NOTCH3 p.R544C and the Thrombophilia Gene Role in Ischemic Stroke: A hierarchical clustering assessment of Vietnamese patients

Huong TT Bui, Quynh N.T.P., Ho Cam Tu, Sinh N.P., Thuy P.T, Thu Vu . T.H, Huyen N.T.T, Lam H.K, Dung N.T

Submitted to: JMIR Bioinformatics and Biotechnology
on: January 29, 2024

Table of Contents

NOTCH3 p.R544C and the Thrombophilia Gene Role in Ischemic Stroke: A hierarchical clustering assessment of Vietnamese patients

Huong TT Bui[1, 2*] MD, PhD; Quynh N.T.P.[3*] MD; Ho Cam Tu[4, 5*] MSci; Sinh N.P.[6] MD, PhD; Thuy P.T[2] MD, PhD; Thu Vu . T.H[4]; Huyen N.T.T[7]; Lam H.K[8] PhD, ING; Dung N.T[7]

[1]Department of Immunology Molecular Genetic Thainguyen National Hospital. Thai Nguyen VN
[2]Department of Biochemistry Thai Nguyen University of Medicine and Pharmacy Thai Nguyen VN
[3]Department of Clinical pharmacy Thai Nguyen University of Medicine and Pharmacy Thai Nguyen VN
[4]Center of Gene and Protein Research Hanoi Medical University Hanoi VN
[5]Institute of Virology, TUM School of Medicine Technical University of Munich Munich DE
[6]Department of Rehabilitation Thai Nguyen University of Medicine and Pharmacy Thai Nguyen VN
[7]Department of Internal Medicine Thai Nguyen University of Medicine and Pharmacy Thai Nguyen VN
[8]Department of Telecomunication Hung Yen University of Technology and Education Hung Yen VN
[*]these authors contributed equally

Corresponding Author:
Dung N.T
Department of Internal Medicine
Thai Nguyen University of Medicine and Pharmacy
284 Luong ngoc quyen, Quang Trung
Thai Nguyen
VN

Abstract

Background: The etiology of ischemic stroke is multifactorial. Several gene mutations have been identified as a leading cause of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). This hereditary disease causes stroke and other neurological symptoms.

Objective: The overall preclinical characteristics, cumulative cutpoint values, and the factors associated with these somatic mutations were analyzed in the uni/multidimensional scaling model.

Methods: We conducted a hierarchical cluster analysis (HCA) study on 100 patients diagnosed with ischemic stroke, identifying the variants of NOTCH3 and the thrombophilia genes by PCT-CTPP and real-time PCR.

Results: We found several critical optimal cutpoints are: 83.67±9.19umol/l(Creatinin), 54±5years old (age), 13.25±0.17 seconds (time of PT), 1.02±0.03 (INR), in which we end up with the significant cutpoint 50 of Glasgow Coma Scale (GCS), the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge, are: 12.77; 2.86±1.21, 9.83±2.85, 7.29±2.04, 6.85±2.90, respectively by Nagelkerke method. The two variants of MTHFR (C677T, A1289C) and NOTCH 3 p.R544C may influence the stroke severity with specific conditions of Prothrombin, Creatinin, INR, and BMI, with the risk ratios of 4.8[1.53,15.04] and 3.13[1.6,6.11], respectively (pfisher<0.05).

Conclusions: It is also interesting that although there are lots of genes linked to increased atrial fibrillation risk, not all of them are associated with the ischemic stroke risk. As stroke risk loci are being detected, more information is gained on their impact and their interconnections, especially in young patients.

(JMIR Preprints 29/01/2024:56884)
DOI: https://doi.org/10.2196/preprints.56884

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

NOTCH3 p.R544C and the Thrombophilia Gene Role in Ischemic Stroke: A hierarchical clustering assessment of Vietnamese patients

B.T. Thu Huong[1,2*], N. T. Phuong Quynh[4*], H. Cam Tu[6,7*], N.P. Sinh[3], P.T. Thuy[1], , V. Thi Hoai Thu[6] , N. Thi Thu Huyen[5], H. Khanh Lam[8], N. Tien Dung[5]

[1] Department of Biochemistry, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam;

[2] Department of Immunology Molecular Genetic Thainguyen National Hospital.

[3] Department of Rehabilitation , Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam

[4] Department of Clinical pharmacy, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam

[5] Department of Internal Medicine, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam

[6] Center of Gene and Protein Research, Hanoi Medical University, Hanoi, Vietnam

[7] Institute of Virology, TUM School of Medicine, Technical University of Munich, 81675 Munich, Germany

[8] Department of Telecomunication, Hung Yen University of Technology and Education, Hung Yen, Vietnam

***these authors contributed equally**

Correspondant:

N. Tien Dung, Department of Internal Medicine, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam Email: dung.nt@tnmc.edu.vn, Phone: 0084-913516863

Funding

This research was funded by Ministry of Education and Training, Vietnam (Grant No. 3813/QĐ-BGDĐT, November 20th 2020).

Ethics approval and consent to participate

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics committee of Thai Nguyen National Hospital (Reference No. #59/HĐĐĐ-

BVTWTN# January 18th 2021). Written informed consent was obtained from the subjects regarding the use of the samples and information for research purpose.

Abstract

Background: The etiology of ischemic stroke is multifactorial. Several gene mutations have been identified as a leading cause of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), a hereditary disease that causes stroke and other neurological symptoms. **Objectives:** We want to identify the variants of NOTCH3 and the thrombophilia genes and their complex interaction with other factors. **Methods:** We conducted a hierarchical cluster analysis (HCA) study on 100 patients diagnosed with ischemic stroke, identifying the variants of NOTCH3 and the thrombophilia genes by PCT-CTPP and real-time PCR. The overall preclinical characteristics, cumulative cutpoint values, and the factors associated with these somatic mutations were analyzed in the uni/multidimensional scaling model. **Results:** We found several critical optimal cutpoints are: 83.67±9.19µmol/l(Creatinin), 54±5years old (age), 13.25±0.17 seconds (time of PT), 1.02±0.03 (INR), in which we end up with the significant cutpoint 50 of Glasgow Coma Scale (GCS), the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge, are: 12.77; 2.86±1.21, 9.83±2.85, 7.29±2.04, 6.85±2.90, respectively by Nagelkerke method. **Conclusion:** The two variants of *MTHFR* (*C677T, A1289C*) and *NOTCH 3 p.R544C* may influence the stroke severity with specific conditions of Prothrombin, Creatinin, INR, and BMI, with the risk ratios of 4.8[1.53,15.04] and 3.13[1.6,6.11], respectively (p_{fisher}<.05). It is also interesting that although there are lots of genes linked to increased atrial fibrillation risk, not all of them are associated with the ischemic stroke risk. As stroke risk loci are being detected, more information is gained on their impact and their interconnections, especially in young patients.

Introduction

Stroke is a medical condition that is the disruption of blood flow, leading to brain cell death. There are several hazard factors for stroke, including high blood tension, smoking, diabetes, and increased cholesterol levels. In 2019, the Global Burden of Disease analysis assessed that there were 12.2 million incident cases of stroke and 101 million prevalent cases

of stroke, with 6.55 million deaths. The burden of stroke is highest in inferior and middle-income countries, where danger factors such as high blood pressure, smoking, and insufficient diet are more prevalent.[1]

The estimated population of Vietnam in 2021 was 98.32 million, of which the young population accounts for the majority. The proportion of people older than 65 years accounts for 7.7% of the total population. This phenomenon is the leading cause of death and disability in Vietnam. The incidence and prevalence of stroke were reportedly 161 and 415 per 100,000 people, respectively.[2] Stroke is classified broadly into three types: ischemic stroke, hemorrhagic stroke, and subarachnoid hemorrhage. Ischemic stroke happens due to blockage of blood vessels, which limits the blood pool to the brain. Approximately 60–80% of all strokes are ischemic. This article focused on acute ischemic strokes and their genetic features. Their Unmodifiable Risk Factors Include Age, Race, Sex, Ethnicity, History of migraine headaches, and Fibromuscular dysplasia. The Hereditary factors are Family history of stroke or transient ischemic attacks (TIAs). The Modifiable Risk Factors Include Hypertension, Diabetes mellitus, Cardiac disease, High cholesterol, Previous stroke, Carotid stenosis, Hyperhomocysteinemia, and Lifestyle issues. The majority of the ischemic strokes seen in patients with cardiovascular disease are embolic. [3]

The etiology of ischemic stroke is multifactorial. Although obtaining minor focus, genetic causalities assemble a considerable contribution to ischemic stroke genesis, particularly in early-onset stroke. Several stroke classification systems were proposed based on genetic information corresponding to various stroke phenotypes. Twin and family history studies and the candidate gene approach are standard methods to discover genetic causes of stroke. However, both have their limitations. Some monogenic disorders (7% of stroke etiology) may generate well-known clinical indications that include stroke. Polygenic disorders are more frequent, causing 38% of ischemic strokes, and their designation is a rapidly evolving field of current stroke genetics. Recent advances in human genetics provide opportunities for personalized stroke prevention and unknown cure options. Some authors boost the application of stroke gene panels for stroke hazard evaluation and stroke research. Moreover, unknown biomarkers for stroke hereditary causes and novel marks for gene therapy are on the horizon.[4]

Machine learning–based models performed better in predicting poststroke outcomes

than regression models using the items of conventional stroke prognostic scores, although they required additional variables, such as laboratory data, to attain improved performance. Further studies are warranted to validate the usefulness of machine learning in clinical settings. [5]

Following our previous HCA research,[6] we studied the overall preclinical characteristics, cumulative cut-point values, and the factors associated with the Thrombophilia genes and *NOTCH3* p.R544C variant test in uni/multi-dimensional analyses from ischemic stroke patients in Vietnam.

Materials and Methods

Study design

We use the convenience sampling choice to collect 100 patients with cerebral infarction (ischemic stroke) (diagnosed as acute ischemic stroke according to the clinical standards of the World Health Organization and the results of diagnostic imaging CT and/or MRI and/or CTA) have been or are being treated at the Stroke Center, Thai Nguyen Central Hospital, are residents of the Northern mountainous provinces, were ≤ 60 years old at the time of first stroke, willing to participate in the research. Exclusion criteria: Patients with cerebral venous sinus thrombosis, intracranial hemorrhage, and subarachnoid hemorrhage. The collected information includes risk factors for stroke in the medical history of patients, such as hypertension, diabetes, coronary artery disease, history of stroke, atrial fibrillation, smoking, headache, hyperlipidemia, valve replacement, thyroid dysfunction, history of abortion, vascular disease, blood disorders, chronic alcohol consumption, and use of oral contraceptives. The patients should undergo routine biochemical and hematological tests, Doppler ultrasound of the carotid and vertebral arteries, magnetic resonance imaging or computed tomography angiography of the brain, coagulation tests, fibrinogen, and homocysteine. Based on the previous study[2], we suppose in 100 ischemic patients with confidence level 95%, the margin of error is within ±7.84% of the polulation size (stroke in general) with 80% are ischemic ones (Formular 1). A sample size with sufficient statistical power is critical to the success of genetic association studies to detect causal genes of human complex diseases, especially in the cases of ischemic stroke. We choose two-tailed test with Type I error is 0.05, because we want to see if the average continous level (preclinical

factors) of patients from different cutpoints. In clinical and biological studies, the effect size d following Cohen's criteria, the degree of difference between two or more groups, are important. Cohen's d is ratio of Δ and σ where σ is the standard deviation; Δ is an influence index of the risk factors (treatment, genotype, ...) on the population phenotype. In our study, we calculated the Cohen's d following the suppose sample size from 50 to 100. With a power of 80%, using a two-sided T test, we estimated d could be from 0.4 (sample size each group is 99) to 0.7 (sample size each group is 45) (Formular 2) That means screening all risk factors may have a medium or higher level of influence on the phenotype (p<.05 is statistical significance) (Table 1)

Genetic testing

Analysis of polymorphisms of *NOTCH3* p.R544C and *FV-H1299R, MTHFR-C677T, MTHFR-A1289C, FII-Prothrombin, FV-Cambridge, PAI1 4G5G, FXIII Val34Leu* is done by PCR-CTPP and the Thrombophilia genetic assay. The peripheral blood of study participants was collected using a standard blood collection procedure and collected in EDTA-containing tubes. Whole-genome DNA was extracted from 2-3 ml of peripheral venous blood using EDTA-containing tubes. The QIAamp DNA mini blood kit (Qiagen, Hilden, Germany) was used for DNA extraction. The quality of the total DNA was checked by electrophoresis on agarose gel and by measuring the absorbance at 260/280 nm wavelengths, then stored at -80°C until use. The *NOTCH3* mutation p.R544C was identified by PCR-CTPP (polymerase chain reaction with confronting two-pair primers). DNA was amplified with the primers 5'GTGGGGTGGAGTGGAAGTAAGTGG (F1) and 5'GAGCAGTCGTCCACGTTGCA (R1) for the C allele; 5'TTGAGGGCACGCTGTGTGATC (F2) and 5'CTAGATGCACCATTCCCAAACCC (R2) for the T allele. The PCR amplification was performed for 40 cycles (denaturation at 95°C for 30 sec, annealing at 62°C for 30 sec, extension at 72°C for 1 minute, and final extension at 72°C for 10 min). PCR products of 479 and 216 bp for the TT genotype, 479, 303, and 216 bp for the TC genotype, and 479 and 303 bp for the CC genotype were shown on 2% agarose gel stained with ethidium bromide. Once the sequence variants were identified, additional steps were taken to confirm the sequence changes of the amplicons. Real-time PCR of SNP biotechnology ®, used for detect *FV-H1299R, MTHFR-C677T, MTHFR-A1289C, FII-Prothrombin, FV-Cambridge, PAI1 4G5G,*

FXIII Val34Leu.

Ethic statement

The study should obtain informed consent from all participants or their legal representatives, ensuring they understand the study's purpose, risks, benefits, and procedures.

Statistical and HCA

The conventional statistical analyses were performed on our data set, including the medical test parameters in 100 stroke ischemic patients, using IBM SPSS Statistics 20. The relationship between clinicopathological factors and presence of *NOTCH3* p.R544C, *FV-H1299R, MTHFR-C677T, MTHFR-A1289C, FII-Prothrombin, FV-Cambridge, PAI1 4G5G, FXIII Val34Leu* mutations were analyzed using Pearson chi-squared test (group size of > 5) or Fisher exact test (group size of ≤ 5), as appropriate. Bonferroni's correction for multiple comparisons was applied. The results were expressed as a percentage or mean ± SD.

Following our previous machine learning study, our multidimensional analysis (MDA) was conducted in R 4.1.0 (R Project for Statistical Computing).[6] We focused on multivariate statistics, using several algorithms of HCA, matrix correlation, Nagelkelker, Kaplan-Meier, and Log Rank Test. The chi-square statistics were computed using Yates's correction for continuity, generating p_{yates}. The Pearson or product-moment correlation coefficient is frequently used as the outcome measure for analyses. Pearson’s method has the advantage when all or most of the nonzero parameters share the same sign. Pearson’s test has proved useful in a genomic setting, screening for age-related genes, which is also our objective. [7] Two alternative criteria are a bias-corrected version of the correlation coefficient (giving p_{uncor}) and Fisher's r-to-z transformed correlation coefficient (giving p_{fisher}). HCA is a cluster analysis concept that creates a dendrogram hierarchy of clusters. The HCPC (Hierarchical Clustering on Principal Components) approach allows us to combine the three standard methods used in multivariate data analyses: Principal component methods (PCA, CA, MCA, FAMD, MFA), Hierarchical clustering and Partitioning clustering, particularly the k-means method. We need calculate the distance between each observation and estimates the cluster's distance to the remaining statements. The distance between the elements can be

complete, single, average, ward, mcquitty, or centroid. The cluster tree is generated by computing the correlation between cophenetic distances and initial distance data. The number of clusters is determined using k-means, which calculates clustering indexes and reallocates observations to the closest cluster. The K-means computation was optimized using 20 indexes to conduct the principal component analysis (PCA) cluster plot, which visualizes the best cluster number. Principal component analysis, or PCA, is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Results

Overview of correlation between the clinic-pathological factors and the mutation prevalence of NOTCH3 p.R544C, *FV-H1299R, MTHFR-C677T, MTHFR-A1289C, FII-Prothrombin, FV-Cambridge, PAI1 4G5G, FXIII Val34Leu in ischemic patients*

The results of a study on 100 patients with cerebral infarction in the northern mountainous region of Vietnam, including 75 Kinh patients (accounting for 75%), 25 patients of the Tay ethnic group (accounting for 25%), showed the average age of study patients is about 60.1 in which: the lowest patient is 24, the highest is 91. (Table 2) The age group from 24 to 49 in 22 patients (accounting for 22%), from 50 to 59 in 23 patients (accounting for 23%), from 60 to 69 were 37 patients (accounting for 37%), from 70 to 91 were 18 (accounting for 18%). (Figure 1).

Male patients accounted for the majority with 62 patients (62%), female patients had 38 patients (accounting for 38), male/female ratio was 1.63. The average BMI of the study group was 22.62, of which three patients had a BMI < 18.5 (accounting for 3%), 56 patients had a BMI in the range of 18.5-22.9 (accounting for 56%), 27 patients with BMI between 23-24.9 (accounting for 27%), 14 patients with BMI in the range of 25-29.9 (accounting for 14%). Male patients accounted for the majority with 62 patients (62%), female patients had

38 patients (accounting for 38), male/female ratio was 1.63. Among the risk factors for stroke, hypertension accounts for the highest proportion of 70% of patients. There are 44% of patients with a family history of stroke. There were 31 patients with a history of smoking (accounting for 31%), a history of alcohol consumption in 29 patients (accounting for 29%), and a history of diabetes in 20 patients (accounting for 20%). Of the 100 studied patients, 35 had a stroke history (accounting for 35%). (Table 2)

Research results show that among the clinical symptoms, motor paralysis, difficulty speaking, and mouth distortion are the most common, with respective rates of 97%, 95%, and 72%. Followed by symptoms of headache accounted for 49%, numbness accounted for 41%, dizziness/vertigo accounted for 27%, round muscle disorder accounted for 21%, and nausea/ vomiting accounted for 8%. For motor paralysis, there were 52% of patients with right hemiplegia, 39% of patients with left hemiplegia, and 6% of total paralysis patients out of 100 patients. For symptoms of dysphasia, mainly Broka-type dysphonic patients (86%), only a tiny number of Wernik-type dyspraxia patients (9%) out of 100 patients. (Table 3)

The study results showed that the average time from the onset of the first symptoms to the patient's admission was 10.94 hours, of which 33 patients were admitted within the first 4.5 hours (accounting for 33%), 26 patients were coming to the hospital within the first 4.5 hours (accounting for 33%), and 26 patients were hospitalized. From 4.6h to 6h (accounting for 26%), the remaining 41 patients were admitted outside the first 6 hours (accounting for 41%). The patient's blood pressure at entry: mean systolic blood pressure was about 148.6 mmHg, and mean diastolic blood pressure was approximately 88.06 mmHg. The average Glasgow score when the patient was admitted to the hospital was 14.72. The average NIHSS score when the patient was admitted to the hospital was 7.14, and after 24 hours of hospital treatment was 6.71 and 3.73 at the time of discharge. The average Rankin score when the patient is discharged from the hospital is about 1.52. The average number of days of treatment was 10.11 days. (Table 4) PCR-CTPP identified the *NOTCH3 p.R544C* and other gene variants were detected by RT-PCR. (Table 1, Figure 2). Result of real-time PCR of SNP biotechnology ®, used for detect *FV-H1299R, MTHFR-C677T, MTHFR-A1289C, FII-Prothrombin, FV-Cambridge, PAI1 4G5G, FXIII Val34Leu* were represented in Figures 3, 4,

5, 6 and Table 2.

Figures 7, Table 1, and Table 5 give the overall view of gene prevalence and some correlations in both negative and positive genes. We confirm the significant correlation of *NOTCH3* p.R544C, *FV H1299R, MTHFR-C677T, MTHFR A1289C, FII Prothrombin, FV-Cambridge, PAI1 4G5G, FXIII Val34Leu* with several factors in ischemic stroke patients. The Pearson correlation coefficient (R) calculates the force and tendency of the relationship between two variables. The relationship strength (effect size) varies between disciplines following the threshold of R, which are 0.5, 0.3, 0, -0.3, and -0.5 for strong, moderately positive, or negative correlation, respectively. (Figure 7; Interactive HTML graph 1, 2, 3, 4; Table 5) The volcano graph shows the most significant correlation pair, especially those containing these gene mutations above. (Figure 8, Interactive HTML graph 5). Overall, the significant medium correlation between the gene mutation prevalence and other factors was shown in the Volcano graph; compared with other genes, *FXIIIVal34Leu* shows the highest positive correlation with the thrombus suction ability (R=0.54, p<.0001,-$\log_{10}p$=8.03).

In clustering step, the dendrograms were built based on the clustering metric “Euclidean,” we selected “average” as the most appropriate linkage model, which had the best correlation between cophenetic distances and the original distance data (Table 6).

We selected the results proposed by Beale methods from 20 different index values, presented with 15 clusters as optimal (Table 7). The PCA cluster plots showed that the cluster number mentioned above was the best number to distinguish the clusters and avoid overlap appropriately. Dendrogram and PCA map complete the overall view of our databases, in which we can see where studied genes could combine and may be necessary for the ischemic stroke outcomes. (Figure 9A, 9B and Interactive Graph 6). We found several clusters of variants that may have a synchronization impact on the outcome of ischemic stroke. This PCA map gives us the first idea of the potential markers that may be important for ischemic score, such as the level of international normalized ratio (INR) and photothrombotic time (PT Second) are in the same cluster with NIHSS and Rankin (Cluster number 9 in Figure 9B, clusters 3 and 14 in Interactive Graph 6) ; the Glasgow locates in the same cluster with the photothrombotic ratio (PT ratio) (Cluster number 12 in Figure 9B, cluster 15 in Interactive Graph 6). We found the studied genes are separated in four different groups: *FII*

prothrombin, *MTHFR C677T*, *Notch 3* are in cluster number 4 (Figure 9B, cluster 4 in Interactive Graph 6); *FV Leiden* and *PAI1 4G 5G* in cluster number 6 (Figure 9B, cluster 7 in Interactive Graph 6); *FV H1299R* and *MTHFR A1298C* in cluster number 11 (Figure 9B, cluster 1 in Interactive Graph 6); *FXIIIVal34Leu* locates in cluster number 13 (Figure 9B, cluster 2 in Interactive Graph 6). We continue to split the data following the significant cutpoint of photothrombotic, international normalized ratio, and the ischemic stroke score. We apply the maximally selected rank statistic to define the optimal threshold of several continuous factors (Creatinin, age, PT time and ratio, INR, LDL C, Number of infarcts via CT or MRI, patient height, MPV) based on the Rankin, NIHSS, and Glasgow score and their related symptom status such as numbness, dizzy, gender, circular muscle disorder, mouth distorted and diabetes status. (Table 8) We found their optimal cutpoint is 83.67±9.19µmol/l(Creatinin), 54±5years old (age), 13.25±0.17 seconds (time of PT), 1.02±0.03 (INR), 4.23±0.89mmol/l(LDL-C), 2 (Number of infarcts via CT or MRI), 99.00±1.96 (ratio of PT), 7.27±1.09fL (MPV), respectively. (Table 8F) Using the Nagelkerke method, we verify which value threshold above could be the significant cut point 50 of these ischemic stroke scores. We end up with only 6/9 factor above (Creatinin, age and height, time and ratio of PT, number of infarcts vis CT) may give us the significant cutpoint 50 of the Glasgow Coma Scale (GCS), the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge, are: 12.77; 2.86±1.21; 9.83±2.85; 7.29±2.04; 6.85±2.90; respectively. With these new ranks of scores, in which we could study the possible influence, including the genotype variant. (Figures 10,11,12,13,14,15,16,17)

Gene variants may be associated with the patient outcome via the ischemic stroke score.

We calculate the risk ratio and the confidence intervals by unconditional maximum likelihood estimation and normal approximation, respectively (Wald), as well as minor sample adjustment by the Mantel Haenszel method generating p_{yates}, p_{uncor} and p_{fisher}. We group these genotype variants following their cluster which give the most relevant risk ratio results. (Table 9, Figure 18 and Interactive Graph 7). The risk ratio in detail for stroke scores were showed in the Table S1. Forest plots represent their RR score of the risk factors combining with remarkable clusters 4 (Figures S1, S2, S3, S4, S5), 6 (Figures S6, S7, S8, S9), 11 (Figures S10. S11, S12, S13), 13 (Figures S14, S15, S16, S17).

The Glasgow Coma Scale (GCS) is connected to a head injury, and they tend to apply scoring ranges to describe the injury severity. The ranges are 13 to 15 for Mild traumatic brain injury (mTBI), 9 to 12 for the moderate TBI, and 3 to 8 is for Severe TBI. The group of non-diabetes patients whose BMI is superior 20.8, with *Notch 3* heterozygous, *MTHFR C677T* and *FI prothrombin* may have the possibility of getting mild traumatic brain injury (mTBI) (Cutpoint 50 of GCS is 12.77) higher than other groups 23% (RR=1.23[0.99,1.54], p_{fisher}<0.01). This ability was decreased significantly in 20% (RR=0.79 [0.61,1.01], *p*fisher<0.01) when BMI is less than 20.8, and *MTHFR A1298C* and *FVH1299R* are wildtype variants.

NIH Stroke Scale (NIHSS) quantifies the impairment caused by a stroke and aids in planning post-acute care disposition, though it was intended to assess differences in interventions in clinical trials. NIHSS 0 is for no stroke symptoms, 1–4 is for minor stroke, 5–15 is for moderate stroke, 16–20 is for moderate to severe stroke, and 21–42 is for powerful stroke. In the NIHSS at admission (NIHSS0h) higher than 9.83 and the NIHSS at 24h (NIHSS24h) higher than 7.92 (moderate stroke), the potential group is the patients who are older than 54 years old, smaller than 161cm, have PT time ≤13.25 senconds; PT ratio ≤99; Creatinin >83.67µmol/l; *FXIII Val34Leu* wildtype (RR=2.72[1.4,5.31] and 2.09[1.1,3.93], respectively, p_{fisher}<.05). For the NIHSS on discharge (NIHSSend) higher than 6.85 (moderate stroke), the most significant risk ratio (RR=4.8[1.53,15.04], p_{fisher}<.05) is in the group of patients who are older than 54 years old, taller than 161cm, have PT time ≤13.25 seconds; PT ratio ≤99; Creatinin >83.67µmol/l; *FII prothrombin* are *MTHFR C677T wildtype* and *Nochth3 p.R544C* is heterozygous.

The Modified Rankin Score (mRS) is an outcome measure in stroke clinical trials. A mRS score is recommended three months (90 days) following hospital discharge. The ranges are 0 for The patient who has no residual symptoms; 1 for the patient who has no significant disability and can carry out all pre-stroke activities; 2 for the patient who has remote disability incapable of carrying out all pre-stroke movements but capable of looking after self without daily help; 3 for the patient has a moderate disability; needing some

external help but capable of walking without the assistance of another individual; 4 for the patient has a moderately severe disability; incapable to walk or follow to physical functions without the aid of another individual; 5 for the patient has a severe disability; bedridden, incontinent, requires continuous care; 6 for the patient has passed away (during the hospital stay or after discharge from the hospital); **7** for incompetent to contact patient/caregiver; 8 for Modified Rankin Score not achieved, OR unable to determine (UTD) from the medical record documentation. At the Rankin higher than 2.86 (moderate disability), the patients have INR higher than 1.02, PT time >13.25 seconds, PT ratio ≤99; Creatinin >83.67µmol/l; *FXIIIVal34Leu wildtype* (in case, number of infarcts via CT is superior to 2) or *MTHFR A1298C heterozygous/wildtype and FVH1299R wildtype*. (RR=3.13[1.6,6.11], p_{fisher}<.05)

Discussion

Some sophisticated techniques for HCA exploit a statistical framework called hierarchical models, or multilevel models. Hierarchical models rather than simpler methods which are useful in a number of contexts. Hierarchical cluster analysis (HCA), also known as hierarchical clustering, is a popular method for cluster analysis in big data research and data mining aiming to establish a hierarchy of clusters As such, HCA attempts to group subjects with similar features into clusters. Clustering is a data science technique in machine learning that groups similar rows in a data set. After running a clustering technique, a new column appears in the data set to indicate the group each row of data fits into best.

Several gene mutations have been identified as a leading cause of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), a hereditary disease that causes stroke and other neurological symptoms. CADASIL accounts for up to 5 percent of all strokes in individuals under the age of 65. The thrombophilia test helps determine the disease's genetic origin to provide appropriate prevention and treatment measures. Hypercoagulation syndrome may be Due to mutations in genes encoding proteins related to blood clotting, meaning Thrombophilia. People with hypercoagulable syndrome tend to form blood clots in blood vessels (primarily veins), thereby causing: Stroke, Heart attack, Repeated miscarriages, Complications during pregnancy: preeclampsia, fetal growth

retardation, and stillbirth.[8]

In our study, eight genes were applied to understand how ischemic stroke genetics could interest the practitioner and be useful for clinical work. They are: *FII Prothrombin; FV Leiden; MTHFR C677T; MTHFR A1298C; FV H1299R; PAI-1 4G/5G; FXIII Val34Leu; FV Cambridge; and Notch 3 p.R544C*.

We visualize how these risk factors and genetic elements could affect ischemic stroke outcomes with the hierarchical analysis strategy. The maximally selected rank statistic helps to define the optimal threshold of several continuous factors (Creatinin, age, PT time and ratio, INR, LDL C, Number of infarcts via CT or MRI, patient height, MPV) based on the Rankin, NIHSS, and Glasgow score and their related symptom status such as numbness, dizzy, gender, circular muscle disorder, mouth distorted and diabetes status. (Table 5) Their optimal cutpoint is fitted with the normal range in both genders. Creatinin at 83.67±9.19µmol/l is among a usual result of 0.7 to 1.3 mg/dL (61.9 to 114.9 µmol/L) for men and 0.6 to 1.1 mg/dL (53 to 97.2 µmol/L) for women. [9] Our age threshold is 54±5years old, which is similar to others worldwide, in which aging is the most robust non-modifiable risk factor for incident stroke, which doubles every ten years after age 55 years. [10] Average prothrombin time (PT second) is recommended for administration of recombinant tissue-plasminogen activator (rt-PA) in stroke. [11] The standard range for PT is 10 to 13 seconds. The usual INR for a healthy individual is 1.1 or below, and the therapeutic range for most patients on Monitoring vitamin K-antagonists (VKAs) is an INR of 2.0 to 3.0. An augmented PT/INR for patients on VKAs may suggest a super-therapeutic status and will need prescription dose adjustments to control bleeding. [12] In our study, the calculated baseline of PT time is 13.25±0.17 seconds (time of PT), 1.02±0.03 (INR), which confirms the moderate outcome cases. Several data on the association between body mass index (BMI) and stroke are scarce. When BMI is between 18.5 and 24.9 – it is in the healthy weight range. Our calculated BMI baseline is 20.85, and it is still associated with genetic factors that influence the GCS. (Table 6F)

The following significant cutpoint 50 by Nagelkerke method of Rankin, NIHSS at admission, after 24h and in the discharged time, are 2.86±1.21, 9.83±2.85, 7.29±2.04, 6.85±2.90, respectively, which suitable with the moderate outcomes of our patients. We found the two variants of *MTHFR* and *NOTCH 3 p.R544C* may influence the most stroke

severity with specific conditions of prothrombin, Creatinin, INR, and BMI.

The *MTHFR* gene provides instructions for the human body to make the *MTHFR* protein, which helps the body process folate, which is important for forming DNA and modifying proteins. The most common variant in the *MTHFR* gene is *MTHFR C677T*. [13]

This mutation will cause a reduction of the capacity to create L-methylfolate. *MTHFR A1298C* SNP has also been suggested to have an impact on *MTHFR* enzyme activity but to a lesser extent than the *MTHFR C677T* polymorphism. They have been recently associated with ischemic stroke. [14]

CADASIL is an autosomal dominant inherited vasculopathy and the most common single-gene disorder causing stroke—more than 200 different *NOTCH3 p.R544C* mutations in patients worldwide, indicating that CADASIL has considerable genetic heterogeneity. The defective 33-exon *NOTCH3 p.R544C* gene is located on chromosome 19, which typically impacts the number of highly conserved cysteine residues among the epidermal growth factor-like repeat domain. [15]

HCA is attractive in exploratory high-throughput data because HCA provides a convenient tool to visualize the similarities of variables and infer the grouping of variables based on the dendrogram structure; hence, HCA facilitates the interpretation of microbiome and other omics data. More critical, biclustering (two-way clustering), a particular case of HCA, can incorporate a correlation method (e.g., Spearman's rank correlation) to cluster rows and columns of the data matrix simultaneously. Thus, biclustering can find features (microbial taxa, genes, metabolites, etc.) that correlate only in a subset of objects but not in the rest of the dataset[16] In this study, we see more clearly the role and interaction of risk factors that influence stroke progression. Genetic mutations become significant in a small range of strongly correlated factors through a PCA chart.

Stroke has multiple modifiable and nonmodifiable risk factors and represents a leading cause of death globally. Understanding the complex interplay of stroke risk factors is thus not only a scientific necessity but a critical step toward improving global health outcomes. [17]

Conclusion

The existence of conventional vascular risk factors may prevent clinicians from suspecting the possibility of gene mutations in stroke patients, especially for those with underlying atrial fibrillation or extensive artery atherosclerosis. As can be seen, a more specific population was chosen in this study. It is also interesting that although there are lots of genes linked to increased atrial fibrillation risk, not all of them are associated with ischemic stroke risk, which might be because those gene variants are too rare to detect their impact on stroke risk so far. Nevertheless, if we find in the future that some of those genes are linked to ischemic stroke, it could be a significant game-changer in the field of stroke prevention. Moreover, as stroke risk loci are being detected and more information is gained on their impact and interconnections, the precision of these scores increases.

Limitations

We found three of nine gene varients have a significant risk ratio. Data settings could help us work with both qualitative and numerical data simultaneously. The main advantage of the HCA clustering concept was showing possible correlations between several factors to provide the reference markers useful for diagnostic control and to improve outcome prevention. Especially finding the association between genetic character and clinical outcomes, which needs several in-vitro studies; however, it had some constraints. It was critical to clean and prepare the data set because HCA and k-means cannot operate with missing or noisy data. We must combine and validate the data with k-means, which provides several options for the optimal cluster number to produce a PCA cluster plot and define the PC position. Since our data had various kinds of statements, it was challenging to calculate the distance matrix in HCA and k-means.

Acknowledgements

We sincerely thank the patients and the participants for consenting and permitting us to publish the data. We thank our colleagues at National Hospital of Thai Nguyen, who support us with patient recruitment and clinical data supply.

Consent to publish

Written informed consent was obtained from all subjects.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. We however cannot provide personal information or data contain identification of the patients in any form.

Conflict of interests

The authors declare that they have no conflict of interests.

Author contributions

BTTH and HCT designed the present study. BTTH received the grant for the study. NPS, PTT, NTPQ, NTTH, VTHT performed the data collection, the experiments. CTH and HKL performed the data mining and HCA study. BTTH and CTN wrote the main manuscript. NTD revised the manuscript and supervise. All authors read and approved the final manuscript.

Abbreviations

LDL-C: Low Density Lipoprotein ***Cholesterol***
HDL-C: High Density Lipoprotein Cholesterol
ALT: Alanine Aminotransferase
AST: Aspartate Transaminase
RBC: ***Red Blood Cell***
WBC: ***White Blood Cell***
NE: Neutrophil
Hb: Hemoglobin
HCT: Hematocrit
PLT: Platelet Count
MPV: Mean Platelet Volume
EF: Ejection Fraction

Reference

[1] V. L. Feigin *et al.*, “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019,” *Lancet Neurol.*, vol. 20, no. 10, pp. 795–820, Oct. 2021, doi: 10.1016/S1474-4422(21)00252-0.

[2] D. T. Mai, X. C. Dao, N. K. Luong, T. K. Nguyen, H. T. Nguyen, and T. N. Nguyen, “Current State of Stroke Care in Vietnam,” *Stroke Vasc. Interv. Neurol.*, vol. 2, no. 2, p. e000331, Mar. 2022, doi: 10.1161/SVIN.121.000331.

[3] C. Chugh, “Acute Ischemic Stroke: Management Approach.,” *Indian J. Crit. Care Med. Peer-Rev. Off. Publ. Indian Soc. Crit. Care Med.*, vol. 23, no. Suppl 2, pp. S140–S146, Jun. 2019, doi: 10.5005/jp-journals-10071-23192.
[4] A. Ekkert, A. Šliachtenko, J. Grigaitė, B. Burnytė, A. Utkus, and D. Jatužis, “Ischemic Stroke Genetics: What Is New and How to Apply It in Clinical Practice?,” *Genes*, vol. 13, no. 1, Dec. 2021, doi: 10.3390/genes13010048.
[5] F. Irie *et al.*, “Predictive Performance of Machine Learning–Based Models for Poststroke Clinical Outcomes in Comparison With Conventional Prognostic Scores: Multicenter, Hospital-Based Observational Study,” *JMIR AI*, vol. 3, p. e46840, Jan. 2024, doi: 10.2196/46840.
[6] T. T. Nguyen, C. T. Ho, H. T. T. Bui, L. K. Ho, and V. T. Ta, “Multidimensional Machine Learning for Assessing Parameters Associated With COVID-19 in Vietnam: Validation Study,” *JMIR Form Res*, vol. 7, p. e42895, Feb. 2023, doi: 10.2196/42895.
[7] Art B. Owen, “Karl Pearson’s meta-analysis revisited,” *Ann. Stat.*, vol. 37, no. 6B, pp. 3867–3892, Dec. 2009, doi: 10.1214/09-AOS697.
[8] E. J. Favaloro, “Genetic Testing for Thrombophilia-Related Genes: Observations of Testing Patterns for Factor V Leiden (G1691A) and Prothrombin Gene ‘Mutation’ (G20210A).,” *Semin. Thromb. Hemost.*, vol. 45, no. 7, pp. 730–742, Oct. 2019, doi: 10.1055/s-0039-1694772.
[9] A. O. Hosten, “BUN and Creatinine.,” in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., Boston: Butterworths, 1990.
[10] M. Yousufuddin and N. Young, “Aging and ischemic stroke.,” *Aging*, vol. 11, no. 9, pp. 2542–2544, May 2019, doi: 10.18632/aging.101931.
[11] R. F. Gottesman, J. Alt, R. J. Wityk, and R. H. Llinas, “Predicting abnormal coagulation in ischemic stroke: reducing delay in rt-PA use.,” *Neurology*, vol. 67, no. 9, pp. 1665–1667, Nov. 2006, doi: 10.1212/01.wnl.0000244493.13898.5b.
[12] W. E. Winter, S. D. Flax, and N. S. Harris, “Coagulation Testing in the Core Laboratory.,” *Lab. Med.*, vol. 48, no. 4, pp. 295–313, Nov. 2017, doi: 10.1093/labmed/lmx050.
[13] K. S. Crider, T. P. Yang, R. J. Berry, and L. B. Bailey, “Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate’s role.,” *Adv. Nutr. Bethesda Md*, vol. 3, no. 1, pp. 21–38, Jan. 2012, doi: 10.3945/an.111.000992.
[14] D. S. Chita *et al.*, “MTHFR Gene Polymorphisms Prevalence and Cardiovascular Risk Factors Involved in Cardioembolic Stroke Type and Severity.,” *Brain Sci.*, vol. 10, no. 8, Jul. 2020, doi: 10.3390/brainsci10080476.
[15] S.-C. Tang *et al.*, “Prevalence and clinical characteristics of stroke patients with p.R544C NOTCH3 mutation in Taiwan.,” *Ann. Clin. Transl. Neurol.*, vol. 6, no. 1, pp. 121–128, Jan. 2019, doi: 10.1002/acn3.690.
[16] Y. Xia, “Chapter Eleven - Correlation and association analyses in microbiome study integrating multiomics in health and disease,” in *Progress in Molecular Biology and Translational Science*, vol. 171, J. Sun, Ed., Academic Press, 2020, pp. 309–491. doi: 10.1016/bs.pmbts.2020.04.003.
[17] S. Lolak, J. Attia, G. J. McKay, and A. Thakkinstian, “Comparing Explainable Machine Learning Approaches With Traditional Statistical Methods for Evaluating Stroke Risk Models: Retrospective Cohort Study,” *JMIR Cardio*, vol. 7, p. e47736, Jul. 2023, doi:

10.2196/47736.

Formular list

Formular 1: The margin of error formula is the following

$$MOE = Z\ value * \sqrt{\frac{p(1-p)}{n}}$$

Where:
The Z-value is the critical Z-value that corresponds to your confidence level.
p is the sample proportion or percentage.
n is the sample size.

Formular 2: The sample size calculation formula used is

$$n = \frac{4(Z_\alpha + Z_\beta)^2 \sigma^2}{\Delta^2}$$

In which: The two-sided confidence level is $Z_{\alpha/2}$. α is the possibility of making a type I error. β is the possibility of making a type II error, the power of the study is 1-β. Cohen's effect size d, with d=Δ/σ.

Table list:

Table 1: Two-sample t test power calculation results with power is 0.80, using a two-sided T test with statistical significant level is 0.05. The general guidelines for interpreting the effect size are as follows: 0.2 to 0.49 = small effect, 0.5 to 0.79 = moderate effect, 0.8 to 1.0 = large effect, >1.0 = very large effect

Sample size each group	Cohen's d
99.08	0.4
63.76	0.5
44.58	0.6
33.02	0.7

Table 2: Distribution of patients according to risk factors and genetic variants

	Risk factors	***N***		***%***
Gender	Male	62		62
	Female	38		38
Age (years old)	24-49	22		22
	50-59	23		23
	60-69	37		37
	70-91	18		18
	Min: 24	Max: 91		Mean±SD: 60.14 ± 12.63
BMI	<18.5	3		3
	18.5-22.9	56		56
	23-24.9	27		27
	25-29.9	14		14
	Min: 12.4	Max: 29.4		Mean±SD: 22.62 ± 2.49
Ethic	Kinh	75		75
	Tay	25		25
Smoking		31		31
Alcohol consuming		29		29
High tesion		70		70
diabete		20		20
Brain stroke		35		35
Brain stroke cases in family		44		44
Total		100		100
Gene Mutation		**Wild type**	**Heterozygous**	**Homozygous**

PAI1_4G_5G	24	44	32
FV_1299	**96**	**4**	**0**
FV_Cambridge	*100*	*0*	*0*
MTHFR_1298	**58**	**37**	**5**
FII_prothrombin	98	1	1
FV_Leiden	**93**	**7**	**0**
MTHFR_677	**55**	**37**	**8**
FXIII_Val34Leu	**98**	**1**	**1**
Notch_3	**6**	**91**	**3**

Table 3: Symptoms on admission

Symptoms at entry		***N***	***%***
Hardly talk	No	5	5
	Broka	86	86
	Wernik	9	9
headache		49	49
dizzy		27	27
Nausea/vomiting		8	8
Distorted mouth		72	72
Circular muscle disorder		21	21
Numbness		41	41
Movement paralysis	No	3	3
	Paralysis of the right half of the body	52	52
	Paralysis of the left half of the body	39	39
	Paralysis of the whole body	6	6

Table 4: Average values of some indicators in the study

		N	***Min***	***Max***	***Mean***	***SD***
Age (years old)		100	24	91	60.14	12.63
BMI	<18.5	100	12.4	29.4	22.62	2.49
		100	1	120	10.94	15.98

Hospitalization time (minutes)	<4.5h	33 (33%)				
	4.6h-6h	26 (26%)				
	>6h	41 (41%)				
systolic blood pressure – SBP (mmHg)		100	90	210	148.6	23.66
Diastolic blood pressure – DBP (mmHg)		100	60	120	88.06	9.5
Glasgow		100	8	15	14.72	1.06
NIHSS	0h	100	0	19	7.14	4.33
	24h	100	0	16	6.71	4.26
	Discharge	100	0	16	3.73	3.87
Rankin at Discharge		100	0	5	1.52	1.35
Time of inpatient treatment at the hospital (days)		100	1	23	10.11	4.33

Table 5: Correlation results between the factors

Factor I	Factor II	Factor III	R(I-II)	p_{value}(I-II)	R(II-III)	p_{value} (II-III)
FV_1299	high_tension	age	-0.23	$4.89x10^{-2}$	0.24	$4.06x10^{-2}$
		diastolic			0.43	$9.97x10^{-5}$
		N_stroke			0.23	$4.69x10^{-2}$
		systolic			0.31	$6.75x10^{-3}$
		agehigh2low1			0.34	$2.81x10^{-3}$
		Aspirin			0.23	$4.69x10^{-2}$
		Ca_channel_blocker			0.42	$1.68x10^{-4}$
		Choles			0.31	$6.83x10^{-3}$
		Circular_muscle_disoder			0.28	$1.66x10^{-2}$
		Creatinin			0.26	$2.46x10^{-2}$
		PT_ratio			0.24	$3.92x10^{-2}$
		strokes_story			0.30	$8.42x10^{-3}$
		thrombus_suction			-0.29	$1.29x10^{-2}$
		UCMC			0.51	$2.78x10^{-6}$
Factor I	**Factor II**	**Factor III**	**R(I-II)**	**p(I-II)**	**R(II-III)**	**p(II-III)**
FV_Leiden	APTT_ratio	APTT_second_	-0.32	$4.87x10^{-3}$	0.34	$2.66x10^{-3}$
		nausea			-0.25	$3.10x10^{-2}$
		PT_ratio			-0.23	$4.70x10^{-2}$
	Diazepam	Glucose	0.28	$1.57x10^{-2}$	0.23	$4.41x10^{-2}$
		LVSF			-0.27	$1.80x10^{-2}$
		NIHSS_0h			-0.24	$3.90x10^{-2}$
		NIHSS_24h			-0.23	$4.48x10^{-2}$
		Meclophenoxat			-0.41	$3.00x10^{-4}$

		MTHFR_1298			-0.24	$4.16x10^{-2}$
		gender	0.24	$4.13x10^{-2}$	0.28	$1.40x10^{-2}$
		weight			0.23	$4.70x10^{-2}$
	hardly_talk_	Fibrinogen			0.31	$6.00x10^{-3}$
		Statin			0.33	$3.97x10^{-3}$
Factor I	**Factor II**	**Factor III**	**R(I-II)**	**p(I-II)**	**R(II-III)**	**p(II-III)**
FXIII_Val34Leu	thrombus_suction	APTT_second_	0.59	$3.17x10^{-8}$	-0.31	$6.73x10^{-3}$
	MTHFR_1298	thrombus_suction	0.33	$4.43x10^{-3}$	0.26	$2.29x10^{-2}$
		ethic			0.28	$1.47x10^{-2}$
		Meclophenoxat			0.30	$8.98x10^{-3}$
		Diazepam			-0.24	$4.16x10^{-2}$
		Actylise			-0.23	4.80E-02
		thrombus_suction			0.26	$2.29x10^{-2}$
	NE	alcolhol	0.31	$6.11x10^{-3}$	0.23	$4.86x10^{-2}$
		Circular_muscle_disoder			0.30	$1.02x10^{-2}$
		gender			0.28	$1.41x10^{-2}$
		Glasgow			-0.24	$3.69x10^{-2}$
		Glucose			0.25	$2.84x10^{-2}$
		Hb			0.26	$2.39x10^{-2}$
		vomit			0.29	$1.24x10^{-2}$
		WBC			0.91	<0.00001
		N_infarctsMRI			0.24	$3.81x10^{-2}$
	WBC	Circular_muscle_disoder	0.30	$8.54x10^{-3}$	0.29	$1.29x10^{-2}$
		Hb			0.28	$1.43x10^{-2}$
		Smoking			0.26	$2.48x10^{-2}$
		Meclophenoxat			-0.26	$2.56x10^{-2}$
		N_infarctsMRI			0.27	$2.03x10^{-2}$
	INR	dizzy	0.29	$1.05x10^{-2}$	-0.35	$2.03x10^{-3}$
		headache			-0.31	$6.11x10^{-3}$
		NIHSS_0h			0.37	$1.10x10^{-3}$
		NIHSS_24h			0.39	$5.50x10^{-4}$
		NIHSS_end			0.25	$2.97x10^{-2}$
		Rankin_end			0.29	$1.15x10^{-2}$
		systolic			-0.27	$1.99x10^{-2}$
		PT_ratio			-0.95	<.00001
		PT_second_			0.58	$5.87x10^{-8}$
		thrombus_suction			0.27	$1.97x10^{-2}$
	diastolic	systolic	-0.28	$1.49x10^{-2}$	0.72	$2.66x10^{-13}$
		Hb			0.31	$7.48x10^{-3}$
		HDL_C			0.35	$1.87x10^{-3}$
		high_tension			0.43	$9.97x10^{-5}$
		RBC			0.32	$5.72x10^{-3}$

		UCMC			0.37	1.21x10^{-3}
		APTT_ratio			-0.23	4.70x10^{-2}
		dizzy			0.36	1.65x10^{-3}
		ethic			-0.26	2.43x10^{-2}
		gender			-0.23	4.96x10^{-2}
		headache			0.31	7.14x10^{-3}
		high_tension			0.24	3.92x10^{-2}
		INR			-0.95	<0.00001
	PT_ratio	nausea	-0.27	2.09x10^{-2}	0.23	4.55x10^{-2}
		NIHSS_0h			-0.36	1.54x10^{-3}
		NIHSS_24h			-0.35	2.23x10^{-3}
		PT_second_			-0.51	2.93x10^{-6}
		Rankin_end			-0.28	1.54x10^{-2}
		systolic			0.27	2.01x10^{-2}
		Choles			0.32	5.54x10^{-3}
		thrombus_suction			-0.30	1.02x10^{-2}
		VD			-0.25	3.14x10^{-2}
Factor I	**Factor II**	**Factor III**	**R(I-II)**	**p(I-II)**	**R(II-III)**	**p(II-III)**
		PT_ratio			-0.30	1.02x10^{-2}
MTHFR_1298	thrombus_suction	high_tension	0.26	2.29x10^{-2}	-0.29	1.29x10^{-2}
		INR			0.27	1.97x10^{-2}
		headache			-0.25	3.05x10^{-2}
		EF			-0.24	3.67x10^{-2}
Factor I	**Factor II**	**Factor III**	**R(I-II)**	**p(I-II)**	**R(II-III)**	**p(II-III)**
		NIHSS_0h			0.24	4.17x10^{-2}
MTHFR_677	Actylise	systolic	-0.23	4.80x10^{-2}	0.24	3.91x10^{-2}
		weight			0.28	1.50x10^{-2}
Factor I	**Factor II**	**Factor III**	**R(I-II)**	**p(I-II)**	**R(II-III)**	**p(II-III)**
	Statin	hardly_talk_	-0.37	1.21x10^{-3}	0.33	3.97x10^{-3}
		gender			0.72	2.70x10^{-13}
		Actylise			0.30	8.32x10^{-3}
		AF			-0.33	3.85x10^{-3}
		alcolhol			0.50	5.78x10^{-6}
		Creatinin			0.28	1.57x10^{-2}
		Hb			0.35	2.39x10^{-3}
Notch_3	height	LVSF	-0.29	1.24x10^{-2}	0.27	2.01x10^{-2}
		N_infarctsMRI			0.31	7.19x10^{-3}
		NIHSS_0h			0.30	9.27x10^{-3}
		NIHSS_24h			0.23	4.35x10^{-2}
		PLT			-0.26	2.57x10^{-2}
		RBC			0.26	2.34x10^{-2}
		Smoking			0.39	5.14x10^{-4}
		weight			0.67	3.68x10^{-11}

	Actylise	height	-0.29	$1.25x10^{-2}$	0.30	$8.32x10^{-3}$
	Aspirin	ALT	0.28	$1.58x10^{-3}$	-0.26	$2.33x10^{-2}$
		AST			-0.31	$7.36x10^{-3}$
		gender			-0.23	$4.95x10^{-2}$
		Glasgow			0.24	$3.95x10^{-2}$
		high_tension			0.23	$4.69x10^{-2}$
		NIHSS_0h			-0.35	$2.22x10^{-3}$
		NIHSS_24h			-0.41	$2.45x10^{-4}$
		NIHSS_end			-0.24	$3.61x10^{-2}$
		Rankin_end			-0.33	$3.83x10^{-3}$
		Ure			-0.23	$4.75x10^{-2}$
	APTT_second_	APTT_ratio	-0.25	$2.95x10^{-2}$	0.34	$2.66x10^{-3}$
		Meclophenoxat			0.31	$6.69x10^{-3}$
		thrombus_suction			-0.31	$6.73x10^{-3}$
	gender	alcolhol	-0.23	4.4610^{-2}	0.47	$2.06x10^{-5}$
		Aspirin			-0.23	$4.95x10^{-2}$
		Choles			-0.27	$2.05x10^{-2}$
		Creatinin			0.43	$1.38x10^{-4}$
		hardly_talk_			0.28	$1.40x10^{-2}$
		Hb			0.41	$2.79x10^{-4}$
		height			0.72	$2.70x10^{-13}$
		LVSF			0.28	$1.64x10^{-2}$
		N_infarctsMRI			0.37	$9.66x10^{-4}$
		NE			0.28	$1.41x10^{-2}$
		NIHSS_0h			0.29	$1.22x10^{-2}$
		NIHSS_24h			0.28	$1.41x10^{-2}$
		NIHSS_end			0.26	$2.68x10^{-2}$
		PT_ratio			-0.23	$4.96x10^{-2}$
		Rankin_end			0.32	$4.74x10^{-3}$
		RBC			0.28	$1.50x10^{-2}$
		Smoking			0.53	$1.04x10^{-6}$
		VD			0.26	$2.22x10^{-2}$
		weight			0.50	$3.85x10^{-6}$

Table 6. Correlation between cophenetic distances and the original distance data

Linkage mode	Correlation between cophenetic distances and the original distance data
ward.D	0.515
ward.D2	0.623
single	0.806
complete	0.537
average	0.813
mcquitty	0.694
median	0.750
centroid	0.797

Table 7: Result of clustering imputation. Abbreviations of clustering methods: CH (Calinski and Harabasz 1974), CCC (Sarle 1983), Pseudot2 (Duda and Hart 1973), KL (Krzanowski and Lai 1988), Gamma (Baker and Hubert 1975), Gap (Tibshirani et al. 2001), Silhouette (Rousseeuw 1987), Hartigan (Hartigan 1975), Cindex (Hubert and Levin 1976), DB (Davies and Bouldin 1979), Ratkowsky (Ratkowsky and Lance 1978), Scott (Scott and Symons 1971), Marriot (Marriot 1971), Ball (Ball and Hall 1965), Trcovw (Milligan and Cooper 1985), Tracew (Milligan and Cooper 1985), Friedman (Friedman and Rubin 1967), Rubin (Friedman and Rubin 1967), Dunn (Dunn 1974).

Methods	Number clusters	Value Index
kl	3	3.0285
ch	77	185.7114
hartigan	77	Inf
cindex	77	0.1155
db	77	0.051
silhouette	77	0.9932
duda	3	0.9519
pseudot2	3	2.9293
beale	15	-11.4778
ratkowsky	-Inf	0
ball	3	33.8538
ptbiserial	6	0.518
gap	2	0.2312
frey	2	1.1939
mcclain	2	0.167
gamma	71	1
gplus	71	0
tau	5	473.3775
dunn	77	1.6721
sdindex	55	1.8494

dindex	77	0.0001

Table 8. The maximally selected rank statistic to define the optimal threashold of several continuous factors following the ischemic stroke score and discrete elements.

A

Rankin at the hospital discharge												
Event	**numbness**		**dizzy**		**gender**		**Circular muscle disoder**		**distored mouth**		**MEDIAN**	**SD**
Factors	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**		
Creatinin (µmol/l)	85.170	3.258	101.350	2.245	102.400	3.135	103.500	3.441	85.170	3.606	101.350	9.477
Age (years old)	54.000	3.374	44.000	2.022	54.000	2.977	51.000	1.143	54.000	2.290	54.000	4.336
PT second (s)	12.700	3.752	13.100	3.177	13.100	3.295	13.100	2.989	13.100	3.304	13.100	0.179
INR	1.000	3.696	1.000	4.047	1.010	3.274	1.000	2.797	1.010	3.370	1.000	0.005
LDL C (mmol/l)	4.590	1.592	2.130	2.501	4.590	2.487	4.740	2.365	2.210	1.913	4.590	1.355
N of infarcts via CT	2.000	2.858	2.000	2.466	2.000	1.578	2.000	2.624	2.000	1.824	2.000	0.000
PT ratio	97.300	3.679	97.300	3.887	97.000	3.504	99.000	2.752	97.000	3.525	97.300	0.841

B

NIHSS at the hospital admission														
Event	**numbness**		**dizzy**		**gender**		**Circular muscle disoder**		**distored mouth**		**diabete**		**MEDIAN**	**SD**
Factors	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**		
Creatinin (µmol/l)	82.170	2.667	75.640	1.819	82.170	1.434	82.170	2.059	82.170	3.328			82.170	2.920
Age (years old)	54.000	2.788	44.000	1.400	54.000	1.400	44.000	1.652	44.000	0.792			44.000	5.477
PT second (s)	13.600	2.936	13.400	3.273	13.400	2.670	14.000	1.358	12.400	2.208			13.400	0.590
INR	1.010	2.988	1.000	3.826	1.060	2.428	0.970	1.593	0.980	2.285	0.990	2.637	0.995	0.032
LDL C (mmol/l)	2.610	1.562	2.130	1.712	3.620	2.001	4.740	1.724	2.230	1.663			2.610	1.106
N of infarcts via CT	2.000	2.047	2.000	1.797	1.000	1.264	2.000	2.235	1.000	1.729			2.000	0.548
PT ratio	103.000	3.050	97.300	3.592	90.300	2.295	103.000	1.627	102.000	2.715			102.000	5.466
Height (cm)			153.000	2.659	155.000	0.982	169.000	2.142	161.000	2.533	162.000	0.951	161.000	6.325
N infarcts via MRI			2.000	2.203	2.000	0.356	2.000	0.851	2.000	1.399	2.000	1.796	2.000	0.000

C

NIHSS after 24h												
event	**numbness**		**dizzy**		**gender**		**Circular muscle disoder**		**distored mouth**		**MEDIAN**	**SD**
Factors	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**		
Creatinin (µmol/l)	82.170	2.726	101.350	2.307	82.170	1.707	82.170	2.024	82.170	3.425	82.170	8.578
Age (years old)	54.000	3.061	44.000	1.666	55.000	2.172	59.000	0.821	54.000	1.424	54.000	5.541

PT second (s)	12.900	3.054	13.400	3.426	13.400	3.423	13.300	2.373	13.400	3.011	13.400	0.217
INR	0.950	3.070	1.000	3.877	1.040	3.052	1.060	1.767	1.070	2.784	1.040	0.049
LDL C (mmol/l)	2.610	1.668	2.130	2.118	4.160	3.095	4.740	2.730	4.160	2.406	4.160	1.125
N of infarcts via CT	2.000	2.540	2.000	2.567	2.000	1.741	2.000	2.850	2.000	2.438	2.000	0.000
PT ratio	114.500	3.302	97.300	3.705	93.000	2.686	99.000	1.725	99.000	2.615	99.000	8.170
MPV					6.000	1.111	6.600	1.144	6.500	1.261	6.500	0.321

D

NIHSS at the hospital admission												
Event	**numbness**		**dizzy**		**gender**		**Circular muscle disoder**		**distored mouth**			
Factors	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**MEDIAN**	**SD**
Creatinin (µmol/l)	77.700	2.608	101.350	2.024	100.500	1.053	77.830	1.755	85.170	3.120	85.170	11.733
Age (years old)	54.000	3.295	44.000	2.083	54.000	3.089	59.000	0.857	54.000	2.052	54.000	5.477
PT second (s)	12.700	3.380	13.100	2.866	13.600	2.895	13.100	1.689	13.800	2.627	13.100	0.439
INR	1.010	2.923	1.000	3.657	1.070	3.045	1.060	1.529	1.070	2.778	1.060	0.034
LDL C (mmol/l)	2.610	1.779	2.130	2.290	4.310	3.056	4.740	2.701	4.740	2.203	4.310	1.244
N of infarcts via CT	2.000	2.552	2.000	2.310	1.000	1.506	2.000	2.736	2.000	2.293	2.000	0.447
PT ratio	114.500	3.346	97.300	3.410	92.000	2.412	99.000	1.179	99.000	2.207	99.000	8.410
MPV (fL)			9.600	1.951					6.500	1.190	8.050	2.192

E

Glasgow						
Event	**Circular muscle disoder**		**ethic**			
Factors	**cutpoint**	**statistic**	**cutpoint**	**statistic**	**MEDIAN**	**SD**
BMI	20.576	3.513	20.576	3.422	20.576	0.000

F

	Median	**SD**
Creatinin (µmol/l)	83.670	9.199
Age (years old)	54.000	5.000
PT second (s)	13.250	0.173
INR	1.020	0.031
LDL C (mmol/l)	4.235	0.890
N of infarcts via CT	2.000	0.000
PT ratio	99.000	1.955
Height (cm)	161.000	6.325
N infarcts via MRI	2.000	0.000
MPV (fL)	7.275	1.096
BMI	20.576	0.000

Table 9: The most relevence factors related with the diagnostic and stroke outcomes (Ischemic Stroeke score) via Risk Ratio results.

		Glasgow >12.77		NIHSS at admission >9.83±2.85	NIHSS at 24h >7.92±2.04	NIHSS at discharged >6.85±2.90	Rankin at discharged >2.86±1.21			
BMI	≤20.58		✓							
	>20.58	✓								
INR	≤1.02±0.03									
	>10.2±0.03						✓	✓	✓	✓
PT time	≤13.25±0.17 seconds									
	>13.25±0.17 seconds			✓	✓	✓	✓	✓	✓	✓
PT ratio	≤99±1.96			✓	✓	✓	✓	✓	✓	✓
	>99±1.96									
Creatinin	≤83.67±9.2 µmol/l									
	>83.67±9.2 µmol/l			✓	✓	✓	✓	✓	✓	✓
N of infarcts via CT	≤2							✓	✓	✓
	>2						✓			
FII prothrombin	wildtype	✓				✓				
	heterozygous									
	homozygouszygous									
MTHFR 677	wildtype	✓								
	heterozygous					✓				
	homozygouszygous									
MTHFR 1298	wildtype		✓						✓	✓
	heterozygous							✓		
	homozygouszygous									
FV1299	wildtype		✓					✓	✓	
	heterozygous									
	homozygouszygous									
Notch3	wildtype									
	heterozygous	✓				✓				
	homozygouszygous									
diabetes	yes									
	no	✓								
FXIII Val34Leu	wildtype			✓	✓	✓	✓			✓
	heterozygous									
	homozygouszygous									
age	≤54±5y.o									
	>54±5y.o			✓	✓	✓				
height	≤161±6.3cm			✓	✓					
	>161±6.3cm					✓				
RR[95%CI]		1.23 [0.99,1.54]	0.79 [0.61,1.01]	2.72 [1.4,5.31]	2.09 [1.1,3.95]	4.8 [1.53,15.04]	3.96 [2.82,5.56]	3.13 [1.6,6.11]	3.13 [1.6,6.11]	3.13 [1.6,6.11]
p_{fisher}		2.68E-03	1.72E-03	2.19X10-2	8.81X10-2	3.47X10-2	1.30E-01	2.64X10-2	2.64X10-2	2.64X10-2
p_{yates}		2.07E-03	9.51E-04	3.73X10-2	9.72X10-2	5.46X10-2	2.91E-01	4.47X10-2	4.47X10-2	4.47X10-2
p_{uncor}		2.90E-04	1.07E-04	1.20X10-2	4.07X10-2	8.49E-03	4.50X10-2	1.13X10-2	1.13X10-2	1.13X10-2

Figures list:

Figure 1: Patient Age distribution. The age group from 24 to 49 in 22 patients (accounting for 22%), from 50 to 59 in 23 patients (accounting for 23%), from 60 to 69 were 37 patients (accounting for 37%), from 70 to 91 were 18 (accounting for 18%).

Figure 2: PCR-CTPP identified the NOTCH3 p.R544C and other gene variants were detected by RT-PCR

Figure 3: RT-PCR identified the FV Leiden mutation

Figure 4: RT-PCR identified the FV H1299R mutation

Figure 5: RT-PCR identified the MTHFR C677T mutation

Figure 6: RT-PCR identified the MTHFR A1298 mutation

Figure 7: Correlation Heatmap of 79 factors on 100 Ischemic stroke patients. The age group from 24 to 49 in 22 patients (accounting for 22%), from 50 to 59 in 23 patients (accounting for 23%), from 60 to 69 were 37 patients (accounting for 37%), from 70 to 91 were 18 (accounting for 18%).

Interactive HTML graph 1: Heatmap A, 79factors-A vs 79factors-A | heatmap made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/9/#/)

Interactive HTML graph 2: Heatmap B, 79factors-B vs 79factors-B | heatmap made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/11/#/)

Interactive HTML graph 3: Heatmap C: 79factors-C vs 79factors-C | heatmap made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/13/#/)

Interactive HTML graph 4: Heatmap D: 79factors-D vs 79factors-D | heatmap made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/15/#/)

Figure 8: The volcano graph shows the most significant correlation pairs.

Related Interactive HTML graph: -log10(P) vs R | scatter chart made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/3/#/)

Figure 9: ***Results of hierarchical cluster analysis (HCA) on overall dataset. (A) Dendrogram (B) PCA map.***

Related Interactive HTML graph: overall cohort, k-means optimal=15,average model,beale index | scatter chart made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/7/#/)

Figure 10: ***Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS after 24h and at the discharge for the Creatinin >83.67±9.19umol/l***

Figure 11: ***Significant cutpoint 50 of the NIHSS at the discharge for the patient ages > 54±5years old***

Figure 12: ***Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge for the Prothrombin time > 13.25±0.17 seconds (time of PT)***

Figure 13: ***Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge for the ratio of Prothrombin > 99.00±1.96 (ratio of PT)***

Figure 14: ***Significant cutpoint 50 of the Modified Rankin Score (mRS) for INR> 1.02±0.03***

Figure 15: ***Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the discharge for the Number of infarcts via CT > 2***

Figure 16: ***Significant cutpoint 50 of the NIHSS at the admission for the patients height > 161±6.3cm***

Figure 17: ***Significant cutpoint 50 of the Glasgow Coma Scale (GCS) for the BMI>20.58.***

Figure 18: Dotplot of ***these genotype variants following their cluster which give the most relevant risk ratio results.***

Related Interactive HTML graph: The most revelant factors | scatter chart made by Hocamtu | plotly (https://chart-studio.plotly.com/~hocamtu/1/#/)

Supplemental Material list:

Table S1: The most relevence factors related with the diagnostic and stroke outcomes (following the ischemic Stroke score) via Risk Ratio results.

Figure S1: Forest plot represents Risk for Rankin > 2.86, of cluster 4 (*FII prothrombin genotype,MTHFR C677T genotype, Notch 3 p.R544C genotype,* diabetes status) with other factors: INR, time and ratio of PT, Creatinin, Number of infarcts via CT.

Figure S2: Forest plot represents Risk for NIHSS 0h > 9.83, of Cluster 4 (*FII prothrombin*

genotype,MTHFR C677T genotype, Notch 3 p.R544C genotype, diabetes status) with other factors: time and ratio of PT, Creatinin, age, height

Figure S3: forest plot of Cluster 4 (*FII prothrombin genotype,MTHFR C677T genotype, Notch 3 p.R544C genotype*, diabetes status) with other factors: Status of Factors : time and ratio of PT, Creatinin, age, height Risk for NIHSS 24h > 7.92

Figure S4: Forest Plot represents the Risk for NIHSS end > 6.85, of Cluster 4 *FII prothrombin genotype,MTHFR C677T genotype, Notch 3 p.R544C genotype*, diabetes status) with other factors Status of Factors: time and ratio of PT, Creatinin, age, height

Figure S5: Forest plot represents the Risk for Glasgow > 12.77, of Cluster 4 (*FII prothrombin genotype,MTHFR C677T genotype, Notch 3 p.R544C genotype*, diabetes status) with BMI

Figure S6: Forest plot represents for the Risk for Rankin > 2.86, of Cluster 6 (*FV Leiden genotype, PAI1 4G 5G genotype, FV Cambridge genotype*) with other factors: INR, time and ratio of PT, Creatinin, Number of infarcts via CT

Figure S7: Forest plot represents the Risk for NIHSS 0h > 9.83, of Cluster 6 *(FV Leiden genotype, PAI1 4G 5G genotype, FV Cambridge genotype*) with other factors : time and ratio of PT, Creatinin, age, height

Figure S8: Forest plot represents the Risk for NIHSS 24h > 7.92, of Cluster 6 (*FV Leiden genotype, PAI1 4G 5G genotype, FV Cambridge genotype*) with other factors : time and ratio of PT, Creatinin, age, height

Figure S9: Forest plot represents the Risk for Glasgow > 12.77, of Cluster 6 (*FV Leiden genotype, PAI1 4G 5G genotype, FV Cambridge genotype*), with BMI

Figure S10: Forest plot represents the Risk for Rankin > 2.86, of Cluster 11(*MTHFR A1298C genotype, FV H1299R genotype*) with others factors : INR, time and ratio of PT, Creatinin, Number of infarcts via CT

Figure S11: Forest plot represents the Risk for NIHSS 0h > 9.83, of Cluster 11 (*MTHFR A1298C genotype, FV H1299R genotype*) with other factors : time and ratio of PT, Creatinin, age, height

Figure S12: Forest plot represents the Risk for NIHSS 24h > 7.92, of Cluster 11 (*MTHFR A1298C genotype, FV H1299R genotype*) with other factors: time and ratio of PT, Creatinin, age, height

Figure S13: Forest plot represents the Risk for Glasgow > 12.77, of Cluster 11(*MTHFR A1298C genotype, FV H1299R genotype*) with BMI

Figure S14: Forest plot represents the Risk for Rankin > 2.86, of Cluster 13 (*FXIII Val34Leu genotype*) with other factors (INR, time and ratio of PT, Creatinin, Number of infarcts via CT)

Figure S15: Forest plot represents the Risk for NIHSS 0h > 9.83, of Cluster 13 (*FXIII Val34Leu genotype*) with the factors: time and ratio of PT, Creatinin, age, height

Figure S16: Forest plot represents the Risk for NIHSS 24h > 7.92, of Cluster 13 (*FXIII Val34Leu genotype*) with other factors : time and ratio of PT, Creatinin, age, height

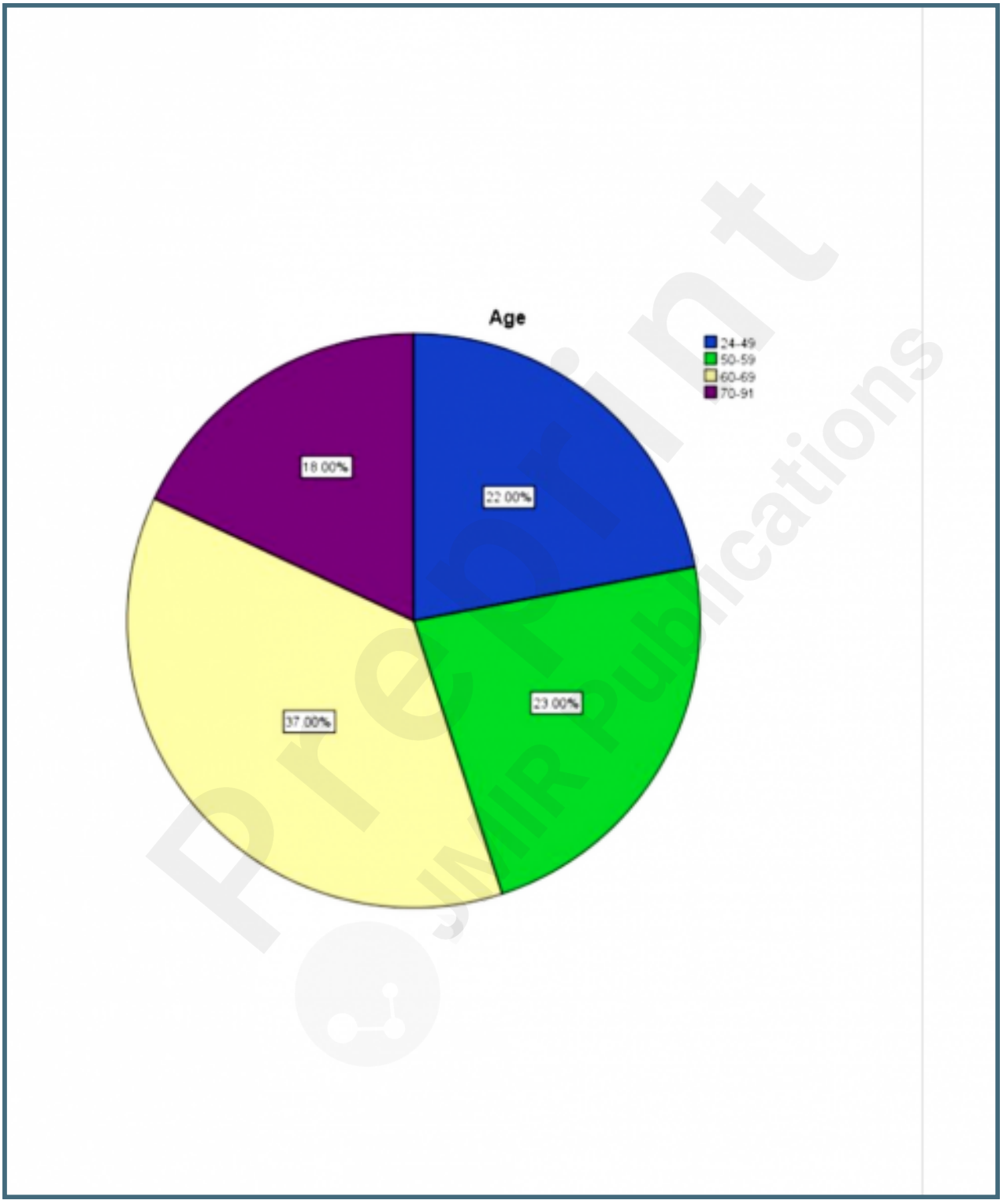
Figure S17: Forest plot represents the Risk for Glasgow > 12.77, of Cluster 13 (*FXIII Val34Leu genotype*) with BMI

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/56f6806cb1d6b998753cc0e21066a361.docx

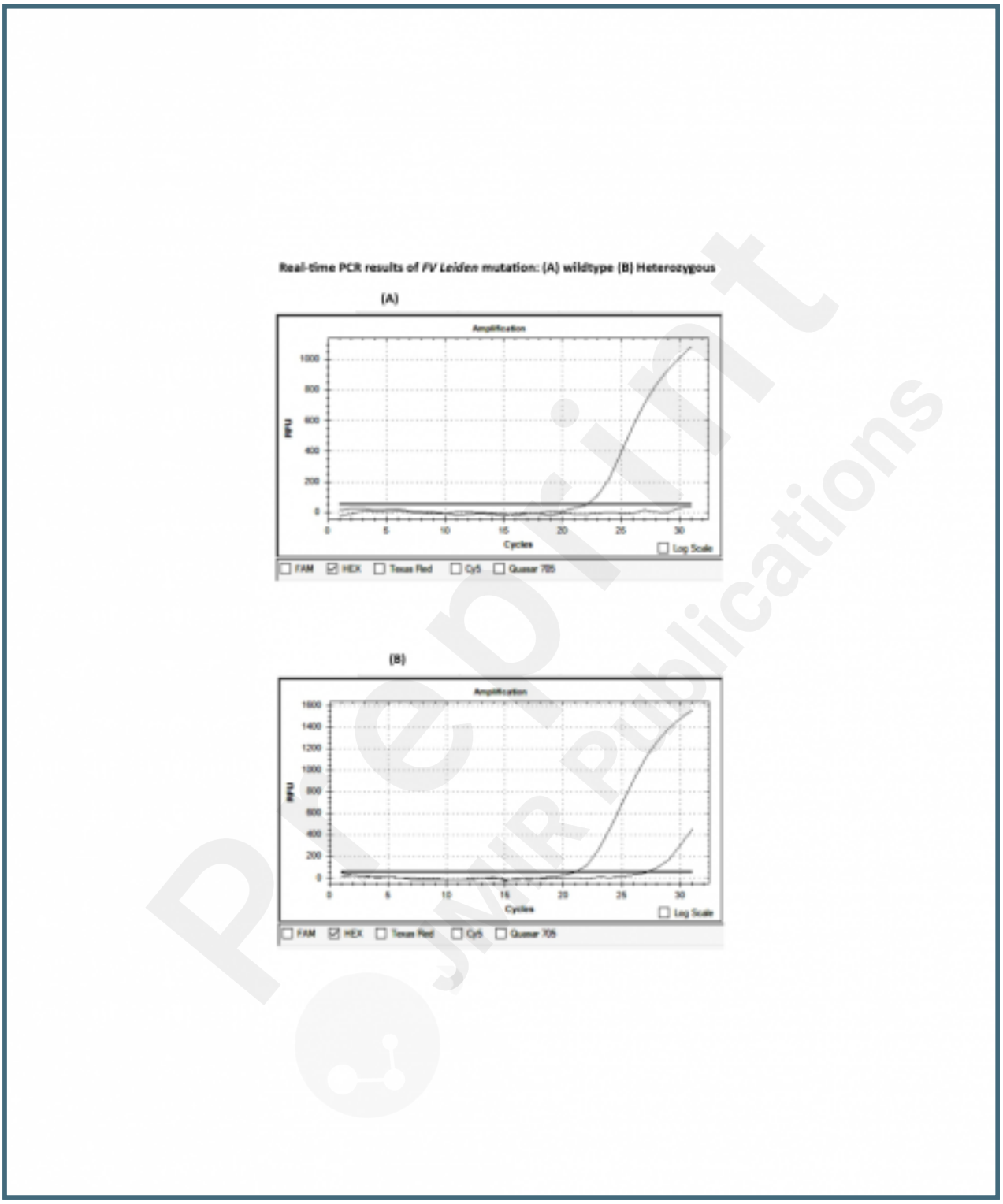
Figures

Patient Age distribution. The age group from 24 to 49 in 22 patients (accounting for 22%), from 50 to 59 in 23 patients (accounting for 23%), from 60 to 69 were 37 patients (accounting for 37%), from 70 to 91 were 18 (accounting for 18%).

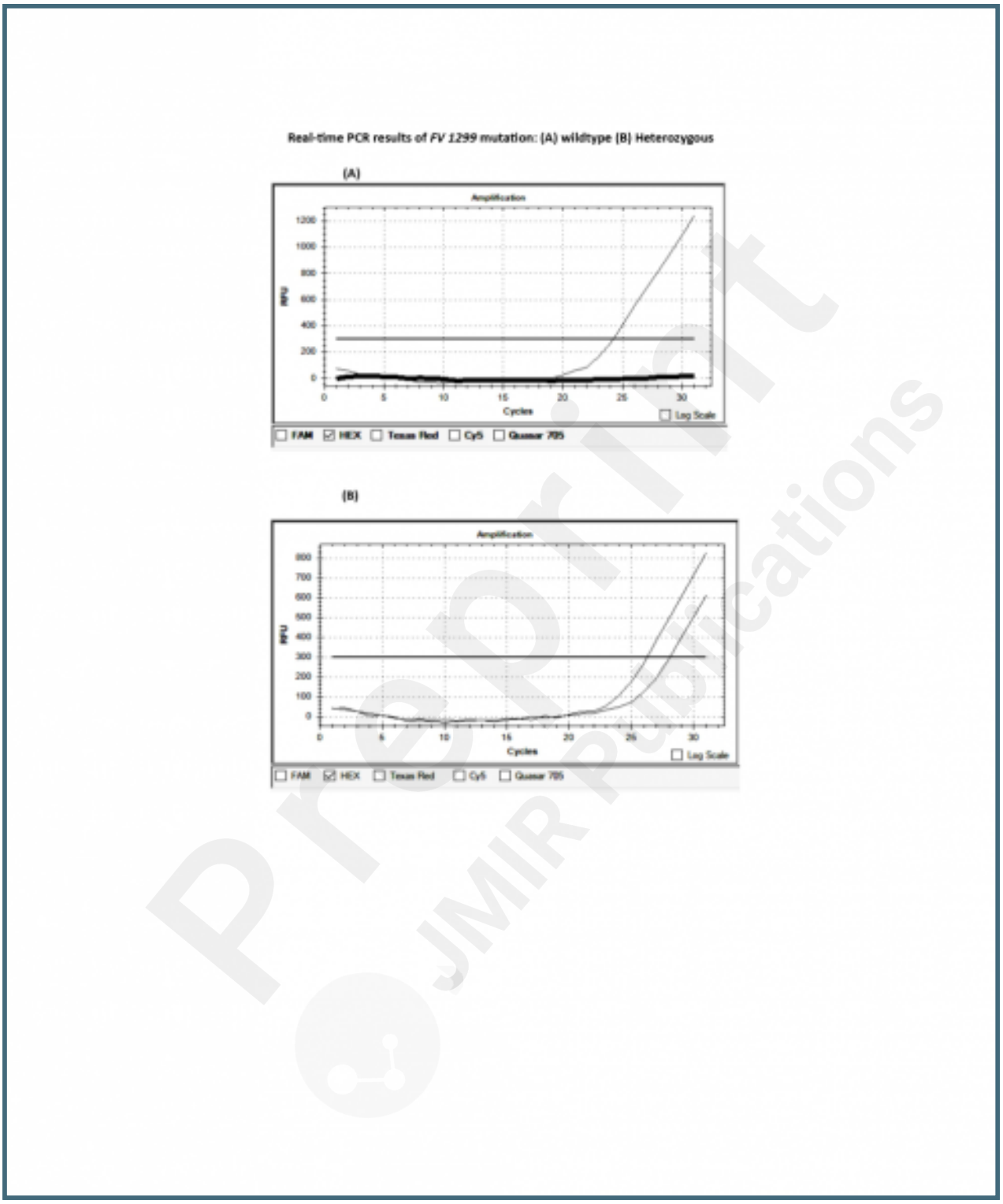


PCR-CTPP identified the NOTCH3 p.R544C and other gene variants were detected by RT-PCR.

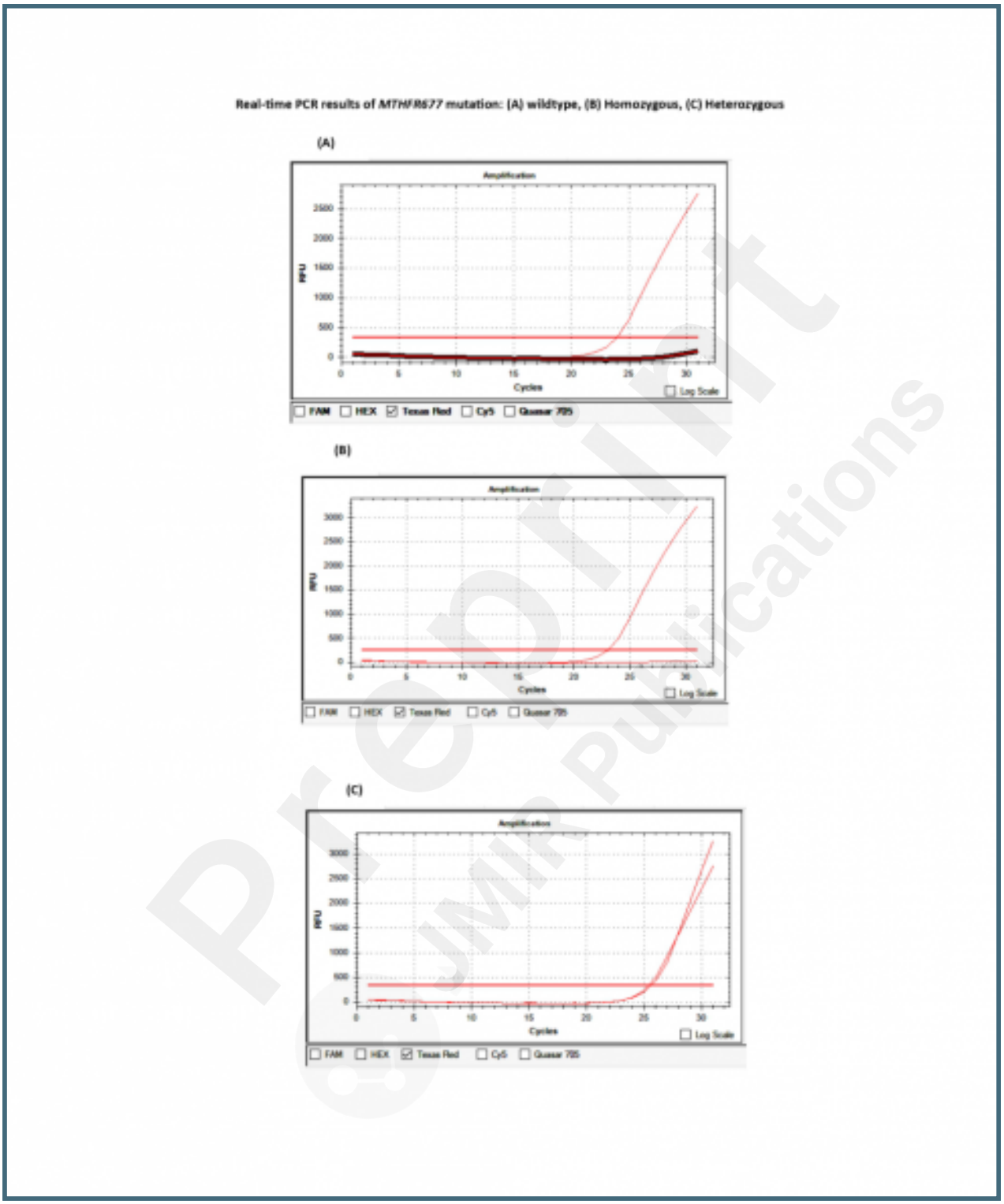
RT-PCR identified the FV Leiden mutation.



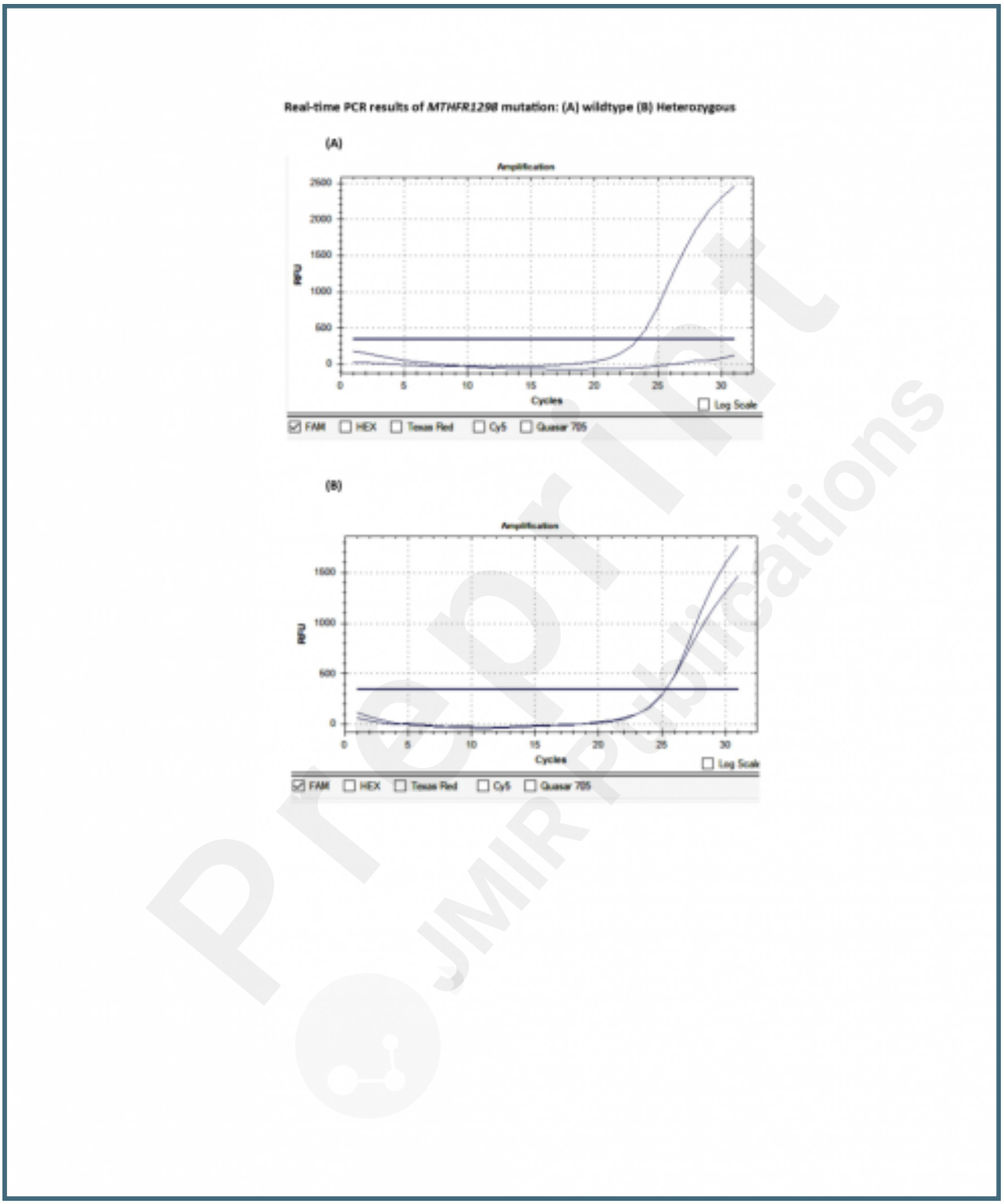
RT-PCR identified the FV H1299R mutation.



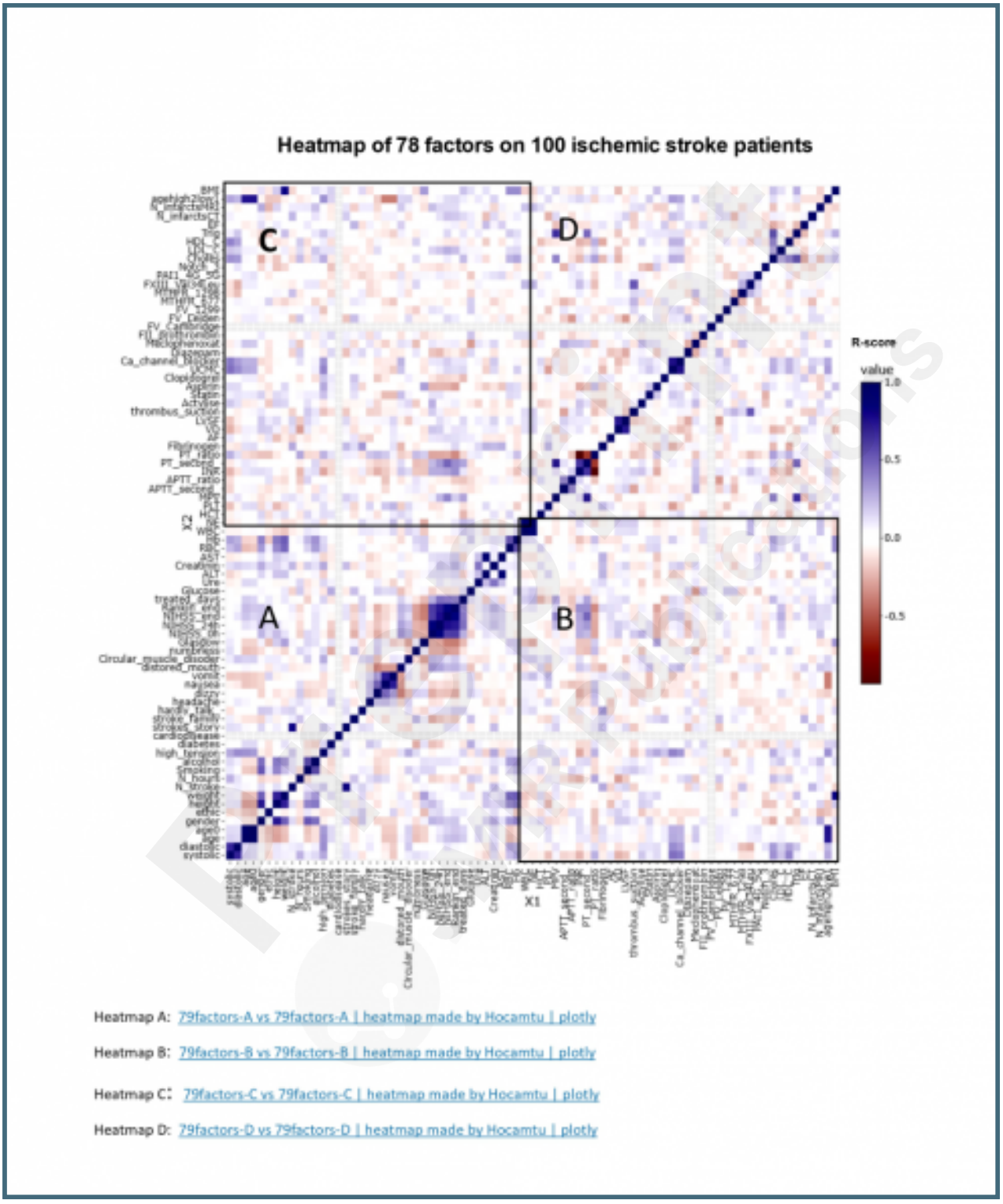
RT-PCR identified the MTHFR C677T mutation.



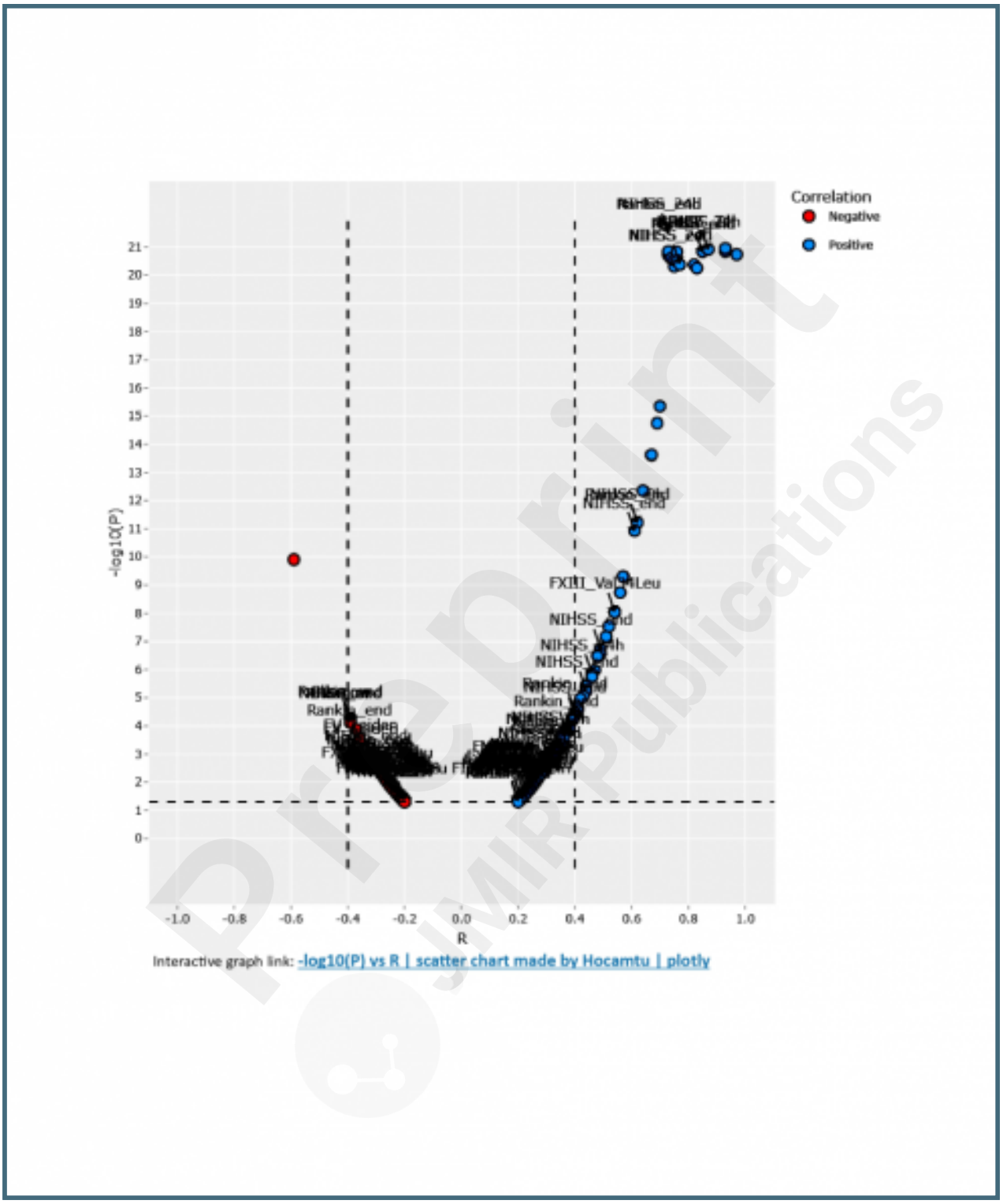
RT-PCR identified the MTHFR A1298 mutation.



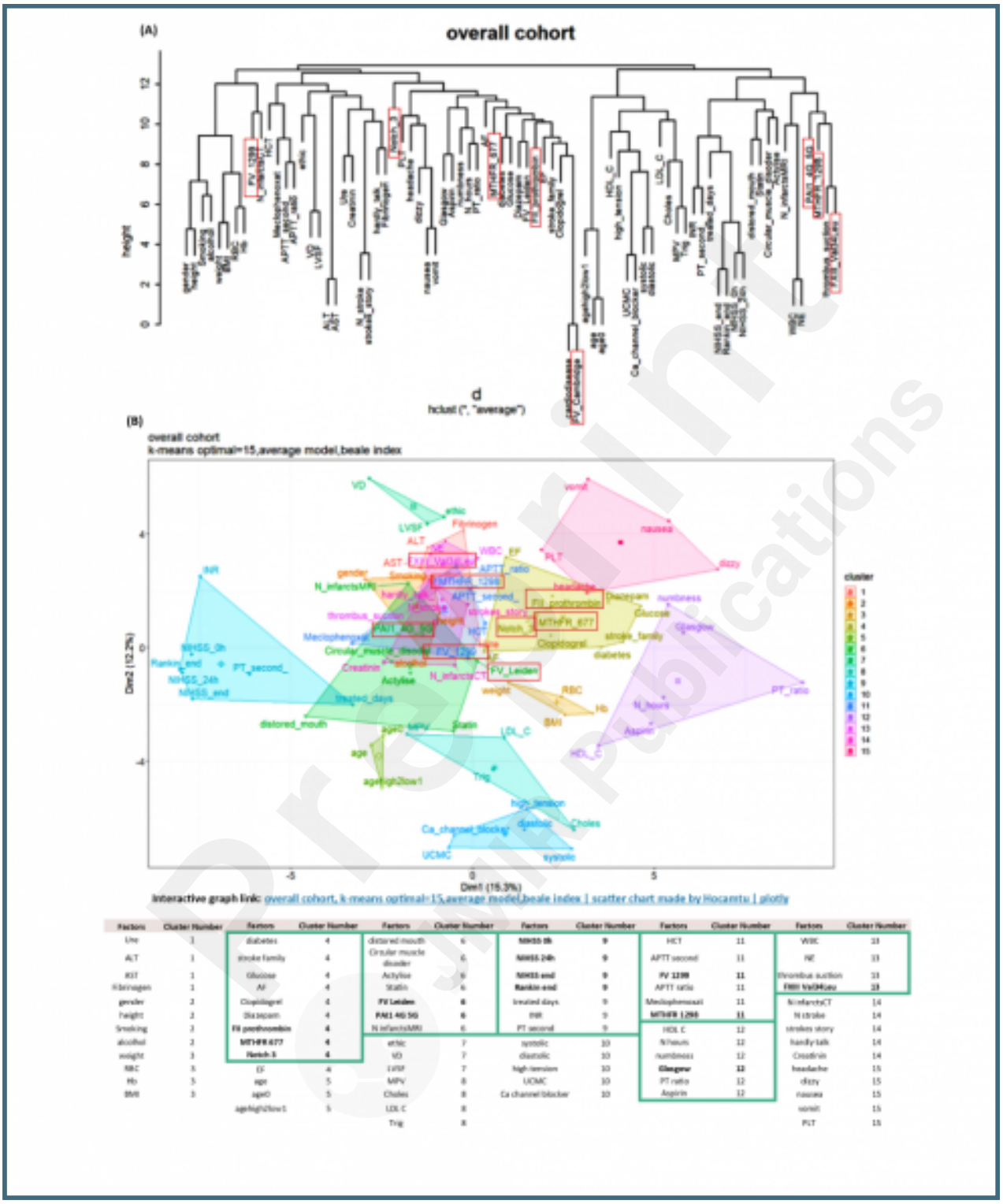
Correlation Heatmap of 79 factors on 100 Ischemic stroke patients. The age group from 24 to 49 in 22 patients (accounting for 22%), from 50 to 59 in 23 patients (accounting for 23%), from 60 to 69 were 37 patients (accounting for 37%), from 70 to 91 were 18 (accounting for 18%).



The volcano graph shows the most significant correlation pairs. .



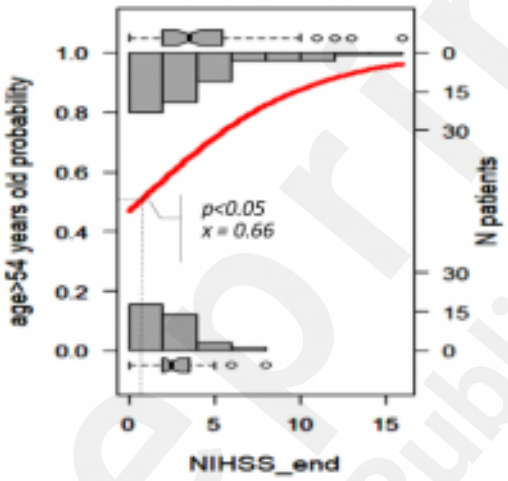
Results of hierarchical cluster analysis (HCA) on overall dataset. (A) Dendrogram (B) PCA map.



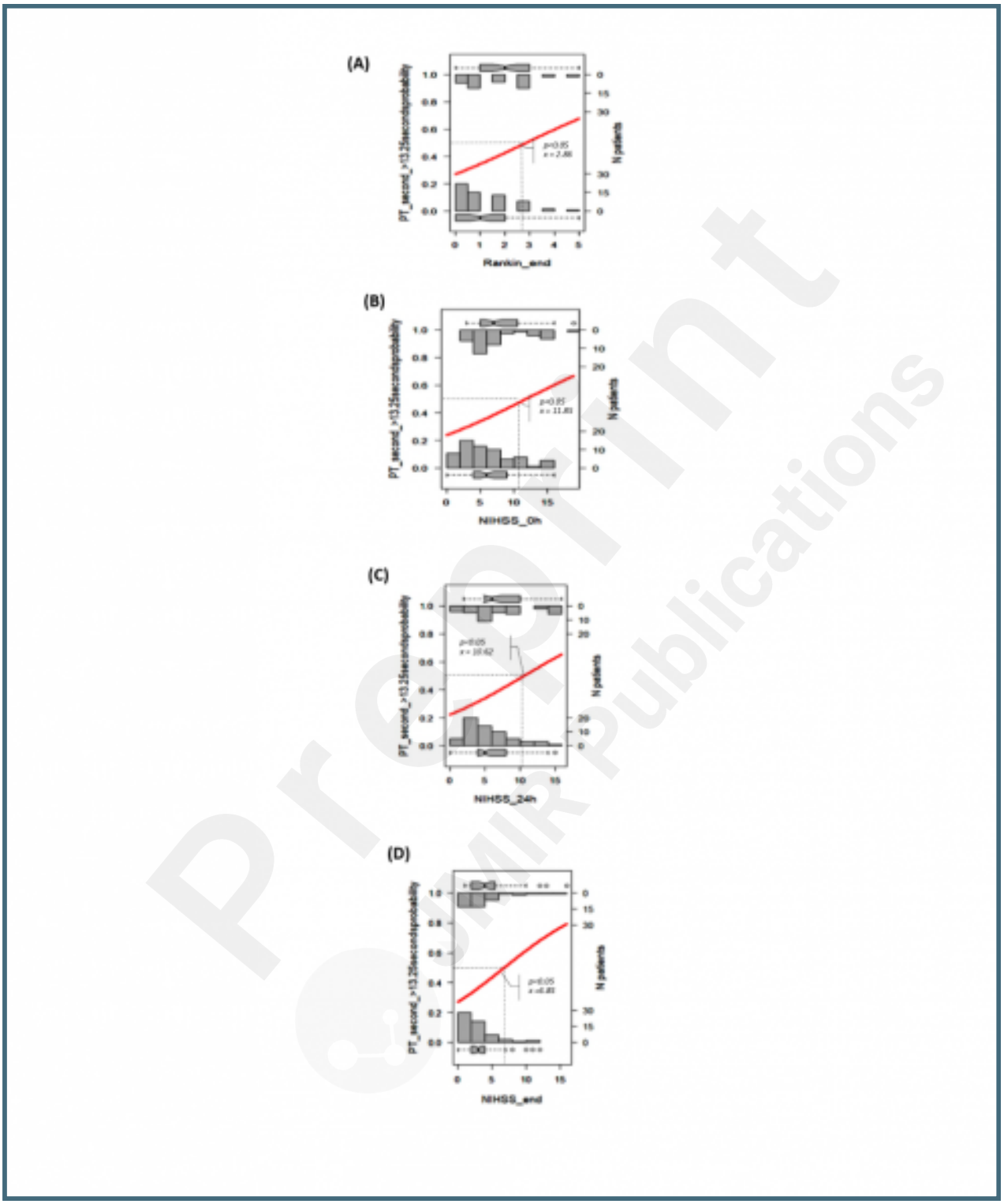
Factors	Cluster Number	Factors	Cluster Number	Factors	Cluster Number	Factors	Cluster Number	Factors	Cluster Number	Factors	Cluster Number
Ure	1	diabetes	4	distored mouth	6	NIHSS 0h	9	HCT	11	WBC	13
ALT	1	stroke family	4	Circular muscle disoder	6	NIHSS 24h	9	APTT second	11	RE	13
INR	1	Glucose	4	Actyllse	6	NIHSS end	9	FV 1299	11	thrombus suction	13
Fibrinogen	1	AF	4	Statin	6	Rankin end	9	APTT ratio	11	FIBI Val34Leu	13
gender	2	Clopidogrel	4	FV Leiden	6	treated days	9	Meclophenoxat	11	N infarctsCT	14
height	2	Diazepam	4	PAI1 4G 5G	6	INR	9	MTHFR 1298	11	N stroke	14
Smoking	2	Fii prothrombin	4	N infarctsMRI	6	PT second	9	HDL C	12	strokes story	14
alcohol	2	MTHFR 677	4	ethic	7	systolic	10	N heurs	12	handy talk	14
weight	3	Notch 3	4	VD	7	diastolic	10	numbness	12	Creatinin	14
MBC	3	Cf	4	LVSF	7	high tension	10	Glasgow	12	headache	15
Hb	3	age	5	MPV	8	UCMC	10	PT ratio	12	dizzy	15
BMI	3	age0	5	Choles	8	Ca channel blocker	10	Aspirin	12	nausea	15
		agehigh2low1	5	LDL C	8					vomit	15
				Trig	8					PLT	15

Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS after 24h and at the discharge for the Creatinin >83.67±9.19umol/l.

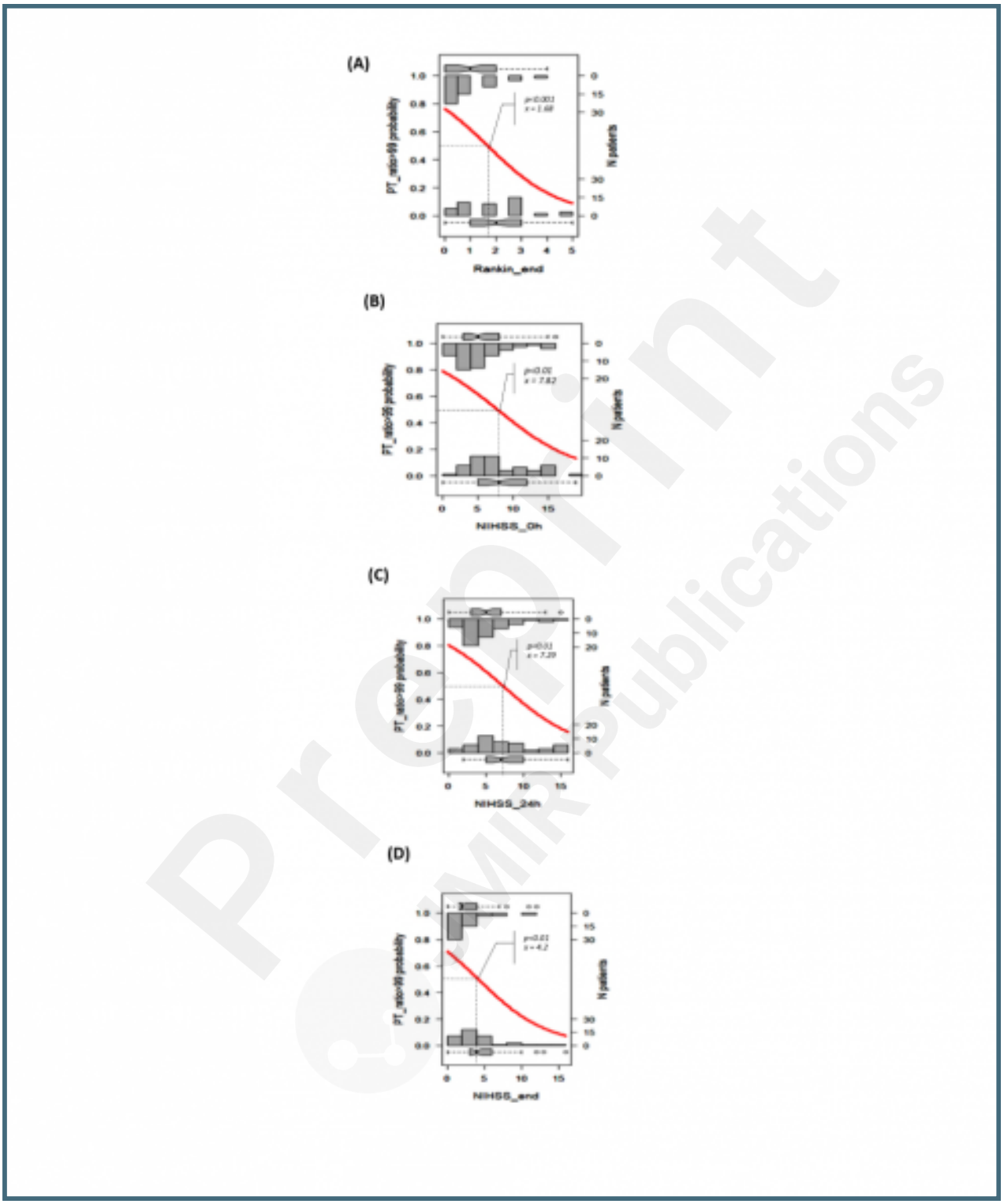
Significant cutpoint 50 of the NIHSS at the discharge for patients ages > 54±5years old.



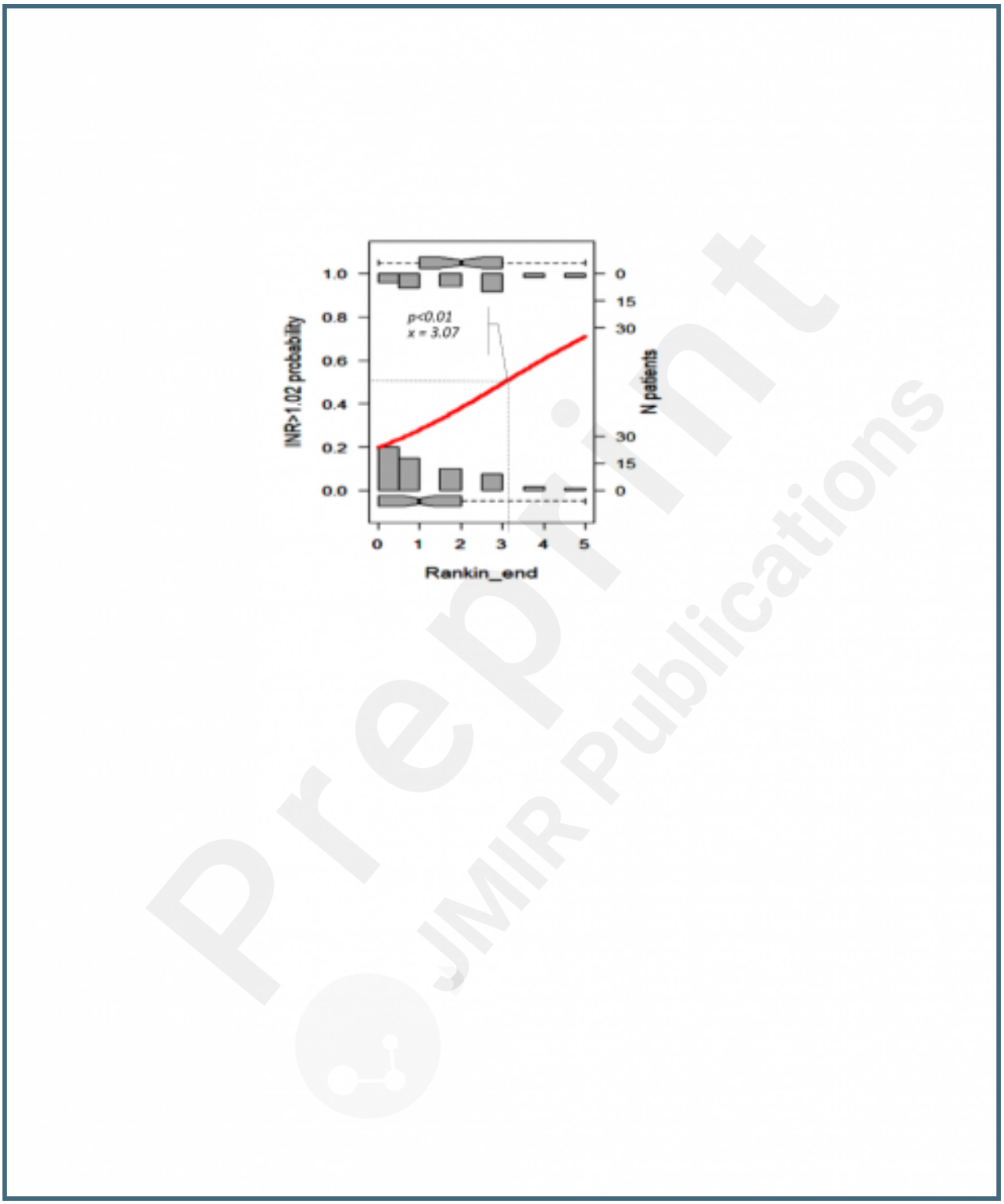
Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge for the Prothrombin time > 13.25±0.17 seconds (time of PT).



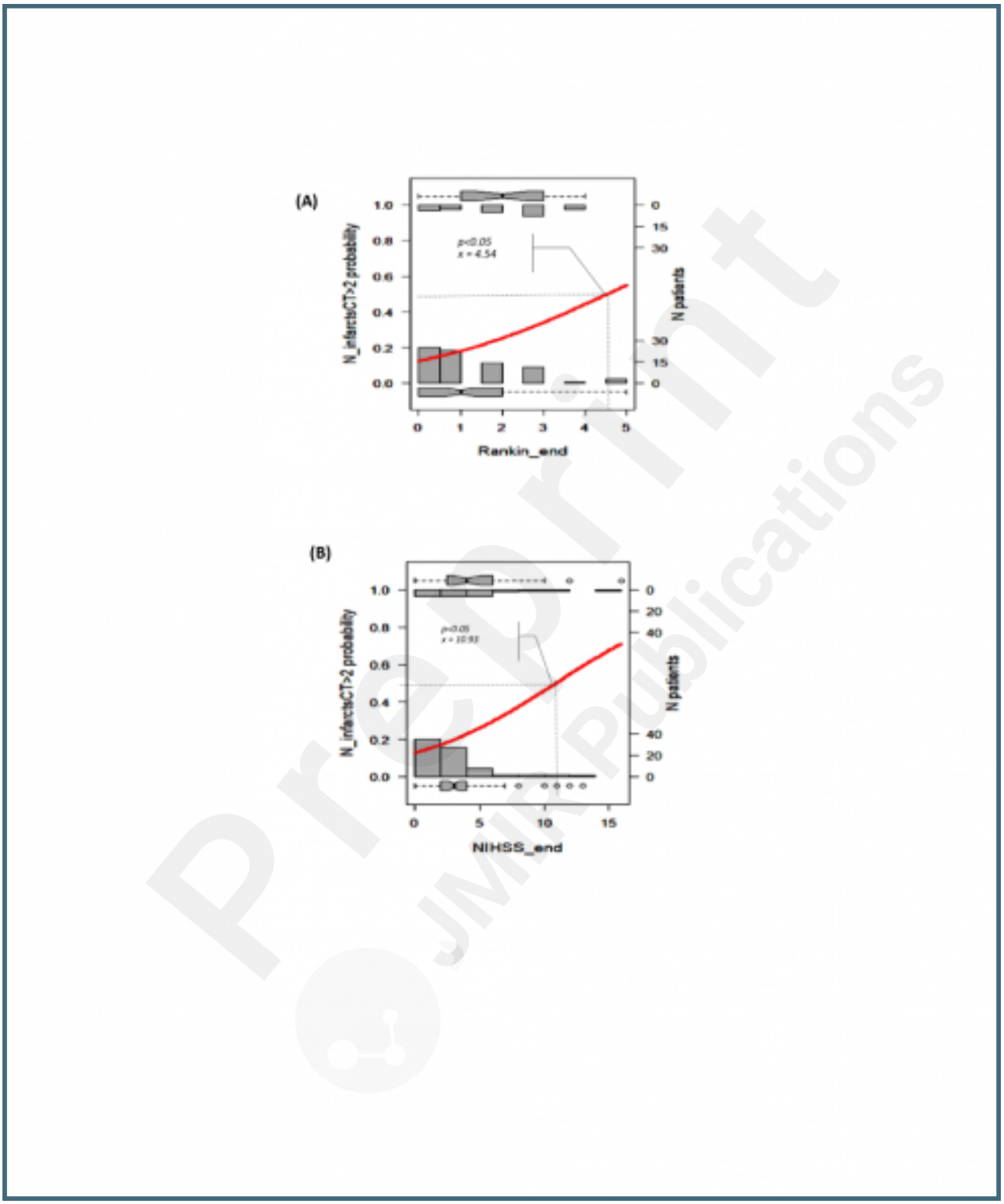
Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the admission, after 24h and at the discharge for the ratio of Prothrombin > 99.00±1.96 (ratio of PT).



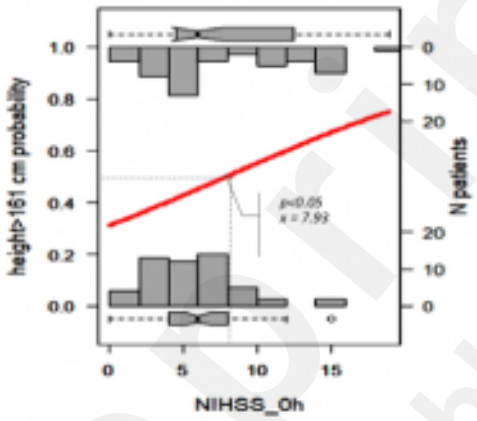
Significant cutpoint 50 of the Modified Rankin Score (mRS) for INR> 1.02±0.03.



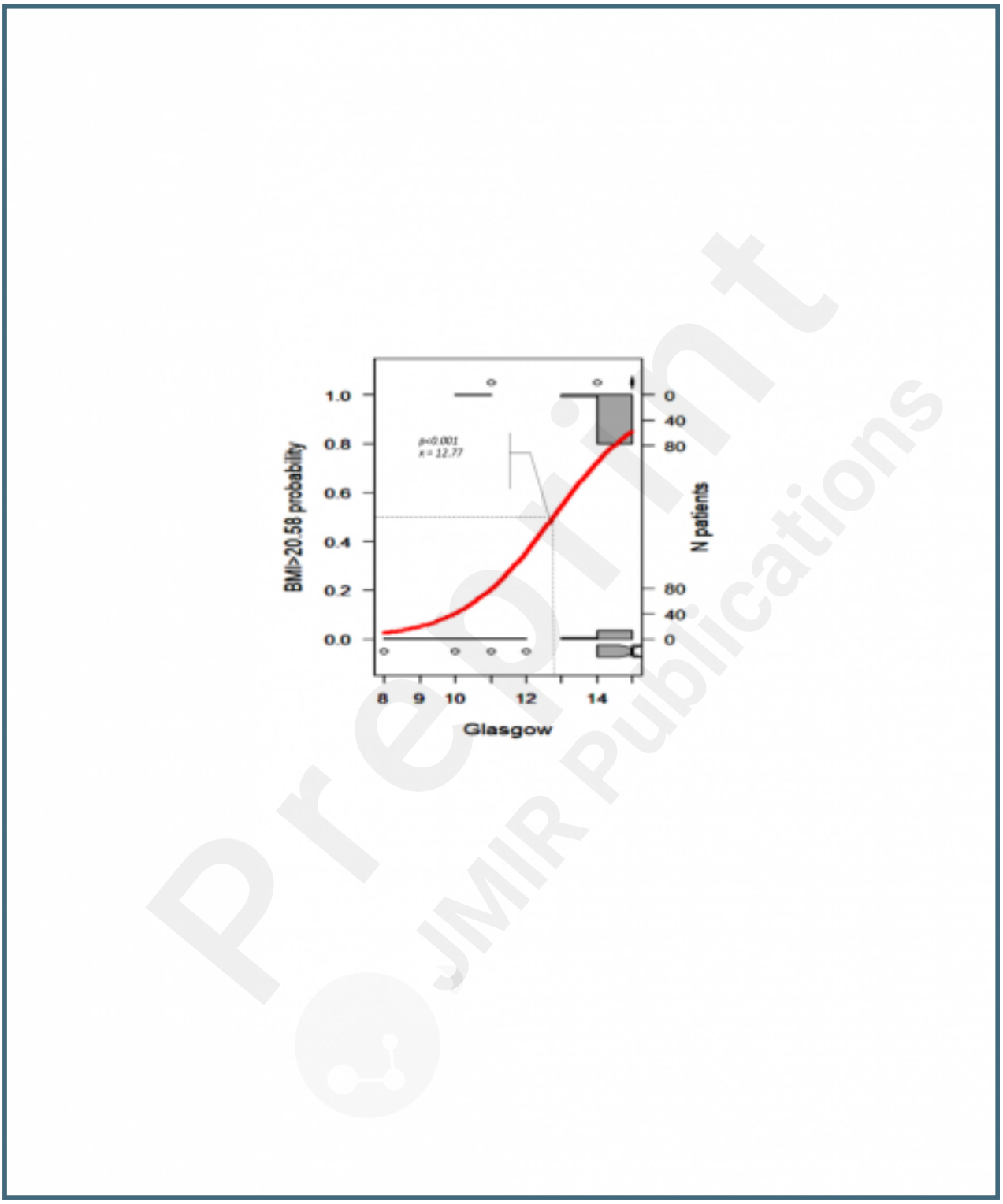
Significant cutpoint 50 of the Modified Rankin Score (mRS); the NIHSS at the discharge for the Number of infarcts via CT > 2
.



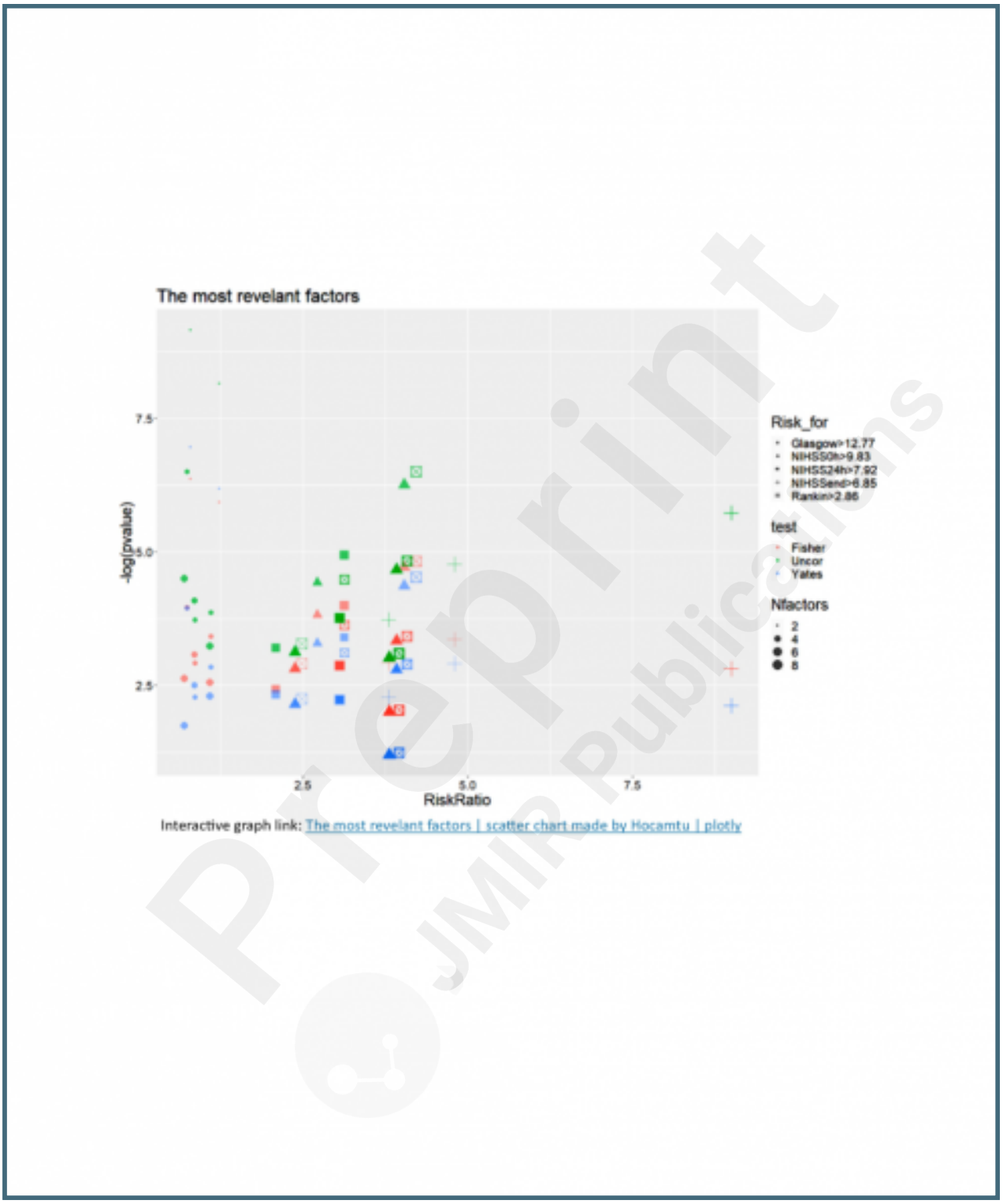
Significant cutpoint 50 of the NIHSS at the admission for the patients height > 161±6.3cm.



Significant cutpoint 50 of the Glasgow Coma Scale (GCS) for the BMI>20.58.



Dotplot of these genotype variants following their cluster which give the most relevant risk ratio results..



Multimedia Appendixes

Supplemental tables and figures.
URL: http://asset.jmir.pub/assets/274191410c593f55acfd4ab52d781de1.pdf

CONSORT (or other) checklists

Clinical Studies Checklist.
URL: http://asset.jmir.pub/assets/bf36b6f59ea903ffa0b8e2e2d98f8943.pdf

STROBE checlist v4 combined.
URL: http://asset.jmir.pub/assets/c86e50d2d2f4d404e8c4cc386297e99c.pdf