ORIGINAL RESEARCH

Correlation Analysis of Persistence and Recurrence of Stroke in Young Patients Based on Big Data in Healthcare

Wei Duan, BM; Xiaojun Fu, BM; Ping Yuan, BM

ABSTRACT

This study aims to analyze the correlation between the persistence and recurrence of stroke in young patients via big data in healthcare. It provides an in-depth introduction to the background of big data in healthcare and a detailed description of stroke symptoms, so as to better apply the Apriori parallelization algorithm based on compression matrix (PBCM) algorithm against the background of big data in healthcare to analyze it. In our study, patients were randomly divided into 2 groups. By observing the different persistent relationships in the groups, the factors affecting the patients' fasting blood glucose (FBG), glycosylated hemoglobin (HbA_{1c}), blood pressure (BP), blood lipids, alcohol consumption, smoking and so on were analyzed. The National Institute of Health Stroke Scale (NIHSS) score, FBG, HbA_{1c}, triglycerides (TG), high-density lipoprotein (HDL), body mass index (BMI), length of hospital stay, gender and high BP, diabetes, heart disease, smoking and other factors affect the recurrence rate of stroke as they all affect the brain, although they are all statistically different ($P < .05$). The recurrence of stroke requires more attention in the treatment of stroke. (*Altern Ther Health Med.* 2023;29(4):110-119).

Wei Duan, BM, Dental Department; **Xiaojun Fu, BM,** Respiratory Medicine Department; **Ping Yuan, BM,** ENT Department; Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China.

Corresponding author: Ping Yuan, BM
E-mail: 245684665@qq.com

INTRODUCTION

The 3 major causes of human death include heart disease, malignant tumors and cerebrovascular disease (CVD). According to recent relevant epidemiological data in China, CVD ranks first on the list of causes of death. Compared with western developed countries, the morbidity and mortality of CVD in China are significantly higher than that of cardiovascular (CV) disease. Among them, ischemic stroke is the most common type of CVD at 70%.

Traditional research considered the most important cause of ischemic stroke to be intracranial atherosclerotic plaque. However, with the development of the Chinese economy and society, the current incidence of ischemic stroke is tending toward an increase in younger individuals. The most common cause of arterial stenosis is intracranial atherosclerosis, and intracranial arterial stenosis in young patients is due to not only intracranial atherosclerosis, but a variety of other causes. However, the primary cause of intracranial artery stenosis in young patients is still unknown, primarily because it is difficult to obtain materials for the pathological analysis of intracranial blood vessels.

In recent years, there has been an increasing number of studies on stroke in the young. Advances in information and communication technology have led to the rapid growth of healthcare data available to various participants in the medical industry. The evolution from paper databases to electronic records shows the continuous pogress of medical information systems. Medical institutions are paying more and more attention to how to apply big data in healthcare, but most of them are struggling to find ways to make full use of it.

Chen's 2018 study made Taiwan's medical industry the research object, used a hybrid method to evaluate, predict and summarize the main applications of big data in healthcare, and established a strategic path that medical institutions should follow in accordance with the different dimensions of the application. First, he reviewed the relevant literature on the application of big data in healthcare and conducted interviews with relevant stakeholders. Then, he conducted content analysis to extract key applications, and use decision-making trial and evaluation laboratory (DEMATEL) to find their network relationship maps (NRM). However, his research methods are not very helpful in targeted areas.[1]

The purpose of the Abanoz Y, et al study was to analyze the relationship between migraine and ischemic stroke in a hospital-based cohort. They included 202 consecutively diagnosed patients with ischemic stroke were between 15 and 50 years of age, old, age and gender matched to a control group of 250 volunteers with no history of stroke. All participants received a migraine questionnaire. According to neuroimaging findings, ischemic lesions are divided into anterior and posterior circulation.[2]

Aigner A et al found that the etiology and risk factors of stroke in young people are different from those in the elderly. The purpose of his research was to study the impact of established potentially modifiable cardiovascular (CV) risk factors on the burden of stroke in young people. A German national case-control study based on the 2007-2010 SIFAP1 study (young patients with Fabry disease with stroke) and the population-based GEDA controlled study conducted from 2009 to 2010 (German Health Update). There were 2155 consecutive cases in 26 clinical stroke centers in patients aged 18-55 years with acute first stroke; the control group (age and gender matched, n = 8500, no stroke) came from a national community sample. The adjusted population attributable risk for 8 risk factors (hypertension [HTN], hyperlipidemia [HLP]), diabetes [DM], coronary heart disease [CHD] smoking, heavy paroxysmal drinking, low physical activity, and obesity) were calculated and their effects on all strokes, ischemic strokes and the combined effect of primary cerebral hemorrhage. However, he did not elaborate on a treatment plan for ischemic stroke in young people.[3]

The propositions of our study are: (1) Combining the background of big data in healthcare, there is more medical data about young patients with stroke, which is conducive to research; (2) Studies are proposed from different angles, and there are many previous studies. Ours is based on analysis of medical treatment in elderly patients with stroke, but is aimed at analysis of medical treatment in young patients with stroke, and strives to solve the recurring problems in this population.

ANALYSIS OF THE PERSISTENCE AND RECURRENCE OF STROKE IN YOUNG PATIENTS BASED ON BIG DATA IN HEALTHCARE

Big Data in Healthcare

With the progress and advances in medical and health undertakings, medical data generated by hospitals is increasing exponentially.[4] Big data in healthcare contains valuable information that needs to be mined, however, the massive amount of data makes traditional data mining methods obsolete.[5] How to use technology to mine and analyze this huge amount of data, tease out the valuable findings, and provide improved disease prevention and treatment has become an urgent problem.[6,7]

The development of the healthcare industry is closely related to people's lives and health. As science and technology are integrated into the medical industry,[8] a series of reforms have been made in medical-related fields.[9] Hospital treat a lot of patients every day, and over time a lot of data are accumulated. Big data in healthcare comes from various aspects such as clinical diagnosis, physical examination, health websites, medical research and so on that are rich sources of information.

How to Use Data Mining Technology to Find Meaningful Information in a Large Number of Medical Data Sets to Provide Physicians with Diagnoses

With certain technical support, big data related-technologies have become an inevitable choice for information mining in medicine.[10,11] In this context, traditional medical data processing technology will inevitably play a role, but when doctors diagnose patients, they can use big data analysis tools to analyze diseases and find a scientific basis for clinical diagnosis.[12] Cloud computing technology provides a brand-new solution for solving such problems.[13] It is considered to be the advanced arrangement and sublimation of parallel computing, grid computing and distributed computing. Cloud computing can combine information on various computers, allowing users to use resources conveniently.[14,15] Among the many platforms for cloud computing, Hadoop has attracted the attention of most enterprises and individual researchers due to its open source and high scalability, which make this computing model very suitable for distributed data mining.[16]

Stroke Disorders in the Young

Cerebrovascular disease (CVD) is a common and frequently-occurring disease of the central nervous system (CNS).[17] Ischemic stroke accounts for the majority of strokes, and is usually caused by cerebral arterial embolism and thrombotic occlusion. Ischemic stroke accounts for approximately 87% of all stroks types, and the annual growth rate is as high as 8.7%.[18,19] CVD, together with coronary heart disease (CHD) and malignant tumors, constitute the 3 major diseases threatening human lives.[20] The disability and fatality rates in patients with recurrence of CVD are higher, and the recurrence rate is extremely high at approximately 14% to 20%.[21] Interventions with regard to the risk factors for ischemic stroke can effectively decrease the recurrence of stroke, which is of great significance for reducing the burden on family and society.

There are many risk factors for recurrent stroke, which are divided into controllable and uncontrollable. Controllable factors include DM, HLP, HHcy, hyperuric acid, carotid atherosclerotic plaque, CHD, atrial fibrillation (AF), hypertension (HTN) and so on.[22] Uncontrollable factors include gender, age, family history of stroke, etc.[23] The recurrence of stroke is a heavy blow to both the patient's family and society as a whole. Many studies have focused on the high-risk factors for stroke recurrence and have reached inconsistent conclusions.[24] This study analyzes the relevant risk factors for recurrence of ischemic stroke by studying the persistence of stroke in young patients with medical treatment, how to prevent stroke recurrence, how to reduce the recurrence and fatality rate of stroke, and reduce the family and social burden.[25]

Parallelization of the Apriori Algorithm Based on Compression Matrix in the Context of Big Data in Healthcare

The Apriori PBCM algorithm is an improved algorithm based on a compressed matrix. The main idea of the algorithm is to describe the transaction database to be processed in the form of a Boolean matrix, in which rows represent transactions and columns represent items. Items in the matrix are not like traditional algorithms. Write-specific items like that, but, etc use 0 and 1. The transformed Boolean matrix occupies less space, because the items in the matrix are all Boolean. Moreover, when seeking the support of the candidate item set, it is no longer necessary to scan the database. You only need to do the inner product of the vector between the columns of the matrix, which greatly simplifies the method of calculating the support of the candidate item set and reduces the data. The set occupies a lot of space.

Suppose there is an arbitrary database "T" with a mapping relationship such as: g: $U \rightarrow T$; that is, g(U) = T_{m*n} where m is the number of transactions, n is the project, and T is defined as:

$$T_{ij} = \begin{cases} 1, P_j \in Q_i \\ 0, P_j \notin Q_i \end{cases} \quad (1)$$

Among these, $1 \le i \le m$, $1 \le j \le n$.

According to the definition of the matrix, the Boolean matrix can be constructed as follows: each row represents a transaction and each column represents an item. When the corresponding item exists, uij=1, otherwise it is equal to 0.

$$G_{m\times n} = \begin{bmatrix} t_1 & t_2 & \cdots & t_{1n} \\ t_2 & t_2 & \cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & \cdots & t_m \end{bmatrix} \quad (2)$$

When there are redundant and repeated transactions in the database, the transformed matrix is as follows:

$$G_{p\times n} = \begin{bmatrix} t_1 & t_2 & \cdots & t_{1n} \\ t_2 & t_2 & \cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{p1} & t_{p2} & \cdots & t_{pn} \end{bmatrix} \quad (3)$$

in which p represents the number of rows of the Boolean matrix, that is, the number of transactions t do not include repetitions, $1 \le p \le m$.

The vector of candidate 2-items set (Ui, Uj) is expressed as:

$$GU_{ij} = GU_i \wedge GU_j = \begin{bmatrix} t_{1i} \wedge t_{1j} \\ t_{2i} \wedge t_{2j} \\ \cdots \\ t_{mi} \wedge t_{mj} \end{bmatrix} \quad (4)$$

Among these, $\wedge$ represents the AND operation. The candidate 1-itemset support count formula is:

$$Support\_count(U_j) = \sum_{i=1}^{q}(t_{ij} \times WE[i]) \quad (5)$$

Among these, UW[i] represents the corresponding transaction weight, and its value represents the number of repeated transactions, $1 \le p \le m$.

The candidate 2-itemset support count formula is:

$$Support\_count(U_i, U_j) = \sum_{l=1}^{q}(t_{li} \wedge t_{lj} \times WE[l]) \quad (6)$$

$$Sup = ((t_{1i} \wedge t_{1j} + t_{2i} \wedge t_{2j} + \ldots + t_{pi} \wedge t_{pj} + t_{pi} \wedge t_{pj}) \times WE[l]) \quad (7)$$

Among these, UW[k] represents the corresponding transaction weight, $1 \le p \le m$.

The candidate k-item set support count formula is:

$$Sup(U_1, U_2, \ldots, U_l) = \sum_{s=1}^{q}(t_{s1} \wedge t_{s2} \ldots \wedge t_{sl} \times WE[s]) \quad (8)$$

Among these, UW[s] represents the corresponding transaction weight, $1 \le p \le m$.

Generate frequent k+1 itemsets according to the above method. According to the theorem, judge whether the number of frequent k-item sets is > or = K+1; if so, continue to repeat the third, fourth and fifth step; otherwise the loop terminates and the algorithm ends.

We divided the data file that needed to be processed into N data blocks of the same size (n1, n2,..., nN) — that is, the data fragment InputSplit — and sent these data fragments to M nodes (m1, m2) ,...,mM and so on.

Scanning of each node according to the definition is converted into matrix A, and according to the matrix column vector operation, the local support count of each node candidate 1-itemset u_1mqjSp(I) was obtained, and the local support count of each node candidate 1-itemset was sent to the host The node statistics obtained the candidate 1-itemset global support count u_1jSpsum(I) and the minimum support count Min_sup for comparison, and get the final frequent 1-itemset L1.

The calculation formula for the local support count of each node candidate 1-itemset is:

$$Sup_{mq\_}1(U_j) = \sum_{i=1}^{Y_q}(t_{ij} \times \mathrm{WE}[i]) \quad (9)$$

Among these, UW[i] represents the corresponding transaction weight, mq represents the qth node $1 \le p \le M$, Tp represents the number of transactions in any data block pn when generating frequent 1-item sets, $1 \le p \le N$. The candidate 1-item set global support count calculation formula is:

$$Sup\_sum1(U_j) = \sum_{mq\text{-}1}^{M} \mathrm{Sup}_{\mathrm{mq\_}}1(U_j) \quad (10)$$

Sp(I) represents the candidate 1-itemset support count of the qth node, where $1 \le p \le M$.

Among these, UW[s] represents the corresponding transaction weight, mq represents the qth node and $1 \leq p \leq M$. Tr represents the number of transactions in any data block rn when generating frequent k-items sets, $1 \leq p \leq N$. The calculation formula for the global support count of candidate k-items set is:

$$Sup_{mp\_}1(U_1, U_2, \dots, U_l) = \sum_{s=1}^{T_l} (t_{s1} \wedge t_{s2} \wedge \dots\ t_{sl} \times WE[s]) \quad (11)$$

Among these, u_mqSpk represents the support count of candidate k-item sets for the qth node, where $1 \leq p \leq M$. The fourth step is to compress the matrix Ak-1. According to the transaction compression theorem: compress the transaction, delete the row vector that does not meet the requirements and delete the column vector corresponding to the itemset whose 12u_kSpsum(kI,I,...,I) value is <Min_sup, and readjust each Node matrix dimension to get a new matrix Ak.

$$Sup\_sum1(U_1, U_2, \dots, U_l) = \sum_{mq=1}^{M} Sup_{mq\_}l(U_1, U_2, \dots, U_l) \quad (12)$$

The fifth step is to loop through the second, third and fourth steps, and count the number of frequent k-itemsets KcountL. If KcountL means that there will be no more frequent k+1-itemsets generated, the algorithm ends; otherwise the algorithm continues to the next time iteration.

Among these, rule support and confidence are 2 very important concepts. For rules X and Y:

The support calculation method of X⇨Y is:

$$Support\ (X \Rightarrow Y) = P(X \cup Y) \quad (13)$$

The calculation method of X⇨Y is:

$$Confidence\ (X \Rightarrow Y) = P(X|Y) = P(XY)P(X)\text{'s confidence is:} \quad (14)$$

CORRELATION ANALYSIS EXPERIMENT OF PERSISTENCE AND RECURRENCE OF STROKE IN YOUNG PATIENTS BASED ON BIG DATA IN HEALTHCARE

General study information

In our study, 120 patients with recurrent stroke who were admitted to the Hospital of Chengdu University of Traditional Chinese Medicine in China from January 2019 to December 2020 were selected as the observation group and 120 young patients with initial stroke were randomly selected during the same period and served as the control group. Recurrent stroke patients (recurrence group) contained 63 women and 57 men, aged 32 to 45 years, with an average age of (33.42 ± 3.9) years. There were 59 men and 61 women patients with primary stroke (initial group), aged 30 to 59 years, with an average age of 31.29 ± 4.5 years. Table 1 show patient inclusion and exclusion criteria.

Table 1. Study Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
Acute CVD (onset within 2 weeks)	Bilateral MCA stenosis or clear Moyamoya disease
Ischemic CVD (cerebral infarction and TIA)	Extracranial carotid or vertebral artery stenosis
No metal stents, pacemakers, dentures; no claustrophobia	Patients with hyperacute onset or undergoing arterial venous thrombolysis
Age 18-50 years	Poor image data obtained
No gender limit	Cerebral embolism
The first MRA after admission can be performed within 3 days after the diagnosis, and patients have complete biochemical, liver function, kidney function, blood sugar, blood lipids tests	Patients who are critically ill or have acute myocardial infarction, severe arrhythmia, etc. and should not be moved
Informed consent of patients and their families	Incomplete data collection

Abbreviations: CVD, cerebrovascular disease; MCA, middle cerebral artery; MRA, magnetic resonance angiography; TIA, transient ischemic attack.

Diagnostic Criteria

The diagnostic criteria for ischemic stroke are in line with the revised diagnostic criteria for CVD from the Fourth National Cerebrovascular Disease Conference (1996). The lesions were confirmed by head computed tomography (CT) or magnetic resonance imaging (MRI) scan. Diagnostic criteria for recurrent stroke were: the original signs or symptoms were aggravated or new neurological deficit symptoms appeared and met the diagnostic criteria for ischemic stroke. Diagnostic criteria for HTN: without the use of antihypertensive drugs, systolic blood pressure (SBP) ≥140 mmHg; diastolic blood pressure (DBP) ≥90 mmHg; currently using antihypertensive drugs; previous history of HTN. DM diagnosis criteria: DM symptoms and random blood glucose ≥11.1 mmol/l; FBG ≥7.0 mmol/l; or OGTT 2h ≥11.1 mmol/l, or a history of DM or use of hypoglycemic drugs or undergoing diet control treatment. Diagnostic criteria for smoking: at least 1 cigarette per day for more than 1 year. Diagnostic criteria for depression: depression as determined by Hamilton score.

Experimental Research Process

A total of 120 young patients with stroke admitted to the Hospital of Chengdu University of Traditional Chinese Medicine in China from January 2019 to December 2020 were selected as the observation group and 120 patients with initial stroke in the same period were randomly selected as the control group according to the random number table method. Inpatient information was obtained, including past medical history, age, gender, current medical history, chief complaint, family history, etc. Imaging examinations included brain parenchymal examination (CT/MRI) and CV examination (magnetic resonance angiography [MRA]/CT/angiography [CTA]). Laboratory tests included FBG, Hb/A1c, BP, depression scale score, etc. All data were statistically analyzed by multivariate logistic regression and chi-square test, and P <.05 was considered statistically significant. The general data in the 2 groups were compared (P>.05).

Relevant informed consent was obtained from the patients or their families.

Judgment Criteria for Risk Factors

(1) HTN. When the patient is in a calm state, monitor blood pressure; frequency should be no less ≥3 times, SBP ≥140 mmHg or DBP ≥90 mmHg; history of HTN. SBP 140-159 mmHg or DBP 90-99 mmHg is considered Grade I; SBP 160-179 mmHg or DBP 100-109 mmHg, Grade II; SBP is not <180 mmHg or DBP <110, Grade III.

(2) DM. According to the research report issued by the World Health Organization (WHO): Patients who have been confirmed to have diabetes and have undergone active and effective treatment before enrollment, or who have been confirmed with acute cerebral infarction after admission. Later, patients' FBG was tested and found to be not <7.0 mmol/L when the patient was fasting or was not <11.1 mmol/L within 120 minutes of eating or random testing.

(3) Heart disease. Specifically related to congenital heart diseases such as AF or atrial flutter, CHD, dilated cardiomyopathy, endocarditis, valvular disease or valve replacement and other congenital heart diseases and heart malformations.

(4) Dyslipidemia. According to the relevant standards in the Cholesterol Index issued by the American Academy of Medicine, TC value is not ≤5.18 mmol/L and/or is not ≤1.7 mmol/L, and/or HDL is not ≤1.04 mmol/L, and/or LDL is not ≤3.38 mmol/L, and/or lipoprotein (a) (Lp[a]) is >1.35 mmol/L, and/or apolipoprotein A1 (Apo-A1) <1.0 g/L, and/or apolipoprotein B100 (Apo-B100) >1.0 g/L, or have taken drugs that have a lowering effect on blood lipids in any case that meets the above standards.

(5) HHcy. Level is not >15 umol/L.

(6) Smoking history. The amount of smoking is determined by the number of smoking pack years = the number of cigarettes smoked per day/20 × the number of smoking years.

(7) **Drinking history**. Divided into occasional, light, moderate and heavy drinking; the standard classification is shown in Table 2.

(8) Carotid artery atherosclerosis. Present if the thickness of the carotid artery intima-media is >0.9 mm; if it is >1.2 mm, it is considered plaque formation.

(9) Family history. A history of stroke in siblings or parents.

(10) MRA/CTA/digital subtraction angiography(DSA). Diagnostic criteria for intracranial vascular stenosis: If the lumen of the vessel on MRA is reduced by >50%, it is considered stenosis. The CTA criteria for vascular stenosis: if the lumen is reduced by >30%, stenosis is present.

(11) Hyperuricemia. Uric acid ≥420 umol/L.

(12) Abnormal rheumatic immune indicators. Erythrocyte sedimentation rate, positive rheumatoid factor, antinuclear antibody, high-sensitivity C-reactive protein.

The etiology, classification and diagnostic criteria of stroke are based on analysis of the patient's past medical history, current disease status, physical signs and laboratory index values, and also comprehensive consideration of any unhealthy living habits and/or other factors that enable a Trial of ORG 10172 in Acute Stroke Treatment (TOAST) determination. The most important group of tests is the cranial MRI+diffusion-weighted imaging (DWI)+MRA examination, which is of great help in confirming the specific location of the patient's lesions, and can also clarify branching and stenosis of the cranial blood vessels.

Specific stroke diagnosis plays an important role in the treatment plan

Large atherosclerotic (LAA type). If the amount of vascular stenosis is ≥50% and the main predisposing factor is atherosclerosis; and if the acute infarction occurs in the responsible blood vessel supply area, it can be sufficient for the diagnosis of an LAA. Generally, if at least 1 monocular transient ischemic attack (TIA) has occurred previously, or the patient has had unilateral internal watershed infarction, especially when there are multiple infarcts around the injured blood vessel, this diagnosis is confirmed. Risk factors include diabetes, high blood pressure, CHD, HLP, drinking alcohol, smoking, etc.

Cardioembolic (**CE**). The patient's cerebral infarction caused by CE can become the source of further thromboembolism. Specifically, sick sinus syndrome, AF, rheumatic mitral valve or aortic valve disease, dilated cardiomyopathy, non-bacterial thrombotic endocarditis, and others are non-thrombotic embolism sources, including left atrial myxoma, infective endocarditis, etc.

Small artery occlusion (SAO). Refers to a single acute responsible lesion found in the basal ganglia or brainstem, with a diameter <20 mm. At the same time, no other symptoms of lesions are found in the main artery. TIA has occurred approximately 7 days before the test, and the patient's symptoms are more consistent with lacunar infarction.

Hypercoagulable state such as intravascular microembolism.

Other causes (OC). The patient's symptoms may be caused by vasculitis, blood disease, genetic metabolism, and Moyamoya disease; there are relatively certain signs of special lesions, etc. This will directly affect the arteries corresponding to the clinical symptoms and cause them to be damaged; there are relatively positive signs of special lesions, and the time and specific location of their appearance are directly related to cerebral infarction, such as from arterial dissection or arterial surgery.

Coronary interventional therapy of uncertain (**UC**) **type**. During the evaluation, no abnormalities were found in the intracranial blood vessels but thrombotic occlusion occurred suddenly; or the same blood vessel changed from the occluded state to the fully recanalized state; or acute multiple infarctions occurred. The time of appearance of the various lesions was very close, but no abnormalities or phenomena were observed in the blood vessels.

Table 2. Alcohol Drinking Classification Standards

Frequency of drinking	Drinking a small amount at a time	Amount of alcohol per drink	Drinking a large amount at a time
Several times a year	Drinking occasionally	Light drinking	Moderate drinking
1-3 times a month	Light drinking	Moderate drinking	Heavy drinking
Every week	Moderate drinking	Heavy drinking	Heavy drinking

Note: A small amount is 0-50 ml white wine or 0-600 ml beer or 0-250 ml red wine; medium amount is 50-100 ml white wine or 600-1200 ml beer or 250-500 ml red wine; large amount is >100 ml white wine or >1200 ml beer or >500 ml red wine

STATISTICAL ANALYSIS

All data were statistically analyzed using IBM® SPSS 17.0 software. The first group was defined as 1, and the relapse group was defined as 0.8 ,indicators corresponding to variables of gender, age, smoking, hyperglycemia, HTN, depression, gender and family history of stroke. Each indicator was assigned a value according to its status, 1 for males, 2 for females: smoking, hyperglycemia, high blood pressure, depression, gender; family history of stroke was assigned a value of 1 and if not present a value of 0. Chi-square test was used for univariate analysis, and $P < .05$ was considered statistically significant. Logistic regression analysis was performed on the collated data with SPSS 17.0 software, and $P < .05$ was considered statistically significant.

Correlation Analysis of Persistence and Recurrence of Stroke in Young Patients Based on Medical Big Data

General data analysis of young patients with stroke

Table 3 shows the general demographic and sociological data of young patients with stroke. Of the patients, 68.67% were men and 31.33% were women, with an age range of 32 to 45 (average age, 33.58 ± 2.74) years. A total of 87.98% of patients were married. In terms of education level, high school or technical secondary school graduate was the highest (36.91% of patients) and elementary school graduate and below was the lowest (17.17% of patients). In terms of current working status, The majority of patients were unemployed or on sick leave, accounting for 42.92% and 46.78%, respectively; most patients had a per capita monthly income of 2000-3000, with city medical insurance accounting for the source of income in 49.36% of patients.

Table 4 shows the disease-related data from the study patients. Among them, ischemic stroke accounted for 82%, hemorrhagic stroke for 18.%; 71 cases showed effects for less than 2 weeks, 42 for more than 1 month, and 120 for 2 weeks to 1 month. The majority of patients were partially aware of their disease status (49.88%), the percentage of patients with no effect after stroke was 48.69%, while the higher percentage of patients with effect was 2 types (46.28%). The degree of patient self-care was mild dependence, accounting for 35.19% and partial dependence, 25.32%. Limb dysfunction was mainly mild (5.50% of patients) and moderate (21.03%). The number of post-stroke patients with type 1 complications was higher by 93 cases (35.16%). In 96.57% of patients, this was first onset of stroke.

Table 3. Demographic and Sociological Data in Young Patients with Stroke

Variable Grouping	Number of Patients	Composition Ratio
Gender		
Male	88	67.68
Female	64	31.32
Marital status		
Unmarried	33	3.25
Married	150	87.94
Divorced	42	10.38
Education level		
≤Elementary school	22	17.69
Junior high school	41	26.58
High school or technical school	38	39.49
>College degree	26	19.73
Current job status		
Unemployed	44	43.69
Sick leave	46	12.34
Still working	88	13.42
Household monthly income per capita		
<2000	24	24.58
2000-5000	38	23.16
>5000	46	7.32
Payment methods		
Own expense	22	9.45
NCMS	51	21.88
City Medical Insurance	115	49.37
Provincial Medical Insurance	45	19.32

Abbreviations: NCMS, New Cooperative Medica Scheme.

Imaging Analysis in Young Patients with Stroke

As a new method in the development of traditional intracranial imaging, vessel wall magnetic resonance imaging (vwMRI) cannot only show the degree of vascular stenosis as well as the gold standard DSA, but also clearly show the structure of the vessel wall, plaque composition and plaque location. It is a non-invasive and non-radiation inspection method, and will become a commonly used technique for CV evaluation. At the same time, it can also assist in the diagnosis of cerebral infarction types that are difficult to determine with traditional imaging, predict the occurrence and development of stroke and of great help in the differential diagnosis of other CVD. In addition, because the pathological examination of intracranial blood vessels is difficult to

Table 4. Disease-Related Data in Young Patients with Stroke

Variable Grouping	Number of patients	Composition ratio
Stroke type		
Ischemic	91	30.49
Hemorrhagic	43	22.49
Sick time		
<2 weeks	72	62.19
2 weeks-1 month	46	18.03
>1 month	22	28.21
Understand the condition?		
No understanding at all	43	21.55
Partial understanding	66	49.88
Full understanding	59	11.26
After effects		
No	52	48.69
1 type	114	32.18
2 types	29	46.28
≥3 types	12	20.18
With or without complications		
None	54	24.13
1 type	93	35.16
2 types	61	25.43
≥3 types	22	15.42
First stroke?		
Yes	122	67.48
No	36	25.36

perform, the imaging findings of intracranial blood vessels cannot truly correspond to the anticipated pathological changes. More research on high resolution MRI (HR-MRI) of intracranial blood vessels is needed to further understand the correlation between blood vessel imaging changes and clinical symptoms, as well as various laboratory indicators, and improve the prevention and treatment of stroke.

Persistence of Medical Treatment in Patients with Stroke

As shown in Figure 2, the duration of medical treatment in young patients with stroke is shorter than in elderly patients with stroke. In terms of long-term treatment, young patients with stroke receive poor medical treatment for various reasons, and the long-term treatment effect is not good. Therefore, young patients with stroke must insist on seeking medical treatment in order to recover as quickly as possible.

Correlation Analysis of Stroke Disease Recurrence

The general clinical data distribution of eccentric and concentric stenosis is shown in Figure 3. Age and atherosclerosis risk factors (HTN, HLP, DM, whether or not there is a family history of stroke, smoking and drinking alcohol) were compared and analyzed. It was found that 65 patients (68.4%) had eccentric stenosis and 30 (31.6%) had centripetal stenosis. The patients with eccentric stenosis were older (35.3 ± 10.6 vs 30.1 ± 6.4 years, respectively; $P < .05$), and of the atherosclerotic factors, HTN (72.3% vs 56.7%; $P = .001$) was more inclined to lead to eccentric stenosis.

Figure 1. Medical images of young patients with stroke. (**1A**) MRA shows irregular mild arterial stenosis near the paramedian pontine artery. (**1B**) DWI image shows acute cerebral infarction of the right pontine. (**1C** and **1D**) T1WI and T2WI of VWMRI showed a low or iso signal density lipid-rich necrotic eccentric plaque on the cross section.

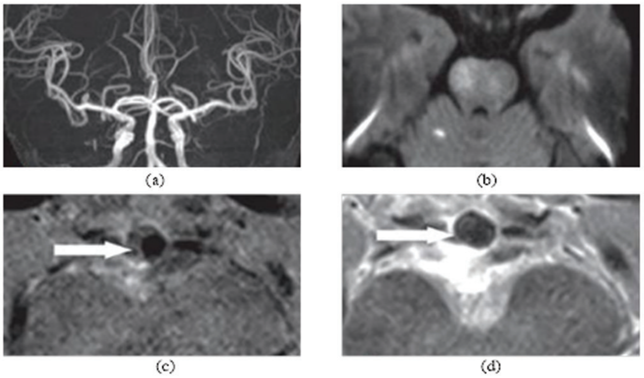


Figure 2. Continuity analysis of medical treatment.

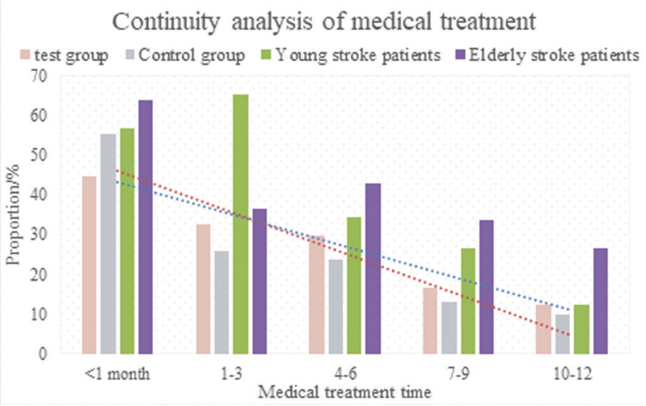


Figure 3. Comparative analysis of different types of stenosis.

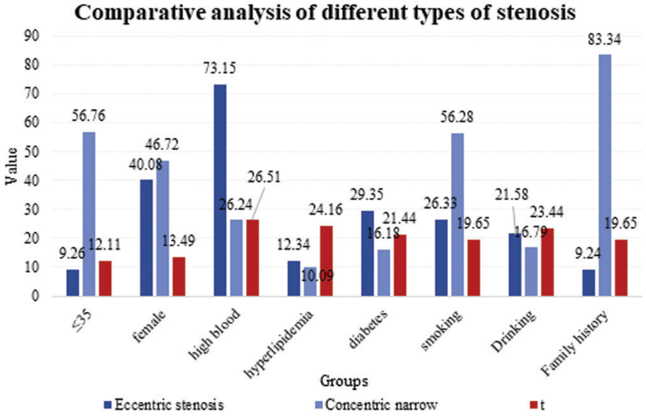


Figure 4. Single factor binary logistic regression analysis of centripetal stenosis.

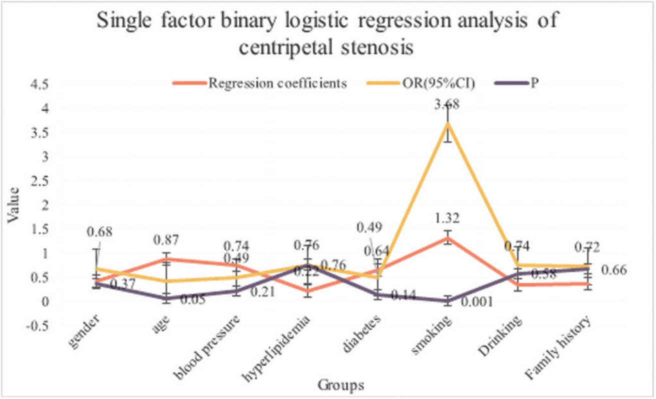


Table 5. Risk Factor Regression Assignment Table

Factor	Variable assignment
Gender	Male=1, female=0
TOAST typing	LAA=1, CE=2, SAO=3, OC=4, UC=5
NIHSS score	0-1=1,2-4=2,5-15=3,16-20=4,21-42=5
Hypertension	None=0, grade I=1, grade II=2, grade III=3
Diabetes	No=0, yes=1
Smoking history	No=0, yes=1
Drinking history	No=0, yes=1
Family history	No=0, yes=1
Coronary Heart Disease	No=0, yes=1
Low-density lipoprotein abnormalities	No=0, yes=1
Cholesterol abnormalities	No=0, yes=1
Triglyceride abnormalities	No=0, yes=1
Hyperhomocysteinemia	No=0, yes=1
Hyperuricemia	No=0, yes=1
Carotid artery atherosclerosis	No=0, yes=1
Intracranial vascular stenosis	No=0, yes=1
Abnormal immune indicators	No=0, yes=1
Cerebral infarction	Youth = 0, middle-aged and elderly = 1

Figure 5. T test results of quantitative data in experimental and control groups.

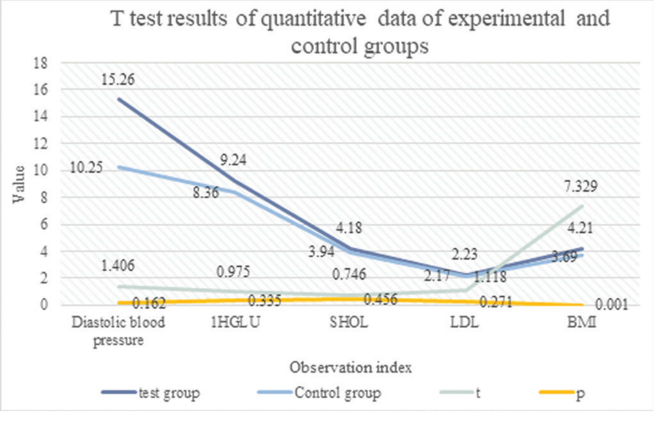
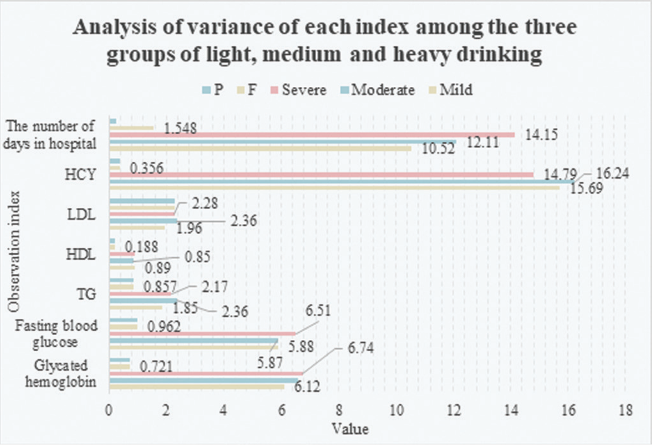


Figure 6. Analysis of variance of each index in the 3 light, medium and heavy drinking groups.



Univariate logistic regression analysis of centripetal stenosis is shown in Figure 4. The regression coefficient of the study age group was -0.87; OR = 0.41 (0.17-0.91), indicating that the age-independent effect and the occurrence of centripetal stenosis are negatively correlated, and are greater ≥35 years of age. The incidence of concentric stenosis in patients was 0.41 times that in patients >35 years of age. The older the patient, the lower the probability of concentric stenosis. The smoking regression coefficient was OR = 3.69 (1.49-9.17), indicating that smoking and centripetal stenosis have a positive correlation. The probability of concentric stenosis in smokers is 3.69 times that in non-smokers.

Table 5 shows univariate logistic regression analysis between young patients with cerebral infarction with regard to gender, TOAST classification, NIHSS score, HTN, DM, smoking history, alcohol consumption, history, family history, CHD, dyslipidemia and HHcy. Values are given for blood disease, hyperuricemia, carotid atherosclerosis, intracranial vascular stenosis and abnormal immune indicators.

It can be seen in Figure 5 that the indicators ($P < .05$) after statistical analysis include NIHSS score, FBG, HbA_{1c}, TG, HDL, body mass index BMI, length of hospital stay and gender. The number of people with high BP, DM, heart disease, smoking, and the above indicators was statistically significant, indicating that these values are different in the two groups. Age, SBP, DBP, oral glucose tolerance test (OGTT) 2-hour blood glucose (2HGLU), total cholesterol (CHOL), LDL and blood homocysteine (HCY) were analyzed by statistical analysis ($P > .05$). There was no statistically significant difference between the two groups.

The 120 patients who were clinically diagnosed as having stroke were then divided into 3 groups according to the severity of obstructive sleep apnea-hypopnea syndrome (OSAHS), and were statistically analyzed. OSAHS severity was mainly based on the Apnea Hypoventilation Index (AHI) score: 5-15 was mild, 15-30 was moderate and >30 was severe. The results are shown in Figure 6. The observation indicators of the homogeneity of variance in the 3 groups of mild, moderate and severe OSAHS were performed by one-

Figure 7. Correlation analysis of recurrent factors in young patients with stroke.

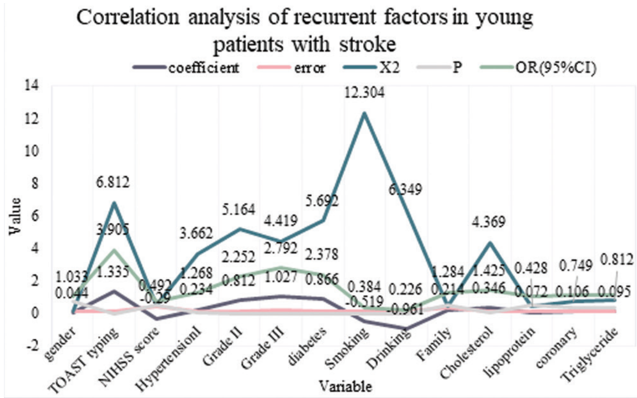
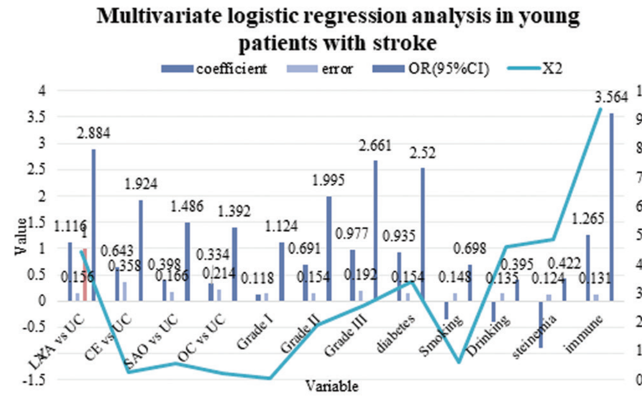


Figure 8. Multivariate logistic regression analysis of stroke in young patients.



way analysis of variance for age, BP, FBG, blood lipids, HCY and hospitalization days, and the $P > .05$ of these indicators was obtained. No statistical difference was found between the groups.

As shown in **Figure 7**, single-factor logistic regression analysis found that TOAST classification of HTN, DM, history of drinking alcohol, smoking history, hyperhomocysteinemia (HHcy), and abnormal immune indicators are associated with risk for cerebral infarction in young and middle-aged individuals. There are statistically significant differences in factors, however; NIHSS score, family history, carotid atherosclerosis, hyperuricemia, intracranial vascular stenosis and CHD were not statistically significant.

Meaningful single-factor values are included in the multi-factor logistic stepwise regression analysis, and the results are shown in **Figure 8**. Through analysis, it was found that HTN, DM, history of drinking alcohol, history of smoking, HHcy and abnormal immune indicators have an impact on the occurrence of cerebral infarction in middle-aged and young people, and the difference was statistically significant. Smoking, drinking, and HHcy had a high incidence rate in young patients with cerebral infarction, and a low incidence in middle-aged and elderly patients with cerebral infarction; the incidence of DM, HTN and abnormal immune indicators is high in middle-aged and elderly patients with cerebral infarction and low in young patients with cerebral infarction.

Our study found that the prevalence of grade I HTN in middle-aged and elderly patients with cerebral infarction was 1.114 times that in young patients with cerebral infarction, grade II HTN in middle-aged and elderly patients with cerebral infarction was 1.994 times that in young patients with cerebral infarction and grade III HTN in middle-aged and elderly patients with infarction 2.660 times that in young patients with cerebral infarction. DM prevalence in middle-aged and elderly patients with cerebral infarction was 2.521 times that in young patients with cerebral infarction; the incidence of smoking history in middle-aged and elderly patients with cerebral infarction was 0.699 that in young patients with cerebral infarctions; the incidence of drinking history in middle-aged and elderly patients with cerebral infarction was 0.394 times that in young patients with cerebral infarction; the incidence of HHcy in middle-aged and elderly patients with cerebral infarction was 0.412 times that in young patients with cerebral infarction; the incidence of abnormal immune indicators in middle-aged and elderly patients with cerebral infarction was 3.555 times that in young patients with cerebral infarction.

The aim of this study is to propose research from a different perspective. Previous studies mostly analyzed medical treatment in elderly patients with stroke, but this study is aimed at the analysis of medical treatment in young patients with stroke, and how to fully resolve the recurrence of stroke in young patients.

Study Limitations

The research framework of this study is relatively limited, and the analysis of influencing factors is not sufficiently comprehensive. Therefore, future studies must explore the factors affecting stroke in young individuals in more depth in order to improve their quality of life and explore effective self-expression in the future.

CONCLUSION

This study primarily analyzes the correlation between the persistence and recurrence of stroke in young patients based on big data in healthcare. NIHSS score, FBG, HbA_{1c}, TG, HDL, BMI, length of hospitalization, high BP, DM, heart disease, smoking and other factors evaluated in this study all affect the brain The recurrence of stroke requires more attention in the overall treatment of stroke.

CONFLICT OF INTEREST

None.

FUNDING

None.

REFERENCES

1. Chen PT. Medical big data applications: intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. *Technol Forecast Soc Change*. 2018;130:150-164. doi:10.1016/j.techfore.2018.01.033
2. Abanoz Y, Gülen Abanoz Y, Gündüz A, et al. Migraine as a risk factor for young patients with ischemic stroke: a case-control study. *Neurol Sci*. 2017;38(4):611-617. doi:10.1007/s10072-017-2810-3
3. Aigner A, Grittner U, Rolfs A, Norrving B, Siegerink B, Busch MA. Contribution of established stroke risk factors to the burden of stroke in young adults. *Stroke*. 2017;48(7):1744-1751. doi:10.1161/STROKEAHA.117.016599
4. Li W, Jiang B, Zhao W. Obstetric imaging diagnostic platform based on cloud computing technology under the background of smart medical big data and deep learning. *IEEE Access*. 2020;PP(99):1-1. doi:10.1109/ACCESS.2020.2988563
5. Godinho TM, Costa C, Oliveira JL. Intelligent generator of big data medical imaging repositories. *IET Softw*. 2017;11(3):100-104. doi:10.1049/iet-sen.2016.0191
6. Yuan Y, Chen W, Yan L, Huang B, Li J. A similarity-based disease diagnosis system for medical big data. *J Med Imaging Health Inform*. 2017;7(2):364-370. doi:10.1166/jmihi.2017.2026
7. Inomata T, Sung J, Nakamura M, et al. Using medical big data to develop personalized medicine for dry eye disease. *Cornea*. 2020;39(suppl 1):S39-S46. doi:10.1097/ICO.0000000000002500
8. Yang A, Han Y, Liu CS, et al. D-TSVR Recurrence prediction driven by medical big data in Cancer. *IEEE Transact Industr Informat*. 2020;PP(99):1-1.
9. Lee D, Park M, Chang S, Ko H. Erratum to: Protecting and utilizing health and medical big data: Policy perspectives from Korea. *Healthc Inform Res*. 2020;26(3):251. doi:10.4258/hir.2020.26.3.251
10. Na HR, Kim HS. Medical ethics in the era of artificial intelligence based on medical big data. *J Korean Diabetes*. 2020;21(3):126-129. doi:10.4093/jkd.2020.21.3.126
11. Hu ZM, Luo LL, Li L, Dai SD, Zhang HG, Liu RZ. Indigenization of the median of markers for Down syndrome screening based on statistical analysis of medical big data. *Taiwan J Obstet Gynecol*. 2020;59(4):556-564. doi:10.1016/j.tjog.2020.05.015
12. Shi M, Jiang R, Zhou W, Liu S, Sciancalepore S. A privacy risk assessment model for medical big data based on Adaptive Neuro-Fuzzy Theory. *Secur Commun Netw*. 2020;2020(9):1-18. doi:10.1155/2020/5610839
13. Gao W, Chen L, Shang T. Stream of unbalanced medical big data using convolutional neural network. *IEEE Access*. 2020;PP(99):1-1. doi:10.1109/ACCESS.2020.2991202
14. Price WN II, Cohen IG. Privacy in the age of medical big data. *Nat Med*. 2019;25(1):37-43. doi:10.1038/s41591-018-0272-7
15. Kessels RP, Eikelboom WS, Schaapsmeerders P, et al. Effect of formal education on vascular cognitive impairment after tsroke: A meta-analysis and study in young-stroke patients. *J Int Neuropsychol Soc*. 2017;23(3):223-238. doi:10.1017/S1355617716001016
16. Ekker MS, Boot EM, Singhal AB, et al. Epidemiology, aetiology, and management of ischaemic stroke in young adults. *Lancet Neurol*. 2018;17(9):790-801. doi:10.1016/S1474-4422(18)30233-3
17. Pedersen A, Stanne TM, Redfors P, et al. Fibrinogen concentrations predict long-term cognitive outcome in young ischemic stroke patients. *Res Pract Thromb Haemost*. 2018;2(2):339-346. doi:10.1002/rth2.12078
18. Reisin RC, Mazziotti J, Cejas LL, et al; AISYF Investigators. Prevalence of Fabry disease in young patients with stroke in Argentina. *J Stroke Cerebrovasc Dis*. 2018;27(3):575-582. doi:10.1016/j.jstrokecerebrovasdis.2017.09.045
19. Patira R, Smith-Benjamin S, Wang JJ. Stroke in a young patient with neurosyphilis and HIV. *Int J STD AIDS*. 2017;28(3):306-309. doi:10.1177/0956462416665029
20. Devaraddi N, Jayalakshmi G, Mutalik NR. CARASIL, a rare genetic cause of stroke in the young. *Neurol India*. 2018;66(1):232-234. doi:10.4103/0028-3886.222859
21. Nakagawa K, Ito CS, King SL. Ethnic comparison of clinical characteristics and ischemic stroke subtypes among young adult patients with stroke in Hawaii. *Stroke*. 2017;48(1):24-29. doi:10.1161/STROKEAHA.116.014618
22. Park S, Lee D, Hong S, Cho K, Lee G. Feasibility of training using full immersion virtual reality video game in young stroke survivor: A case report. *NeuroRehabilitation*. 2021;48(1):1-8. doi:10.3233/NRE-201501
23. Razzak MA, Kawnayn G, Naznin F, Rahman QAA. A case of young stroke due to Moyamoya disease. *J Armed Forces Med College Bangladesh*. 2020;15(1):110-113. doi:10.3329/jafmc.v15i1.48659
24. Thakre PI, Qureshi MI, Naqvi WM. Neuro developmental techniques with functional electrical stimulation reduces shoulder dysfunction in young stroke population: A quasi-experimental novel rehabilitative approach. *Int J Res Pharma Sci*. 2020;11(SPL4):1650-1656.
25. Razzak MA, Badsha S, Rahman Q. A case of young stroke due to Takayasu's arteritis. *J Armed Forces Med College Bangladesh*. 2020;14(2):209-213. doi:10.3329/jafmc.v14i2.45912