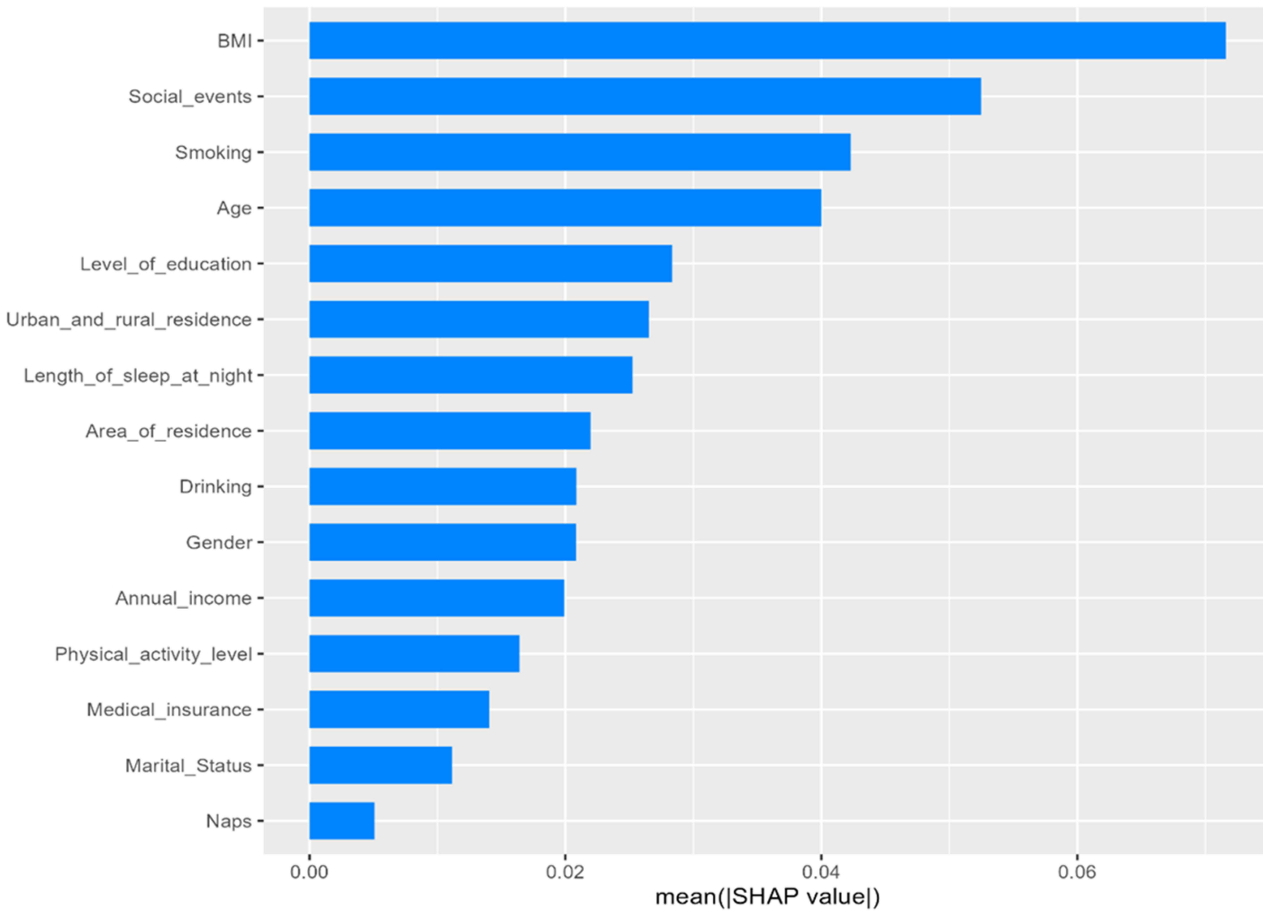

Figure 5 Ranking of feature importance(n=218).

scores based on the information gain (Gain) of the features as they split at the tree nodes. The cumulative Gain value of each feature will be normalized and ranked to measure how much it contributes to the model prediction. By analyzing the Gain values of the features, we can determine which features are most important for prediction in the XGBoost model.[37] The SHAP explanatory model utilizes the concept of game theory by removing one feature in turn and then calculating the contribution of the remaining features. The contribution of each feature to the outcome in different combinations is considered and the marginal contribution of each feature is derived accordingly.[38] Combining the results of the three methods, the characteristics of BMI, social activity, level of education, age, smoking, drinking, and urban/rural residence had a greater impact on the predicted results.

Contribution of Single Features to Predicted Values

Taking BMI as an example, the relationship between the four classifications of BMI and the corresponding SHAP values is visualized, and a SHAP dependence plot is drawn using the sv_dependence function. The relative influence of BMI in different classifications on the prediction results is presented. SHAP values are positive when the patient's BMI is normal and overweight. This means that in both cases, BMI has a positive effect on the predicted outcome. On the contrary, when the patient's BMI was obese and underweight, the SHAP value was negative. This means that in both cases, BMI has a negative impact on predicting outcomes. See Figure 6.

Discussion

In this study, the health utility value reflects the HRQoL of elderly patients with hypertensive stroke through the health utility value, and the median health utility value of stroke patients was measured to be 0.427, which is lower than that of

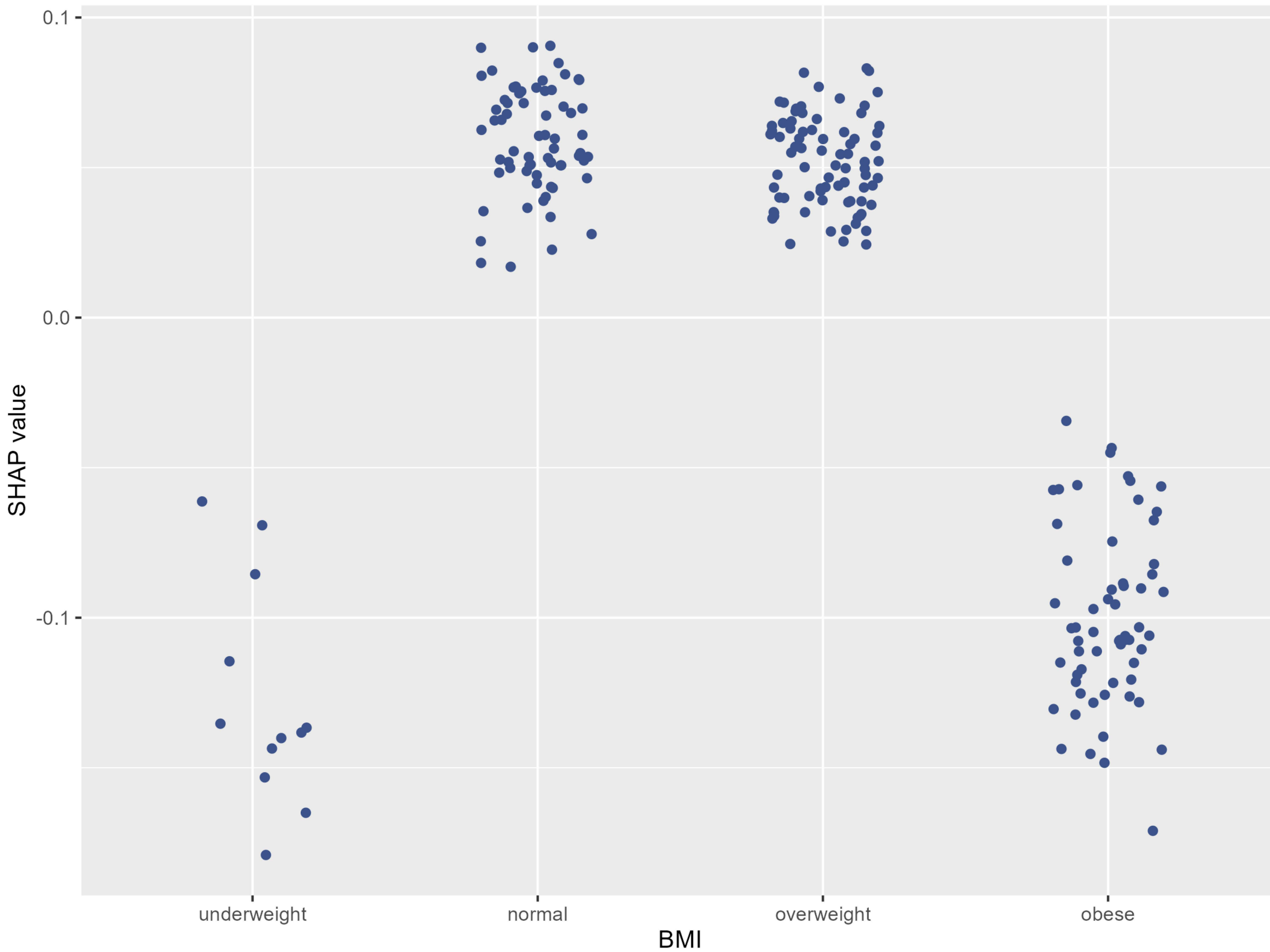

Figure 6 BMI Contribution Map(n=218).

the study of Deng et al in this study, probably because of the difference in the health utility system used, and the population of the study.[39] Since the research object of this study is mostly rural population, the health utility value system used is Liu et al added rural comparison on the basis of urban, which is more in line with China's national conditions. Moreover, the higher proportion of participants with controlled blood pressure can be attributed to their educational background. About 80% of these individuals had received education, enabling them to comprehend and adhere to their doctor's instructions, such as timely medication intake and regular blood pressure monitoring. This educational advantage likely facilitates better hypertension management, thereby enhancing overall blood pressure control success. Some studies have shown that patients with higher levels of education are better able to control their hypertension.[40,41] Additionally, it's noteworthy that a significant portion of patients in our study reported low levels of weekly physical activity. This can be partly attributed to the fact that our study cohort comprises elderly patients with stroke concomitant with hypertension. Older adults may encounter certain constraints in physical activity, compounded by the impact of conditions such as stroke and hypertension, thereby leading to diminished overall activity levels. After Tobit regression and machine learning analysis, it was finally determined that the characteristics that have a greater impact on the health utility value include BMI, social activities, age, smoking, education level, alcohol consumption, urban and rural residence, annual income, physical activity level, and nighttime sleep duration.

Obese and underweight patients have worse HRQoL, consistent with the findings of Craig et al, informing patients that they should eat a healthy diet and maintain a healthy weight.[42] HRQoL was better in patients with monthly social activities, in line with the findings of Gudina et al, family and social support is crucial for patients' recovery and mental health.[43] Patients are encouraged to stay in touch with family, friends and community and participate in social activities to reduce feelings of anxiety/depression and improve emotional state. As the body's bodily functions deteriorate with age, the HRQoL of hypertensive stroke patients deteriorates, in line with the findings of Sadlonova M.[44] HRQoL is higher in patients who have undergone smoking cessation, consistent with the findings of McClave et al.[45] HRQoL was higher in patients with higher levels of education, consistent with the findings of Liao et al. HRQoL is higher in patients with high levels of education, consistent with the findings of Liao et al.[46] Providing relevant health education and information helps patients to understand their diseases, treatments, and rehabilitation measures. Increasing patients' knowledge of their condition helps them to better manage their health. HRQoL was better in patients who drank alcohol. The GBD published in The Lancet suggests that for people aged 40 years and over, there is instead a slight benefit from small daily alcohol consumption, consistent with the results of this study.[47] HRQoL was better in patients residing in urban areas, which may be due to the fact that urban areas provide more healthcare resources, better education, in line with the findings of Liu et al.[48] Patients with annual incomes of ¥50,000 or less have better HRQoL. Most studies suggest that people with lower incomes may be more likely to be at risk for cardiovascular health.[49,50] In our study, we observed higher weekly activity ratings among patients with lower incomes, possibly because they held jobs with more physical labor or were more focused on maintaining health through exercise. This association reflects a complex relationship between income and health behaviors that needs to be explored in depth by further research. Patients with moderate or heavy weekly physical activity levels had better HRQoL, consistent with the findings of Chen et al.[51] Some studies have shown that people who walk at least 5 times a week for 30 min are of higher health-related quality than those who do not exercise, and patients are advised to get more moderate or heavy aerobic exercise each week.[52] HRQoL was better in patients with normal hours of sleep at night, consistent with the findings of Levine et al.[53] Adequate sleep maintains normal metabolic function, balances hormone levels, and helps lower blood pressure.

There are some limitations of this study. First, the selection of characteristics did not cover all factors affecting the HRQoL of elderly hypertensive stroke patients, and there may be omissions. Second, the statistical significance of the findings is questionable due to the small sample size, while bias in sample selection may have limited overall representativeness. Furthermore, our study lacks a control group for comparative analysis, which limits the ability to draw direct causal inferences. Therefore, we need to be cautious in interpreting the study findings and realize that the results are only applicable to the current sample. To ensure the reliability and external applicability of the results, future studies should consider increasing the sample size and incorporating control groups for comparative analysis. These steps will help strengthen the robustness of our findings and contribute to a more comprehensive understanding of HRQoL in elderly hypertensive stroke patients.

Conclusion

The above results suggest that the HRQoL of hypertensive stroke patients is influenced by many factors, including many acquired ameliorative and controllable lifestyle factors, in addition to uncontrollable or unregulated factors such as age. These factors include BMI, social activities, smoking, level of education, alcohol consumption, urban/rural residence, annual income, level of physical activity, and number of hours of sleep at night. By modifying these factors and making lifestyle adjustments, it is possible to have a positive impact on people's HRQoL. For example, maintaining a healthy weight (controlled by BMI), being socially active, quitting smoking, improving living conditions, increasing physical activity levels, and getting enough sleep are all examples of healthy lifestyles that may be beneficial. However, it is important to note that each individual's specific circumstances and medical recommendations may vary slightly. Therefore, it is recommended to seek guidance and advice from medical professionals in developing and implementing lifestyle changes.

Acknowledgments

We are grateful to the China Center for Economic Research at Beijing University for providing us with the data, and we thank the CHARLS research and field team for collecting the data.

Author Contributions

All authors made significant contributions to the reported research, from initial concept and design to execution, data collection, analysis and interpretation. They were involved in writing, reviewing, and editing the manuscript, approved the final manuscript for publication, agreed on the journal selected for submission, and agreed to take responsibility for all aspects of the work. In addition, all authors have reviewed and agreed to the published version of the manuscript.

Funding

This work was supported by the Natural Science Key Project of Anhui Province Education Fund (Grant No. KJ2017A223), the Natural Science Key Project of Bengbu Medical University (Grant No.2020byzd018) and Postgraduate Research and Innovation Program Projects of Bengbu Medical University (Grant No.Byycx23044, Byycx23038).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Wang YJ, Li ZX, Gu HQ, et al. China Stroke Statistics: an update on the 2019 report from the national center for healthcare quality management in neurological diseases. China national clinical research center for neurological diseases, the Chinese stroke association, national center for chronic and non-communicable disease control and Prevention, Chinese center for disease control and prevention and institute for global neuroscience and stroke collaborations. *Stroke Vasc Neurol*. 2022;7(5):415–450. doi:10.1136/svn-2021-001374
2. Owolabi MO, Thrift AG, Mahal A, et al. Primary stroke prevention worldwide: translating evidence into action. *Lancet Public Health*. 2022;7(1): e74–e85. doi:10.1016/S2468-2667(21)00230-9
3. Ma Q, Li R, Wang L, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990–2019: an analysis for the global burden of disease study 2019. *Lancet Public Health*. 2021;6(12):e897–e906. doi:10.1016/S2468-2667(21)00228-0
4. Lou Y, Ma WJ, Wang ZJ, et al. Writing protocols for the Chinese clinical practice guidelines of hypertension. *Chinese J Cardiol*. 2022;50 (07):671–675. doi:10.3760/cma.j.cn112148-20211126-01021
5. Furie K. Epidemiology and Primary Prevention of Stroke. *Continuum (Minneap Minn)*. 2020;26(2):260–267. doi:10.1212/ CON.0000000000000831
6. Castilla-Guerra L, Fernández-Moreno MDC. Update on the management of hypertension for secondary stroke prevention. *Eur Neurol*. 2012;68 (1):1–7. doi:10.1159/000336836
7. le LA, Ji Y, Zhu S, et al. Risk probability and influencing factors of stroke in followed-up hypertension patients. *BMC Cardiovasc Disord*. 2022;22 (1):328. doi:10.1186/s12872-022-02780-w
8. Wang C, Lang J, Xuan L, Li X, Zhang L. The effect of health literacy and self-management efficacy on the health-related quality of life of hypertensive patients in a western rural area of China: a cross-sectional study. *Int J Equity Health*. 2017;16(1):58. doi:10.1186/s12939-017-0551-9
9. Joundi RA, Patten SB, Lukmanji A, Williams JVA, Smith EE. Impairment in health-related quality of life among community-dwelling stroke survivors. *Can J Neurol Sci*. 2022;49(3):373–380. doi:10.1017/cjn.2021.119
10. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health*. 2009;12(8):1194–1200. doi:10.1111/j.1524-4733.2009.00581.x
11. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the china health and retirement longitudinal study (CHARLS). *Int J Epidemiol*. 2014;43 (1):61–68. doi:10.1093/ije/dys203
12. Huang QB, Wang XH, Chen G. Reliability and Validity of 10-item CES-D among middle aged and older adults in China. *China J Health Psychol*. 2015;23(07):1036–1041. doi:10.13342/j.cnki.cjhp.2015.07.023
13. Domínguez F, Fuster V, Fernández-Alvira JM, et al. Association of sleep duration and quality with subclinical atherosclerosis. *J Am Coll Cardiol*. 2019;73(2):134–144. doi:10.1016/j.jacc.2018.10.060
14. Huang L, Gao X. The interaction of obesity and craniofacial deformity in obstructive sleep apnea. *Dentomaxillofac Radiol*. 2021;50(4):20200425. doi:10.1259/dmfr.20200425
15. Ahn S, Han K, Lee JE, Jeun SS, Park YM, Yang SH. Associations of general and abdominal obesity with the risk of glioma development. *Cancers*. 2021;13(12):2859. doi:10.3390/cancers13122859
16. National Health and Family Planning Commission of the People's Republic of China. *WS/T 428-2013 Criteria of Weight for Adults[S]*. Beijing: China Quality Inspection Press, Chinese Standards Press; 2013.
17. Qu NN, Li KJ. Study on the reliability and validity of international physical activity questionnaire (Chinese Vision, IPAQ). *Zhonghua Liu Xing Bing Xue Za Zhi*. 2004;25(3):265–268.
18. Thayabaranathan T, Andrew NE, Stolwyk R, Lannin NA, Cadilhac DA. Comparing the EQ-5D-3L anxiety or depression domain to the Hospital Anxiety and Depression Scale to identify anxiety or depression after stroke. *Top Stroke Rehabil*. 2022;29(2):146–155. doi:10.1080/ 10749357.2021.1895494
19. Almeida Lins WM, Tura BR, Kasal DA. The association between physical performance and health-related quality of life based on the EQ-5D-3L questionnaire in patients with Chagas disease. *Value in Health Regional Issues*. 2021;26:191–196. doi:10.1016/j.vhri.2021.09.002
20. Wu TT, Lu Y, AI DD, Chen C, Chang F. Analyzing the health-related quality among the elderly with multiple chronic conditions in Jiangsu: estimation based on the EQ-5D Value Sets. *Chin Gen Pract*. 2020;23(S1):47–51.

21. Garratt AM, Furunes H, Hellum C, et al. Evaluation of the EQ-5D-3L and 5L versions in low back pain patients. *Health Qual Life Outcomes*. 2021;19(1):155. doi:10.1186/s12955-021-01792-y
22. Liu G, Guan H, Jin X, Zhang H, Vortherms S, Wu H. Rural population's preferences matter: a value set for The EQ-5D-3L health states for china's rural population. *Review*. 2021;2021:1. doi:10.21203/rs.3.rs-746682/v1
23. Vainiola T, Roine RP, Suojaranta-Ylinen R, Vento A, Sintonen H. Can factors related to mortality be used to predict the follow-up health-related quality of life (HRQoL) in cardiac surgery patients? *Intens Crit Care Nurs*. 2013;29(6):337–343. doi:10.1016/j.iccn.2013.04.003
24. BreimanLeo. Bagging predictors. Machine Learning; 1996. Available from: https://xueshu.baidu.com/usercenter/paper/show?paperid=1w650mn0bc760p70yf0x0gs0hm116084&site=xueshu_se. Accessed May 27, 2023.
25. Fang KN, Wu JB, Zhu JP, Xie BC. A Review of Random Forest Methods. *Statist Inform Forum*. 2011;26(3):32–38.
26. Chen T, Guestrin C. XGBoost: a Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016:785–794. doi:10.1145/2939672.2939785.
27. Bjerregaard SS. Exploring predictors of welfare dependency 1, 3, and 5 years after mental health-related absence in Danish municipalities between 2010 and 2012 using flexible machine learning modelling. *BMC Public Health*. 2023;23(1):224. doi:10.1186/s12889-023-15106-y
28. Boo Y, Choi Y. Comparison of mortality prediction models for road traffic accidents: an ensemble technique for imbalanced data. *BMC Public Health*. 2022;22(1):1476. doi:10.1186/s12889-022-13719-3
29. Patel SS. Explainable machine learning models to analyse maternal health. *Data Knowl. Eng*. 2023;146:102198. doi:10.1016/j.datak.2023.102198
30. Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med*. 2018;6(11):216. doi:10.21037/atm.2018.05.32
31. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016:1135–1144. doi:10.1145/2939672.2939778.
32. Sundararajan M, Najmi A. The Many Shapley Values for Model Explanation. In: Proceedings of the 37th International Conference on Machine Learning. PMLR; 2020:9269–9278. Available from: https://proceedings.mlr.press/v119/sundararajan20b.html. Accessed May 29, 2023.
33. Wei P, Lu Z, Song J. Variable importance analysis: a comprehensive review. *Reliab Eng Syst Saf*. 2015;142:399–432. doi:10.1016/j.ress.2015.05.018
34. Gromping. Variable importance assessment in regression: linear regression versus random forest. AMER STATIST; 2009. Availbale from: https://xueshu.baidu.com/usercenter/paper/show?paperid=033d60c22092b43a73121ab8c3d9d3e5&site=xueshu_se. Accessed May 27, 2023.
35. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One*. 2018;13(7):e0201016. doi:10.1371/journal.pone.0201016
36. Grekousis G, Feng Z, Marakakis I, Lu Y, Wang R. Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: a geographical random forest approach. *Health Place*. 2022;74:102744. doi:10.1016/j.healthplace.2022.102744
37. Chang W, Ji X, Xiao Y, et al. Prediction of hypertension outcomes based on gain sequence forward tabu search feature selection and XGBoost. *Diagnostics*. 2021;11(5):792. doi:10.3390/diagnostics11050792
38. Molnar C. Interpretable Machine Learning. Available from: https://christophm.github.io/interpretable-mL-book/. Accessed July 23, 2023.
39. Deng QW, Liu WB. Analysis of health-related quality of life and influencing factors in patients with stroke. *J Nanj Med Univer*. 2020;20(5):459–463.
40. Bruthans J, Mayer O, De Bacquer D, et al. Educational level and risk profile and risk control in patients with coronary heart disease. *Eur J Prev Cardiol*. 2016;23(8):881–890. doi:10.1177/2047487315601078
41. Wahab KW, Kolo PM, Salawu FK, Sanya EO. Blood pressure control among hypertensive stroke survivors in nigeria. *J Stroke Cerebrovasc Dis*. 2017;26(6):1222–1227. doi:10.1016/j.jstrokecerebrovasdis.2017.01.011
42. Craig A, Mtintsilana A, Mapanga W, Dlamini SN, Norris SA. Socioeconomic position, perceived weight, lifestyle risk, and multimorbidity in young adults aged 18 to 35 years: a Multi-country Study. *BMC Public Health*. 2023;23(1):1360. doi:10.1186/s12889-023-16234-1
43. Gudina AT, Cheruvu VK, Gilmore NJ, et al. Health related quality of life in adult cancer survivors: importance of social and emotional support. *Cancer Epidemiol*. 2021;74:101996. doi:10.1016/j.canep.2021.101996
44. Sadlonova M, Wasser K, Nagel J, et al. Health-related quality of life, anxiety and depression up to 12 months post-stroke: influence of sex, age, stroke severity and atrial fibrillation – a longitudinal subanalysis of the Find-AFRANDOMISED trial. *J Psychosom Res*. 2021;142:110353. doi:10.1016/j.jpsychores.2020.110353
45. McClave AK, Dube SR, Strine TW, Mokdad AH. Associations between health-related quality of life and smoking status among a large sample of U.S. adults. *Preventive Med*. 2009;48(2):173–179. doi:10.1016/j.ypmed.2008.11.012
46. Liao W, Liu X, Kang N, et al. Associations between healthy lifestyle score and health-related quality of life among Chinese rural adults: variations in age, sex, education level, and income. *Qual Life Res*. 2023;32(1):81–92. doi:10.1007/s11136-022-03229-5
47. GBD 2020 Alcohol Collaborators. Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020. *Lancet*. 2022;400(10347):185–235. doi:10.1016/S0140-6736(22)00847-9
48. Liu N, Zeng L, Li Z, Wang J. Health-related quality of life and long-term care needs among elderly individuals living alone: a cross-sectional study in rural areas of Shaanxi Province, China. *BMC Public Health*. 2013;13(1):313. doi:10.1186/1471-2458-13-313
49. Sun YA, Kalpakavadi S, Prior S, et al. Socioeconomic status and health-related quality of life after stroke: a systematic review and meta-analysis. *Health Qual Life Outcomes*. 2023;21(1):115. doi:10.1186/s12955-023-02194-y
50. Rizal H, Said MA, Abdul Majid H, et al. Health-related quality of life of younger and older lower-income households in Malaysia. *PLoS One*. 2022;17(2):e0263751. doi:10.1371/journal.pone.0263751
51. Chen S, Ling J, Cheng Y. Physical activity and body mass index were interactively related to health-related quality of life among older adults. *Arch Gerontol Geriatrics*. 2023;104:104833. doi:10.1016/j.archger.2022.104833
52. Choi MJ, Park YG, Kim YH, Cho KH, Nam GE. Association between type of exercise and health-related quality of life in adults without activity limitations: a nationwide cross-sectional study. *BMC Public Health*. 2020;20(1):599. doi:10.1186/s12889-020-08699-1
53. Levine D, Poole C, Hostetler C, et al. The association between sleep and health-related quality of life in stroke survivors. *Arch Phys Med Rehabil*. 2021;102(10):e28–e29. doi:10.1016/j.apmr.2021.07.541